UNIVERSIDADE
BEIRA INTERIOR

Large-Scale Integration of Variable Renewable Energies in Electric Power Markets — Optimization Models, Options and Challenges

Pedro Miguel Rocha Bento

Tese para obtenção do Grau de Doutor em
Engenharia Eletrotécnica e de Computadores
(3º ciclo de estudos)

Orientador: Prof. Sílvio José Pinto Simões Mariano

Março de 2026

Provas públicas realizadas em 24 de fevereiro de 2026

Composição do júri

Presidente

Doutor António João Marques Cardoso
Professor Catedrático da Universidade da Beira Interior

Vogais

Doutor Pedro Manuel Santos de Carvalho
Professor Catedrático do Instituto Superior Técnico da Universidade de Lisboa

Doutor Sílvio José Pinto Simões Mariano
Professor Catedrático da Universidade da Beira Interior

Doutor João Francisco Alves Martins
Professor Catedrático da Universidade Nova de Lisboa

Doutor Miguel Filipe Narciso Miguens Louro
Especialista da E-REDES – Distribuição de Eletricidade, SA

Declaração de Integridade

Eu, Pedro Miguel Rocha Bento, que abaixo assino, estudante com o número de inscrição D2263 do 3º ciclo de estudos em Engenharia Eletrotécnica e de Computadores da Faculdade de Engenharia, declaro ter desenvolvido o presente trabalho e elaborado o presente texto em total consonância com o **Código de Integridades da Universidade da Beira Interior**.

Mais concretamente afirmo não ter incorrido em qualquer das variedades de Fraude Académica, e que aqui declaro conhecer, que em particular atendi à exigida referenciação de frases, extratos, imagens e outras formas de trabalho intelectual, e assumindo assim na íntegra as responsabilidades da autoria.

Universidade da Beira Interior, Covilhã 03 /03 /2026

Assinado por: **Pedro Miguel Rocha Bento**
Num. de Identificação: 14618076
Data: 2026.03.03 17:19:07+00'00'
Localização: Covilhã

(assinatura conforme Cartão de Cidadão ou preferencialmente

assinatura digital no documento original se naquele mesmo formato)

This work was supported by the Portuguese Foundation for Science and Technology (FCT) under grant number SFRH/BD/140371/2018, co-funded by the Ministry of Education, Science and Innovation (MECI), by the European Social Fund (FSE) through the 'Programa Operacional Regional Centro' (POR-Centro), and by the Human Capital Operational Programme (POCH).

All research activities associated with this work took place in the Power Systems - Cv group of the Institute of Telecommunications (University of Beira Interior).

Dedication

To my parents and grandparents.

(Aos meus pais e avós.)

Acknowledgements

Overture: By looking at the etymology of the word “Professor”, we get the definition ”one who teaches a branch of knowledge, especially in a university”. Dating back to the late 14th century, this definition derives from the French noun “Professeur”. Personally, I would alter it to: “one who knows how to sensibly teach a branch of knowledge to different types of audiences”.

Having reflected on the meaning of such an ancient and noble profession, in my experience, nobody embodies this definition better than my supervisor, Full Professor Sílvio Mariano at the University of Beira Interior. So, first and foremost, I would like to express to him my outmost gratitude for accompanying me throughout my academic journey, for the encouragement and patience, and most importantly, for being a source of wisdom. Looking back on these years, I am certain that all the conversations (including those on unrelated subjects), the different perspectives regarding “our cherished” Electric Power Systems, the agreements, and the occasional critiques were all vital for what has been a very fruitful relationship. A collaboration whose original goals have been towered, thus, allowing me to expand the research boundaries of my thesis.

Second, to Assistant Professor José Pombo, for being a “compass” in the course of my thesis. Thank you for all the support, the prolific discussions, shared knowledge, on an academic level but also regarding different programming and formulation subjects. An excursion that started during my bachelor’s degree project, and it gradually progressed into endless hours of research guidance. Professor José Pombo is an extremely valuable asset for any academic institution, establishing an important bridge between investigation and teaching.

Third, to Associate Professor w/Habilitation Maria do Rosário, someone who, like me, shares a great interest in the theoretical (physics-related) aspects of Electrical Engineering, and to whom I would like to thank the expertise, rectitude, and promptness. A conduct that was always present during our research/project collaborations.

Fourth, my thesis would not have been possible without the comradery and the support from all my lab colleagues and co-authors. Sharing with me a vast assortment of skills and knowledge. A special word for those who became my friends, namely Hugo Nunes, João Faria, Rui Mendes, João Fermeiro and Bianca Magalhães. I am sure I will treasure the memorable moments we have shared together.

On a side note, I would like to express my gratitude to the University of Beira Interior, Instituto de Telecomunicações (in the person of Doctor Rui Paulo) and Fundação para a Ciência e Tecnologia for directly and indirectly funding my research and providing the

conditions conducive to the successful realization of my thesis.

Last but not least, a word to my parents and family members whose unwavering support has been crucial during my academic adventure!

Shibboleth: “Tell me and I forget. Teach me and I remember. Involve me and I learn.” – Benjamin Franklin (Polymath)

Tenet: “I stand upon my desk to remind myself that we should constantly look at things in a different way. See, the world looks very different from up here… Just when you think you know something, you have to look at it in another way. Even if it may seem silly, or wrong, you must try.” - John Keating (Fictional English Professor)

Peroration: ”A hero can be anyone. Even a man doing something as simple and reassuring as putting a coat around a young boy’s shoulders to let him know that the world hadn’t ended.” - Bruce Wayne (Fictional Philanthropist)

Abstract

Electrical Power Systems recognized as vast and complex entities, are the true ”engine” of contemporary societies. Currently, social and environmental concerns dominate the political agenda, notably highlighted by the Paris Agreement (2015), the EU’s Energy Roadmap 2050, and the UN’s 2030 Agenda for Sustainable Development. These initiatives stem from the detrimental impacts of persistent greenhouse gas emissions, requiring a significant contribution from these entities.

This scenario foreshadows a substantial increase in renewable energy penetration, leading to increasingly decentralized and variable generation. Consequently, the Iberian Power Systems, as a leading adopter, faces additional and considerable transformations, requiring a reevaluation of otherwise outdated predicates and in more micro level, decision-making processes. Key areas demanding special attention include the future energy policy, which can only be effectively shaped after reviewing the evolution of the power generation structure within a market-integrated context, as well as the multitude of related options for providing system flexibility—a key enabler of this profound change.

Moreover, a prevailing trend across nearly every field is the increasing reliance on data proliferation, machine learning models, and advanced optimization tools to address challenges that were once considered highly complex and less accessible. This is particularly evident in the context of electrical power systems design and operation, where such tools are now being applied to solve micro-level tasks with significant accuracy. As a result, to effectively integrate these tools, it is essential to first delve into the theoretical bedrock and practical intricacies — including the selection and tuning of hyperparameters, model architectures, and optimization strategies. This knowledge has become indispensable for modern power systems engineers, enabling them to better understand and tackle the micro-level challenges that often stem from broader energy policy decisions.

In this context, the optimal power flow and short-term price forecasting problems — both of vital importance to electricity market participants — will serve as experimental test beds to rigorously evaluate and demonstrate the application of these data-driven methodologies.

Keywords

Electrical power systems, iberian power systems, energy policy, large-scale renewable penetration, flexibility in power systems, hybrid energy systems, optimization, neural networks, metaheuristics, load forecasting, electricity price forecasting, ocean wave power forecasting, optimal power flow.

Resumo alargado

Ao longo da história, os Sistemas de Energia Elétrica foram alvo de constantes e profundas transformações. Em meados da segunda década do século XXI, estima-se que a produção mundial de energia elétrica tenha rondado os 25.000 TWh. Este é um número impressionante, sobretudo se considerarmos que a primeira central elétrica, em Pearl Street (Nova Iorque), foi inaugurada em 1882 com a capacidade de alimentar apenas 400 lâmpadas.
Dada a severidade dos impactos da contínua emissão de gases com efeito de estufa, torna-se crucial acelerar as transformações na operação dos SEE. Particular relevo é dado à incorporação de fontes de energia renováveis variáveis em larga escala. Dentro do espectro de tecnologias de geração renovável consolidadas, observa-se e prevê-se uma predominância da produção Eólica e Solar. Para este cenário, contribuíram significativamente políticas "agressivas" que mitigaram os elevados custos de investimento e de integração. No contexto português, estas políticas materializaram-se numa forte aposta na produção em regime especial, no fomento da microgeração e, mais recentemente, na legislação do autoconsumo.
No entanto, a incerteza e variabilidade associadas a estes recursos (dependentes de fatores meteorológicos e sazonalidades) contrastam fortemente com a previsibilidade da geração baseada num "mix" tradicional de tecnologias. Consequentemente, os sistemas existentes enfrentam um conjunto de desafios, alguns já identificados e outros emergentes, que exigem reflexões holísticas e soluções mais pragmáticas que priorizem a questão da segurança energética.
Desta forma, e em consonância com o título delineado para esta tese, a primeira parte deste trabalho contempla a análise e evolução integrada dos sistemas de energia na Península Ibérica. O foco recai numa análise crítica e fundamentada da política energética conduzida até então, partindo da evolução do sistema e propondo soluções adequadas de flexibilidade. Uma flexibilidade adaptada às condições atuais e futuras, que pode e deve ser vista como o grande motor para a integração massiva de energias renováveis, mitigando os custos de integração e os problemas operacionais adjacentes.
Como complemento, e com uma maior substância teórica e de implementação de *software*, uma segunda grande linha de desenvolvimento deste trabalho dará continuidade ao percurso do autor no mestrado. "Bebendo" da experiência adquirida nas áreas da otimização e das redes neuronais artificiais, e dada a sua relevância como ferramentas de Data Analytics, serão desenvolvidas novas estratégias de otimização na operação do sistema. Isto inclui previsões de carga, preços e geração de origem renovável, bem como a operação e o despacho de sistemas renováveis híbridos, culminando na solução para o clássico problema de trânsito de energia.
Este contributo permite não só expandir o conjunto de ferramentas analíticas, como também proporcionar ao autor uma compreensão mais profunda de como decisões orientadas para uma maior penetração de renováveis, necessariamente de âmbito mais macro, influenciam as características das séries temporais, os desafios dos problemas de otimização e

os requisitos dos próprios modelos de apoio à decisão baseados em *machine learning*.

Palavras-chave

Sistemas de energia elétrica, sistema elétrico ibérico, política energética, integração de energias renováveis em larga escala, flexibilidade em sistemas elétricos, sistemas de energia híbridos, otimização, redes neuronais, metaheurísticas, previsão de carga, previsão do preço da eletricidade, previsão de energia das ondas, fluxo de potência ótimo.

Contents

Dedication ix

Acknowledgements xi

Abstract xiii

Resumo alargado xv

Contents xvii

List of Figures xxi

List of Tables xxix

List of Algorithms xxxi

List of Abbreviations xxxiii

1 Introduction 1

1.1 Electrical power systems: a background 1

1.1.1 Transmission line modeling 5

1.1.2 Imbalances and frequency control 7

1.1.3 Transmission in restructured and interconnected electric power systems 12

1.1.4 Electricity markets 13

1.2 Motivation and Objectives 22

1.3 Structure of the thesis 28

2 Large-scale penetration of renewables in the Iberian power system 31

2.1 Introduction 31

2.2 The power energy mix in Iberia: characterization and trends in installed capacity and annual energy mix 33

2.2.1 An insight into the thermal dispatchable capacity in Peninsular Spain: policies, regulation, and market environment 37

2.2.2 An insight into the thermal dispatchable capacity in Peninsular Portugal: policies, regulation, and market environment 40

2.3 A comprehensive strategy for the future grid: options to increase system flexibility 42

2.3.1 Energy storage systems 45

2.3.2 Interconnections 48

2.3.3 Other renewable energy sources 49

2.3.4 Demand response and self-consumption . . . 54
2.3.5 More efficient/cleaner thermal generation . . . 56
2.4 Discussion . . . 66
2.4.1 The leading problem with the current RES integration approach . . . 66
2.4.2 Challenges posed by electrification . . . 69
2.4.3 Lessons and policy insights . . . 71

3 Soaring electricity prices in the day-ahead Iberian market: policy insights, regulatory challenges and lack of system flexibility 75
3.1 Energy policy insights: a catalyst for large-scale renewable adoption and its trade-offs . . . 75
3.1.1 Spiking electricity prices in Iberia: events and timeline . . . 76
3.1.2 Study Methodology . . . 77
3.2 European regulatory environment: an overview of policies and their implications for MIBEL . . . 78
3.3 Evolution of Iberian power system: the case of the day-ahead electricity prices in MIBEL . . . 83
3.4 Discussion . . . 90

4 Optimization 95
4.1 Background . . . 95
4.2 Classification of Optimization Algorithms . . . 97
4.2.1 Multivariate optimization conditions and critical points . . . 102
4.2.2 Metaheuristics . . . 104
4.3 Constrained problems and the penalty method . . . 107
4.4 Multi-objective problems . . . 110
4.5 From metaheuristics to surrogate-assisted evolutionary algorithms . . . 113
4.5.1 Surrogate Models . . . 115
4.5.2 Slime Mould Algorithm . . . 118
4.5.3 Proposed Methodology . . . 120
4.5.4 Case Study and Benchmarking . . . 126

5 Exploring Machine Learning: a deep-dive into artificial neural networks 135
5.1 Introduction . . . 135
5.2 Types of learning . . . 137
5.3 Artificial Neural Network: a prominent supervised learning model . . . 140
5.3.1 Forward Propagation (Feedfoward processing) . . . 142
5.3.2 Training and backpropagation . . . 147
5.3.3 Architectures . . . 171
5.4 Other ML models/architectures . . . 177
5.4.1 Deep Belief Neural Network . . . 177
5.4.2 Adaptive Neuro-Fuzzy Inference System . . . 178

5.4.3 Support Vector Machine . . . 178
5.4.4 Random Forest . . . 179
5.5 A qualitative assessment of ML models . . . 180

6 Power Flow Studies **183**
6.1 Introduction . . . 183
6.2 Fundamentals . . . 187
6.2.1 Types of Buses and Transmission System Devices . . . 187
6.2.2 Power flow equations and the admittance matrix . . . 194
6.3 The Optimal Power Flow Problem . . . 197
6.3.1 Objective Function(s) . . . 199
6.3.2 Equality Constraints . . . 201
6.3.3 Inequality Constraints . . . 203
6.3.4 Iterative Methods for the power flow solution . . . 204
6.3.5 Penalty Method . . . 212
6.4 OPF: A Test Case . . . 213

7 Forecasting in the Electric Power Systems **217**
7.1 Introduction . . . 217
7.2 Load Forecasting . . . 218
7.2.1 Theoretical Background . . . 219
7.2.2 Proposed Methodology . . . 223
7.2.3 Case Study: New England Load . . . 227
7.3 Electricity Price Forecasting . . . 232
7.3.1 Case Study I: Iberian electricity market . . . 235
7.3.2 Case study II: Daily operational optimization of a hybrid energy system using an STPF-based approach . . . 242
7.4 Generation Forecasting . . . 250
7.4.1 Ocean Wave Forecast: A state-of-the-art . . . 251
7.4.2 Wave power and sea state parameters . . . 252
7.4.3 Methodology: theory and implementation . . . 255
7.4.4 Short-term Ocean Wave Forecasting . . . 258

8 Conclusions **263**
8.1 Why the starting point matters . . . 263
8.2 Summary of Key Findings . . . 265
8.3 Future Works . . . 267

A SASMA a novel surrogate assisted metaheuristic **333**
A.1 Case Study: an in-depth look at mathematical test functions . . . 334
A.2 Optimizations Results for 100D . . . 336

B A modified Harris Hawks Optimization (mHHO) 341
B.1 Harris Hawks Optimization (HHO) . . . 341
B.1.1 From Exploration to Exploitation . . . 341
B.1.2 Exploration stage . . . 342
B.1.3 Exploitation stage . . . 342
B.1.4 Modifications . . . 343
B.1.5 mHHO: The Pseudo-code . . . 345

C Proposed Methodology for STPF: Pseudo-code 347

D Using STLF as an exogenous input for STPF: a case study on the MIBEL data 349
D.1 Intermediate STLF in Peninsular Spain and Portugal: Methodology . . . 349
D.2 Intermediate STLF in Peninsular Spain and Portugal: Results . . . 349

E HES: proposed operational strategy, data analysis and modeling 353
E.1 Proposed operation strategy . . . 353
E.2 A brief statistical analysis of the datasets . . . 357
E.3 Hybrid system modeling . . . 358
E.3.1 Kinetic Battery Model . . . 358
E.3.2 Pumped-Storage Hydropower . . . 359
E.3.3 Wind Power . . . 361
E.3.4 Solar Photovoltaic . . . 362
E.3.5 Diesel Generator . . . 363
E.3.6 Power Electronic Converters . . . 364
E.4 A weekly results analys: costs, dispatch and STPF . . . 365

List of Figures

1.1 Typical EPS architecture (overview by voltage levels): From generation to consumption. 2
1.2 Main components of a overhead double-circuit transmission line with two cross-arms. Adapted from Ron Clausen [Photograph], CC0, via Wikimedia Commons. 6
1.3 Single-phase nominal π(Pi) model of a transmission line. 7
1.4 Typical EPS frequency response during a under-frequency contingency: control time-frame. 9
1.5 Frequency coordination: conceptual view of EPS frequency constraints and main response mechanisms. 11
1.6 Structure (Hierarchy) of most European Wholesale Electricity Markets. . . 15
1.7 Cumulative (aggregated) supply and demand curves derived from consumer bids (magenta) and generator offers (blue), illustrating the typical piecewise linear clearing method used in the day-ahead market for a specific hour. 16
1.8 Overview of the operational time-frame and settlement type for the different wholesale electricity markets in the EU (adapted). 20
1.9 Lock-ins in large-scale penetration of renewable energies (adapted). 24
1.10 Transformation in the Electric Power Systems: Thesis scope. 27

2.1 Peninsular Spain: Evolution of the installed capacity in GW. Bar plot of the main generation technologies share (y-right) and the $renewablescap_{ratio}$ (dashed pink y-left). 34
2.2 Peninsular Spain: Annual generation mix share by technology: 2016 vs 2021 (comparison). 35
2.3 Peninsular Portugal: Evolution of the installed capacity in GW. Bar plot of the main generation technologies share (y-right) and the $renewablescap_{ratio}$ (dashed pink y-left). 36
2.4 Peninsular Portugal: Annual generation mix share by technology: 2016 vs 2021 (comparison). 36
2.5 Main vectors of action to increase power system flexibility. 43
2.6 Most adequate energy storage technologies for VRE integration support. . 46
2.7 Effect of large-scale penetration of ESSs: breakdown of electricity generation. 46
2.8 Interconnections in Iberia: installed and planned locations, voltage levels and total capacity. 50
2.9 Promising RES to complement the existing generation portfolio in Iberia. . 50
2.10 Classification: Main DR mechanisms. 55
2.11 Main modern solutions (technologies) for a “greener” and more efficient thermal generation. 57

2.12 Evolution of coal power plants technology. 60
2.13 Evolution of nuclear reactor technologies. 62
2.14 From waste heat energy sources to electricity. 64
2.15 CSS modalities for thermal power plants. 65
2.16 Wind and PV generation (hourly) power capacity factors by seasons of the year (2017-2021). 67
2.17 Key differences between synchronous generation and inverter-based resources. 69
2.18 Electrification impacts on the system load (scenarios): Current ”duck” load profile (lilac); Uncontrolled electrification with sizable peak load increase (maroon); Electrification plus system flexibility, especially EV-based (mint); PV self-consumption downward effect (olive). 71

3.1 Methodology: literature/documentation review and market/grid operator data analysis (downward); extract result inferences and policy insights (upward). 77
3.2 European Power Systems Transformation: threats, actors and drivers of change, policies and operational challenges. 80
3.3 Annual Electricity Demand in TWh: the Peninsular Spanish system (yellow) and the Peninsular Portuguese system (red). 83
3.4 Annual Net Generation Mix – Share (%) of Renewables (green bar) and Non-Renewables (dark grey bar) [left axis] vs Cumulative Annual Generation (TWh) of Renewables (lime green dashed line) and Non-Renewables (grey dashed line) [right axis]: data for the Peninsular Spanish and Portuguese systems, respectively. 84
3.5 Evolution of the generation installed (available) capacity (GW) in the Iberian Peninsula: Nuclear (purple bar), Non-Renewables (dark grey bar), Renewables (green bar). 85
3.6 Annual Energy Trade in TWh: Peninsular Spain Imports (purple bar) and Exports (light-blue bar) versus Peninsular Portugal Imports (violet bar) and Exports (green bar). 86
3.7 Wholesale day-ahead (daily) average price in MIBEL (sky-blue) from 2018 to 2021 (Spain prices in the market-spliting instances) and the linear fitting trends (pink). 88
3.8 Relative frequency analysis of buying and selling day-ahead electricity market (hourly) orders, together with the absolute frequency, above the prices classes defined by the new price ruling: Month prior (07/06/2021-06/07/2021) in blue vs Month after (07/07/2021-05/08/2021) in magenta. 89
3.9 The evolution of the “energy security” concept in the context of the European power system’s transformation. 91

4.1 Optimization functions: a) convex function; b) non-convex and non-concave function. 98
4.2 Hierarchical classification of optimization algorithms. 101
4.3 Classification of metaheuristic algorithms. 106
4.4 Quadratic penalty term value for several penalty parameters. 109
4.5 Pareto Front (green curve) in a two-objective space, Non-Dominated/Pareto-optimal solutions (green circles), Dominated solutions (white circles), and Infeasible region (delimited in blue). 113
4.6 Online surrogate model working scheme (generic). 116
4.7 Surrogate Model: (rough) smoothing effect. 116
4.8 RBFNN: a default forward ML architecture. 117
4.9 Slime mould foraging mechanism. 118
4.10 SMA: balance between different space search mechanisms. 119
4.11 Proposed Surrogate Assisted SMA algorithm (SASMA): Overview of the flow between the main mechanisms. 121
4.12 Detailed flowchart of the global database management: DB update and integrity check mechanisms. 123
4.13 Selected subspace for surrogate modeling (spread coefficient effect). 126
4.14 Convergence curves (fitness value): CALSAPSO (dashed orange), GORS-SSLPSO (dashed red), SAE-ARFS (dashed blue), SHPSO (dashed cyan), TL-SSLPSO (dashed brown), TLSAPSO (dashed purple) and SASMA (solid pink) on 30D unimodal test functions (F1-F4,F6,F7 and F14). 132
4.15 Convergence curves (fitness value): CALSAPSO (dashed orange), GORS-SSLPSO (dashed red), SAE-ARFS (dashed blue), SHPSO (dashed cyan), TL-SSLPSO (dashed brown), TLSAPSO (dashed purple) and SASMA (solid pink) on 30D multimodal test functions (F5,F8 and F10-F13). 132
4.16 Convergence curves (fitness value): CALSAPSO (dashed orange), GORS-SSLPSO (dashed red), SAE-ARFS (dashed blue), SHPSO (dashed cyan), TL-SSLPSO (dashed brown), TLSAPSO (dashed purple) and SASMA (solid pink) on 30D very complicated multimodal test functions (F9 and F15-F17). 133
4.17 Relative time comparison between SASMA and the other SAEAs for 30D and 100D. 134
4.18 SASMA Database: Result Analysis for 30D (first line) and 100D (second line). 134

5.1 Deep learning in the artificial intelligence and big data landscape (adapted). 136
5.2 ML traditional stages with the manual features extraction stage accentuated in a dashed line to signal its absence in DL models (adapted). 137
5.3 Supervised learning: regression vs classification. 139
5.4 Working mechanism of ANNs: a) basic structure of a biological neuron; b) ANN layered structure. 141
5.5 Modern perceptron representation. 143

5.6 Activation functions as the primary (and for most of the architectures the sole) source of non-linearity in ANNs: a key enabler of their ability to capture and model complex data patterns (highlighted in a chain of key ANN characteristics/traits). 145
5.7 Common activation functions: Sigmoid, Tanh and ReLU considering $\alpha = 1.5$.146
5.8 Forward and backward propagation in a generic 2-hidden-layer, single-node-output (layer) ANN: an iterative training process. 148
5.9 Gradient descent for training loss minimization illustration: weight update and convergence. 149
5.10 ANN stochastic gradient descent training: epoch, iterations and batch size. 151
5.11 Gradient descent behavior for the different batch size variants. 152
5.12 Effect of different learning rate parameters η (adapted): Loss function as a function of a given weight and Loss function as function of the epochs. . . 153
5.13 Bias-Variance trade-off illustration: underfitting, balanced fitting and overfitting. 161
5.14 Training and validation convergence curves. The concept of early stopping for ANN training. 163
5.15 K-fold cross validation: data partition procedure. 166
5.16 Overfitting mitigation/prevention options in ANN context. 168
5.17 Commonly used Feed Forward type architectures: Generic overview of connections and nodes. 172
5.18 Commonly used recurrent type architectures: Generic overview of connections and nodes. 175
5.19 CNN architecture: Generic overview of connections and nodes. 177
5.20 DBNN architecture: Generic overview of connections and nodes. 178
5.21 ANFIS architecture: Generic overview of connections and nodes. 179
5.22 RF architecture: Generic overview of connections and nodes. 180

6.1 OPF literature: conventional vs trendy approaches. 186
6.2 Types of Buses: name and most common representations. 188
6.3 Tap-changing transformer: a) EPS branch representation; b) simplified π-type equivalent circuit. 192
6.4 Nodal current injections from interconnected buses (power system illustration). 195
6.5 OPF objective function value vs number of control actions. 199
6.6 Valve point loading effect. 200
6.7 IEEE 30-bus system (test case). 213
6.8 Convergence curve limits and mean value: HHO vs mHHO. 215

7.1 Deep Forward Neural Network (DFNN). 223
7.2 Proposed methodology: overview. 224

7.3 Predictions models: outline of the different stages from data selection to ARIMA Forecasters. 226
7.4 New England region (US) — Electrical Load (MW) 2016. 228
7.5 Load forecasts vs. real load (New England region), where the ARIMA forecasters area in yellow represents the region mapped by the forecasts of 15 different ARIMA models, the stacking ensemble forecast is illustrated in a dashed magenta line and the real (target) load is illustrated in a solid blue line. 230
7.6 Box-plots of the MAPE (%) and RMSE (MW) error distributions (NN without exogenous inputs-blue, NN with exogenous inputs-green and the proposed ensemble methodology—magenta over the course of 50 runs). . . . 232
7.7 Predicted loads vs. real load and the respective mean deviation (signed error) of each method: 1st week of April 2016. 233
7.8 Predicted loads vs. real load and the respective mean deviation (signed error) of each method: 3rd week of April 2016. 234
7.9 MIBEL Day-Ahead Electricity Prices Time-Series (2016-2019) and Correlation Analysis. 236
7.10 Proposed Methodology: An overview. 239
7.11 Tree of possible permutations between the selected exogenous variables, symbolized by its respective index. 240
7.12 Day-ahead price forecast [€/MWh] (MIBEL) for the month of February 2019: Real Price (in violet) vs Forecasted Price (in teal green). 241
7.13 Day-ahead price forecast [€/MWh] (MIBEL) for the month of August 2019: Real Price (in violet) vs Forecasted Price (in teal green). 243
7.14 Schematic representation of the proposed hybrid system architecture. . . . 245
7.15 Illustration of the employed price forecasting: FFNN input data selection mechanism. 246
7.16 Day-ahead Electricity Prices and Prosumer's Load Profile. 247
7.17 Weather Data: Wind Speed, Solar Irradiance and Temperature. 248
7.18 Energy generation share and load levels for each week. Where yellow, green and blue represent the PV, wind and hydro generation, respectively, light blue represents pumping periods, brown buying periods and orange selling periods. In turn, purple represents discharge periods, and lilac periods of charging the battery bank. 250
7.19 Kernel joint probability density estimation of the collected observations. The colour scale represents the wave energy per metre of wave front (in MWh/m). The numbers within the graphs indicate the occurrence of sea states (in number of hours per year) and the isolines refer to the wave power J (in kW). (a) Buoy ID: 41048, (b) Buoy ID: 42056, (c) Buoy ID: 44014. . . 254
7.20 Convolutional Neural Network (CNN or ConvNet) architecture. 255
7.21 Wave pre-processing and input data selection. 257

7.22 Proposed Methodology (Overview): CNN architecture fed with several 1D wave input data. 258
7.23 Box plots of the error metrics distributions, considering all forecasting horizons (buoy 41048). 261
7.24 Test Results illustration for all horizons: Actual and forecasted wave power flux, J, respectively, blue solid line and maroon dashed line (left axis) and absolute error, canary yellow dot-dashed line (scaled right axis) ample graph with blue (dotted), green (solid) and red (dashed) lines (a) Buoy 41048 Spring Week, (b) Buoy 42056 Fall Week. 262

A.1 Convergence curves (fitness value): GORS-SSLPSO (dashed red), SAE-ARFS (dashed blue), SHPSO (dashed cyan), TL-SSLPSO (dashed brown), TL-SAPSO (dashed purple) and SASMA (solid pink) on 100D unimodal test functions (F1-F4,F6,F7 and F14). 338
A.2 Convergence curves (fitness value): GORS-SSLPSO (dashed red), SAE-ARFS (dashed blue), SHPSO (dashed cyan), TL-SSLPSO (dashed brown), TL-SAPSO (dashed purple) and SASMA (solid pink) on 100D multimodal test functions (F5,F8 and F10-F13). 338
A.3 Convergence curves (fitness value): GORS-SSLPSO (dashed red), SAE-ARFS (dashed blue), SHPSO (dashed cyan), TL-SSLPSO (dashed brown), TL-SAPSO (dashed purple) and SASMA (solid pink) on 100D very complicated multimodal test functions (F9 and F15-F17). 339

B.1 $E_{escprey}$ modified behavior illustration. 344

D.1 Day-ahead load forecast [MW] of the Portuguese System for the month of February 2019: Real Load (in violet) vs Forecasted Load (in teal green). . . 350
D.2 Day-ahead load forecast [MW] of the Spanish System Load for the month of October 2019: Real Load (in violet) vs Forecasted Load (in teal green). . 350

E.1 Equation (2) co-domain illustration: PSH and battery bank costs for four different SOCs. 357
E.2 The two-well kinetic battery model (KiBaM). 359
E.3 Head-flow ranges of small hydro turbines. 360
E.4 Wind Turbines Power Curves: BWC Excel-10 (blue line) and Norwin 24-STALL-150 (dashed red line). 361
E.5 Diesel Generator: fuel consumption curves as a function of power output. . 364
E.6 Dispatched energy mix and its dispatch costs for the summer week. Where yellow, green and blue represent the PV, wind and hydro generation, respectively, light blue represents pumping periods, brown buying periods and orange selling periods. In turn, purple represents discharge periods, and lilac periods of charging the battery bank. Black represents the existing load level. 366

E.7 Price Forecast vs. Real Prices and State of Charge evolution for the summer week. 366

List of Tables

1.1 Voltage classes: defined ranges as per ANSI C84.1 standard. 3

2.1 Coal-fired generation: Gross installed capacity in Peninsular Spain (available in 2022). 38
2.2 Nuclear power generation: gross installed capacity in Peninsular Spain (active in 2022). 40
2.3 Main CCGT thermal generation: gross installed capacity in Peninsular Portugal (available in 2022). 41
2.4 Estimated unweighted LCOE and LCOS energy resources entering service in 2027. 44

3.1 Bivariate correlation of related variables with the day-ahead electricity prices time-series. 87

4.1 Parameter settings for the benchmark algorithms and the proposed SASMA algorithm. 127
4.2 Case Study: List of unimodal and multimodal test functions. 128
4.3 Statistical results of the proposed SASMA and the benchmark SAEAs on 30D mathematical test functions. 130
4.4 Mean rank scores (based on the Friedman Test) of the SASMA and the benchmark SAEAs on 30D mathematical test functions. 131

6.1 Power System Bus Types: Classification 188
6.2 IEEE 30-Bus System: Main Parameters and Fuel Cost Constants. 214
6.3 Control variables bounds, optimized values and total cost for the best run. 216

7.1 One day-ahead NE Load Forecasting results: MAPE (%). 229
7.2 One day-ahead NE Load Forecasting results: RMSE (MW). 230
7.3 Comparison of one day-ahead NE Load Forecasting results: MAPE (%). . . 231
7.4 Comparison of one day-ahead NE Load Forecasting results: RMSE (MW). 231
7.5 One-day ahead electricity MIBEL price forecast (random 50 days – 2018). 236
7.6 Exogenous variables: statistical measures. 238
7.7 Pearson correlation coefficients between the electricity prices and the selected exogenous variables. 238
7.8 Grouped (list) of exogenous variables. 239
7.9 STPF of the MIBEL day-ahead electricity prices: error metrics analysis. . . 242
7.10 HES Simulation: Main results. 249
7.11 Test weeks statistics and auxiliary STPF forecasting accuracy. 249
7.12 Statistics of the different wave power datasets [$\mathrm{kW\,m^{-1}}$] 254
7.13 Forecasting: Testing Dates. 259

7.14 Forecasting errors: buoy 41048 (mean results). 259
7.15 Forecasting errors: buoy 42056 (mean results). 260
7.16 Forecasting errors: buoy 44014 (mean results). 260

A.1 Case Study I: Benchmark test functions. 335
A.2 Mean rank scores (based on the Friedman Test) of the SASMA and the benchmark SAEAs on 100D mathematical test functions. 336
A.3 Statistical results of the proposed SASMA and the benchmark SAEAs on 100D mathematical test functions. 337

D.1 STLF of the Spanish and Portuguese System Load : Error Metrics Analysis. 350

E.1 Descriptive statistics of the different datasets. 357
E.2 KiBaM model: battery bank specifications. 359
E.3 PSH considered specifications. 360
E.4 Wind Turbine Modeling: Bergey and Norwin. 362
E.5 Specifications of the PV solar module. 363

List of Algorithms

1 Pseudo Code of the Surrogate Assisted Slime Mould Algorithm (SASMA) and the employed parameters 333
2 Pseudo-code of mHHO algorithm 346
3 Ensembled DFNN-based Short-Term Load Forecasting (STLF) Procedure 347
4 Operational strategy. 356

List of Abbreviations

ACER	Agency for the Cooperation of Energy Regulators
ADALINE	Adaptive Linear Neuron
AE	Auto Enconder
AI	Artificial Intelligence
ANFIS	Adaptive Neuro-fuzzy Inference System
ANN	Artificial Neural Network
AVR	Automatic Voltage Regulator
BWR	Boiling Water Reactor
CC	Combined Cycle
CCGT	Combined Cycle Gas Turbine
CCS	Carbon Capture and Sequestration
CG	Conjugate Gradient
CNMC	Spanish National Markets and Competition Commission
CNN	Convolutional Neural Architecture
CPP	Coincident Peak Pricing
CPU	Central Processing Unit
CS	Cuckoo Search (Algorithm)
CSP	Concentrating Solar Power
DBNN	Deep Belief Neural Network
DE	Differential Evolution
DER	Distributed Energy Resource
DFNN	Deep Feedforward Neural Network
DGEG	Portuguese Directorate-general for Energy and Geology
DL	Deep Learning
DOE	U.S. Department of Energy
DR	Demand Response
EA	Evolutionary Algorithm
EC	European Community
ED	Economic Dispatch
EDP	Energias de Portugal
EHV	Extra High Voltage
EIA	U.S. Energy Information Administration

ELM	Extreme Learning Machine
EPR	European Pressurized Reactor
EPS	Electric Power System
ERSE	Portuguese Energy Regulator
ESNN	Echo State Neural Network
ESS	Energy Storage System
EU	European Union
EV	Electric Vehicle
FDLF	Fast Decoupled Load Flow
FE	Fitness Evaluation
FiT	Feed-in Tariff
FORATOM	European Atomic Forum
GHG	Greenhouse Gas
GPU	Graphics Processing Unit
GP	Gaussian Process
GRU	Gated Recurrent Unit
GS	Gauss-Seidel
GSA	Gravitational Search Algorithm
GWO	Gray Wolf Optimization (Algorithm)
HHO	Harris Hawks Optimization (Algorithm)
HV	High Voltage
HVAC	Heating, Ventilation, and Air Conditioning
HVDC	High-Voltage Direct Current
IAEA	International Atomic Energy Agency
IEA	International Energy Agency
IGCC	Integrated Gasification Combined Cycle
ISO	Independent System Operator
LCOE	Levelized Cost of Electricity
LCOS	Levelized Cost of Storage
LHS	Latin Hypercube Sampling
LMS	Least Mean Squares Algorithm
LSTMs	Long Short Term Memory
LV	Low Voltage
MAE	Mean Absolute Error

MCO	Market Coupling Operator
MIBEL	Iberian Electricity Market
ML	Machine Learning
MLP	Multilayer Perceptron
MRE	Mean Relative Error
MSA	Moth Swarm Algorithm
MSE	Mean Squared Error
MV	Medium Voltage
NEMO	National (or Regional) Electricity Market Operator
NR	Newton-raphson
OMIE	Spanish-controlled side of the Iberian Energy Market Operator
OMIP	Portuguese-controlled side of the Iberian Energy Market Operator
OPF	Optimal Power Flow
OTC	Over-the-counter
PCA	Principal Component Analysis
PHS	Pumped Hydroelectric Storage
PPA	Power Purchase Agreement
PQ	Load Type Bus
PSH	Pumped-storage Hydro
PSO	Particle Swarm Optimization
PTEL	Portuguese Electricity Market Zone
PV	Photovoltaics Or Generator Type Bus (context dependent)
PVPC	Voluntary Price for Small Consumers (Spanish ”regulated” market-based electricity price scheme)
PWR	Pressurized Water Reactor
RBF	Radial Basis Function
RBFNN	Radial Basis Functions Neural Network
RBM	Restricted Boltzmann Machines
REE	Spanish Electric Grid Operator
RES	Renewable Energy Source
RF	Random Forest
RMSE	Root Mean Squared Error
RNN	Recurrent Type Architecture

RTO	Regional Transmission Organization
RTP	Real-Time Pricing
SAEA	Surrogate-Assisted Evolutionary Algorithm
SASMA	Surrogate Assisted Slime Mould Algorithm
SGD	Stochastic Gradient Descent
SMA	Slime Mould Algorithm
SOM	Self-Organizing Map
SPEL	Spanish Electricity Market Zone
SRP	Special Production Regime (for RESs in Portugal)
SSR	Sum of Squared Residuals
STD	Standard Deviation
SUBCR	Subcritical Units
SUPERCR	Supercritical Plants
SVM	Support Vector Machine
SVR	Support Vector Regression
TES	Thermal Energy Storage
TSO	Transmission System Operator
UC	Unit Commitment
UHV	Ultra High Voltage
ULTRSC	Ultra-Supercritical
VEGA	Vector Evaluated Genetic Algorithm
VRE	Variable Renewable Energy
WOA	Whale Optimization Algorithm

Chapter 1

Introduction

Emerging in the late 19th century, electric power systems (EPSs) have played a pivotal role in the overall development of societies. Responsible for generating, transmitting, and distributing electrical energy to homes, businesses, and industries, their importance has only grown over the course of more than a century. They are an integral part of today's modern world, often overlooked or only discussed through major news cycle talking points.
Before delving into the challenges facing EPSs, it is crucial to lay a strong foundation by revisiting the fundamentals. This will ensure a common framework encompassing power generation, electricity markets, and load demand, with an emphasis on the intricacies and interdependencies involved. With a strong grasp of the basics, one can more easily understand how various traits and operational characteristics are reviewed, altered, maintained, or neglected in the context of system-wide restructuring, which is currently spearheading the advancement towards a carbon-free generation portfolio.
Prioritizing system reliability and affordability will provide key insights for the analytical lenses employed throughout this thesis.

1.1 Electrical power systems: a background

Electrical power systems are comprised of thousands of individual components that are dispersed through vast areas of both urban and rural landscape, connecting generation (classically synchronous generators from thermal plants), with consumption (load) centers [1]. Therefore, fulfilling, for 24 hours a day, 7 days a week, and 365 days a year, the power balance equation:

$$\sum_{i=1}^{n} p_{G_i}(t) - P_{Load}(t) = 0 \Leftrightarrow P_G(t) - P_{Load}(t) = 0 \quad (1.1)$$

translating, the most basic but also most important premise of any EPS. Where $P_G(t)$ stands for the total power generation at a certain time instant, t, i.e, the contribution of all n individual power generation outputs in an EPS, $p_{G_i}(t)$; while $P_{Load}(t)$ is the current load (also referred to as electricity demand) to serve.
The large-scale and complex nature of these interconnected systems can be decomposed into various hierarchical levels [2], particularly in terms of voltage levels at which electricity is transported [3], as is synthetically illustrated in Figure 1.1. Covering the main voltage conversion legs of the traditional electric energy journey between generation and consumption (from primary energy sources, which may be renewable or non-renewable, to secondary energy [fluxes] right after conversion and during transport, accounting for

all respective losses, and finally to the delivery of final energy for end use [4]).

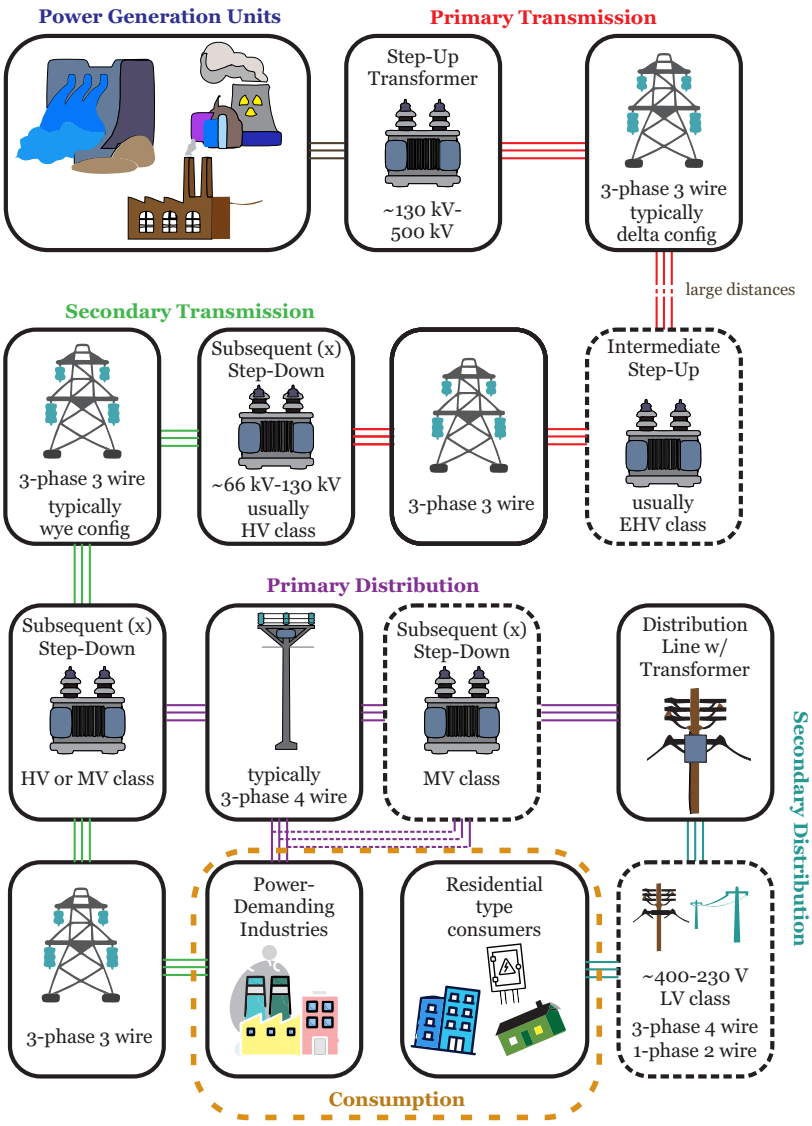


Figure 1.1: Typical EPS architecture (overview by voltage levels): From generation to consumption.

First, it is important to underscore the large range of output voltages of different power generating stations (usually comprised of a couple or even thousands of generation units). These range from a DC output as low as 400V to 1kV (with series and parallel connections of the 24 to 48V panels) for photovoltaic farms; 690V to 3.3kV or even higher AC outputs for wind farms; to 3.3kV, 6.6kV, and 11kV AC outputs for most of the thermal, nuclear and hydro power plants; some instances can go even north of 22kV (very large hydropower and nuclear installations). This disparity is explained by the different nature of the energy source (calorific value, power density, etc.), efficiency, scalability and modularity, and the "custom-made" nature of the inherent energy conversion technology (design). For instance, hydropower is well-known for their scalability (range). The world's largest hydroelectric plant, the Three Gorges Dam in China, with its installed capacity of roughly

22500MW, surpasses by a factor of 2.8 the capacity of the largest nuclear power plant, and even more so the capacity of the largest gas-fired and coal-fired thermal plants.

This ample range of generated output voltages constitutes the "starting" point of the transmission network, representing the bulk transfer of electrical energy [5]. Operating at the highest voltage levels to significantly reduce power losses, as defined by the first two classes in Table 1.1, where transmission substations with step-up transformers usually connected to EHV transmission lines (UHV transmission is still relatively scarce on a global scale) — bare conductors with large cross-sectional areas and greater conductivity, mostly overhead — establish the bridge between generation sites and the distribution network, in what is often designated as primary transmission.

Table 1.1: Voltage classes: defined ranges as per ANSI C84.1 standard [6].

Voltage Class	**Range**
Ultra High Voltage (UHV)	>1000 kV
Extra High Voltage (EHV)	>230 kV and ≤ 1000 kV
High Voltage (HV)	>100 kV and ≤ 230 kV
Medium Voltage (MV)	>1 kV and ≤ 100 kV
Low Voltage (LV)	<1 kV

With no need for what would be a costly neutral wire — as in a Wye configuration, where the extra wire provides access to two different voltage levels, such as supplying single-phase loads directly using the phase voltage — the Delta configuration stands out as a more cost-effective option, relying only on a three-wire system. Despite its simplicity, it is highly reliable, as it can continue operating even if one phase is lost. Additionally, the closed-loop nature of the Delta configuration inherently manages unbalanced currents by redistributing any imbalance across the three phases, allowing them to circulate internally without affecting external components. This meshed connection between the windings also suppresses — or more precisely, prevents — certain harmonic interferences from propagating in the power network, particularly the third harmonic. These advantages make the Delta-Delta three-phase transformer connection the preferred choice for long-distance EHV transmission systems.

As it travels through the transmission lines, electrical energy experiences voltage drops due to the resistance of the conductors and other factors. To counteract these drops and maintain voltage levels within an optimal range, additional step-up transformers may be placed at intermediate substations along the primary transmission network, as is also depicted in Figure 1.1. Besides, primary transmission is also responsible for the connection between two regional or national power grids, forming an interconnection [7].

Secondary transmission or sub-transmission is a second leg of the electricity's journey at HV levels, covering shorter distances to accommodate regional load centers, where voltages are further transformed for delivery in subsequent sub-transmission substations. Typically, voltages are stepped down (possibly several times) to feed distribution substations, but not exclusively, as is the case of energy-intensive (power-demanding) industries, requiring HV levels, as depicted in Figure 1.1.

In the same manner, the distribution network can also be divided into two operational levels. Primary distribution operates at the MV range, marking the transition from transmission to distribution. It involves several voltage reductions at various distribution substations before reaching load centers through different feeders, such as those in metropolitan areas, certain industrial and commercial sites, as well as more rural or dispersed consumption spots.

In transmission systems, the absence of a neutral wire simplifies the system by eliminating the need for grounding at the highest voltage levels, as imbalances at this bulk transport level are typically small. This is why most voltage classification tables for HV, EHV, and UHV do not specify 4-wire systems. However, when connecting transmission systems to lower voltage distribution networks, a Wye connection on the secondary side is often paired with a Delta primary (Delta-Wye configuration). This configuration is commonly used for step-down transformers interfacing with high voltage networks, providing MV output. The Wye connection offers several advantages: it provides a neutral point, allowing access to different voltage magnitudes; it serves as a grounding reference for stability and safety, essential for handling fault currents; and it reduces insulation requirements, as the phase-to-neutral voltage is lower than the phase-to-phase voltage. When the grid is already configured to a four-wire setup, a Wye-Wye configuration is then employed for sub-sequential transformation needs. As such, the last leg of this journey concerns secondary distribution, as depicted in Figure 1.1, which is typically defined by the point where a grid connection is established for LV ranges and its respective consumption, such as in residential areas.

System feeders can be arranged in radial, parallel, ring or mesh topologies [8]. The first is the simplest and cheaper option, which is commonly used in more rural/suburban areas at a LV level. While the last one, making use of several ring and/or parallel topologies, is more common at the transmission level. Given that it provides added degrees of redundancy, for a much greater load (and potentially impacted customers in the case of failure).

Nevertheless, the traditional and more stratified EPS architecture, as represented in Figure 1.1, where large-scale generation plants feed into high voltage (HV) transmission networks, transporting power to medium voltage (MV) distribution networks and on to low voltage customer levels [9], is already being moderately challenged by persistent developments in sector coupling and the proliferation of distributed energy resources (DERs), which may or may not use inverter-based resources to interface with the AC grid. It is important to recognize that these resources are not exclusively generation-related, as commonly assumed. They also include storage systems, connections with high voltage direct current systems, such as transmission lines in the EHV and UHV classes (Table 1.1), and the interface with static synchronous compensator and static volt-ampere reactive compensators [10].

In other words, the grid architecture is constantly evolving and adapting to accommodate the large-scale penetration of variable renewable energies, adding to the existing EPS complexity [11]. Specifically, the wide integration of technologies such as solar photo-

voltaic and energy storage solutions at both utility (connected to the bulk power system, usually at sub-transmission voltage levels) and residential (secondary distribution) levels, as well as the foreseeable challenging electric vehicle loads at the medium to lower end of the voltage spectrum [12], does not benefit from starting with a ”blank canvas” for optimal grid design, as these networks were not originally designed for distributed generation. As many of the existing traditional large-scale power stations did (in the sense of a more long-term planning with the dedicated transport infrastructure and carefully conducted power flow studies). Instead, renewable technologies must adapt to the current grid architecture and take into account its rigidity, including the availability and sizing of coupling points. This directly ties into operational issues like voltage and frequency stability. Effectively representing a paradigm shift from the described (old) centralized view towards a more distributed one, creating alternatives/opportunities in terms of energy, firm capacity and operating reserves services, and it can even defer some transmission or distribution network investments which otherwise would be needed, just to accommodate the expected loads (not only in terms of peak demand but more so in terms of how it may re-shape the daily load diagram, as we will explore) [13].

A daring transition characterized by increased interconnectedness, where generation becomes more atomized (more numerous and with relatively smaller capacities), situated closer to or even within the load centers themselves. Additionally, end-users shift from a passive role to a more active one [14], bringing forth a variety of challenges and decisions for energy players and policymakers to consider and study, as will be explored in depth in later chapters.

1.1.1 Transmission line modeling

Transmission lines are a major component of the EPS infrastructure. In terms of the tower itself — the conventional way of illustrating the transport of electric energy — there are several possible configurations, varying in the number circuits, cross-arms, frame disposition, etc. For instance, in Figure 1.2, a fairly common double-circuit transmission line with two cross-arms is shown.

Power losses, P_{Losses}, are a major issue in the transport of energy, accounting for approximately 4-15% of total generation when aggregating transmission, distribution, and transformer losses [15]. Generation sources must also cover transmission losses to ensure that demand is met. Taking this factor into account, the power balance equation 1.1 can be rewritten as shown in equation 1.2. This additional factor incurs an associated cost, which is directly reflected in end-user tariff prices [16], as per regulatory decisions.

$$P_G(t) - (P_{Load}(t) + P_{Losses}(t)) = 0. \quad (1.2)$$

Essentially, there are two main types of transport losses: technical and non-technical. The first type refers to inherent (inevitable) losses, which primarily occur due to energy

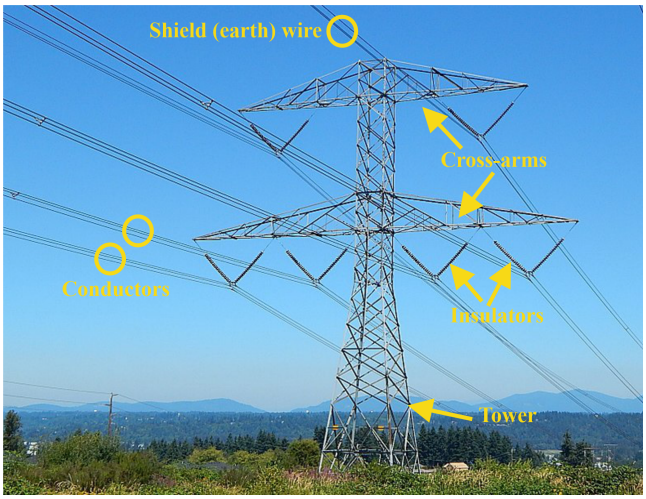


Figure 1.2: Main components of a overhead double-circuit transmission line with two cross-arms. Adapted from Ron Clausen [Photograph], CC0, via Wikimedia Commons.

dissipation phenomena. These losses can be categorized as either variable or fixed (no-load) losses. In contrast, the second type pertains to already consumed but unpaid electric energy. Unlike the first type, these losses are avoidable. Common examples include wrong consumption estimates, metering and billing errors, faulty and illegal connections [17].

Due to the conductor's internal resistance, (technical) variable losses are the primary source of losses associated with the transmission of electric energy, [18]. Joule's heating, given by equation 1.3 is dependent on the (changing) power flow transversing the conductors, that comprise the grid transport infrastructure, e.g., copper wires in the overhead lines or coils in transformers windings.

$$P_{Joule} = (I_{\text{rms}})^2 \operatorname{Re}(Z)\,. \quad (1.3)$$

Where I_{rms} is the root mean square (equivalent DC) current, that dissipates the same power in a purely resistive load, represented by $\operatorname{Re}(Z) = R$. A subsequent negative effect of this heat loss is the increase in the resistivity of the conductor, a variable that is directly proportional to its temperature.

As mentioned earlier, this type of loss is mitigated via the increasing HV spectrum used in transmission lines, requiring better electrical insulation. The same explains why data from 35 European countries show losses as a percentage of injected energy in transmission, varying between 0.5% and 3%, while in distribution, they range from 2% to 14%. Together with the inherent losses due to the higher currents and reduced transformer efficiencies, at the distribution level, the fact that virtually all the non-technical losses also take place at this level, explains the differing power loss percentage ranges [19]. Capacitive losses occur between the ground and the power lines (conductors) due to the electric field created, effectively storing a tiny fraction of the energy that cannot be fully transmitted. This is one of the reasons, beyond safety, why power lines are positioned high

above the ground, as capacitance is inversely proportional to the distance between conductors $\left(C = \frac{\varepsilon \cdot A}{d}\right)$. Iron losses in the transformer's core stand out as a prime example of no-load losses. Nevertheless, there are other technical losses that occur in the form of heat or noise, with a relevant impact on transmission and distribution networks, that are also worth mentioning, namely radiation losses, coupling losses, dielectric heating losses, corona losses, skin effect and Ferranti effect [20].
To model these technical losses/effects, which are mostly observed for medium to long transmission lines [21], and before moving on to power flow analysis, and understand why transmission losses are an added component of economic dispatch studies, it is essential to understand the reactance components of the transmission line models as well. There is no established single line model suitable for all situations, however, nominal π model is considered a reference equivalent circuit for average length transmission lines, that allows us to transform the distributed line parameters (resistance, inductance, and capacitance) into a set of discrete parameters [22], as shown in Figure 1.3.

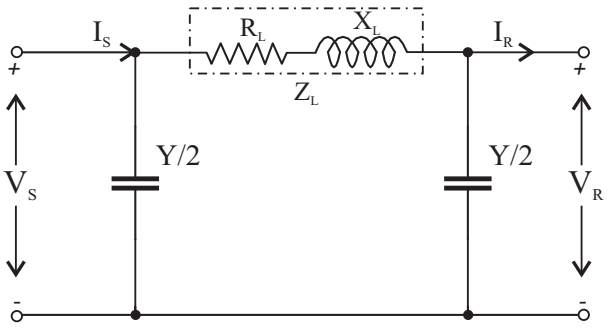


Figure 1.3: Single-phase nominal π(Pi) model of a transmission line.

This lumped-parameter model consists of a series impedance (with an inductive reactance), $Z_L = R_L + jX_L$, which models series voltage drop, concentrated at the center of the circuit; while a shunt (parallel) capacitance, is divided equally and placed at either end of the circuit, denoted by the admittance, $\frac{Y}{2} \simeq \frac{jB_C}{2}$(with neglectable shunt conductance) [23], thus accounting for line-charging currents and properly modeling the dielectric losses. So, by simplifying the transmission line into a circuit with three discrete and concentrated parameters—a series impedance and shunt admittance—the π line model provides a straightforward way to represent what is otherwise a continuously distributed system along the entire line length. This simplification enables the analysis of voltage drops, reactive power flow, and line losses, thereby making it easier to evaluate the steady-state behavior of AC EPSs.

1.1.2 Imbalances and frequency control

The largest blackout in the US history, which also impacted bordering eastern parts of Canada, is an extreme example of how vital the transmission network integrity is for the permanent power balance and how frequency coordination mechanisms acted in this instance to prevent equipment damage. The widespread power outage occurred, when on

August 14th, 2003, overgrown trees came in contact with overhead 345 kV transmission circuits, on a particular hot day, with a large cooling load. In a domino effect, the fault in these lines resulted in a major overload event with an immediate (under)frequency disruption, that triggered a dramatic contingency response. Nearly 50 million people were impacted for several hours, with a staggering 62 GW of load shedding. While, in terms of generation trip, more than 250 generating plants, including 11 nuclear units were off-line at some point during this disturbance period [24].

The utility (nominal) frequency, f_n, in most countries, including in Europe, is 50 Hz, while in North America and a few other countries is 60 Hz. A value that can be measured in all parts of the synchronous area with minimal sub-decimal (mHz) deviations. The existence of power imbalances results in a period of deviation from the standard utility frequency, $\Delta f = f_n - f$. The maximum frequency deviation and the rate at which frequency changes, $\frac{df}{dt}$, depends not only on the magnitude of the power imbalance itself, but also on the inertia of all the synchronized turbine-generators in the system. In practice, this frequency deviation results in a variation of the synchronous machine angular speed, $\frac{d\omega}{dt}$. Thus, for a rotating mass with a total moment of inertia J [kgm^2], and a kinetic energy $E_k = \frac{1}{2}J\omega^2$, the relationship between the speed deviation and the power balance, $P_m(t) - P_e(t)$, is given by equation 1.4 [25]:

$$\begin{aligned} P_m(t) - P_e(t) &= \frac{dE_k}{dt} \Leftrightarrow \\ P_m(t) - P_e(t) &= \omega J \frac{d\omega}{dt} \simeq \omega_{grid} J \frac{d\omega}{dt} \end{aligned} \tag{1.4}$$

where the active power balance component represents the difference between the mechanical power imparted by the prime mover (typically a steam, combustion, or hydro turbine), and the electrical power drawn as a load. While the mean angular speed of the rotating mass, ω, can be approximated by the grid angular frequency, $\omega_{grid} = 2\pi f_n$.

This inertia can be measured in power-time product units (say, GWs), but more often is normalized to its nominal electrical power and so is described in time units instead. A quantity known as inertia constant [s]. Assuming a well-defined EPS, where the load is served by a group of N_{sg} synchronous generators and inverter-based generators (from variable renewable energy generators and battery storage), the inertia constant of the synchronous machines, H_{sync}^{gen} (equation 1.5), can be computed as follows [26]:

$$H_{sync}^{gen} = \frac{\sum_{i=1}^{N_{sg}} H_i S_{G,i}}{P_{Load}} \tag{1.5}$$

where H_i and $S_{G,i}$ are the individual generator inertia constant and its respective nominal power; P_{Load} is again the system load.

In other words, the inertia response highlighted by the first period following a frequency

disruption in Figure 1.4, is largely an indirect measure of the total stored rotational kinetic energy in the spinning generators. Considering (adding) the inertial contribution of rotating loads as well, H^{load}, the total grid inertia is obtained, H^T. As such the total kinetic energy stored in these synchronous generators (including loads), E_k^T (equation 1.6), is defined in power-time product units as:

$$E_k^T = \left(H_{sync}^{gen} + H^{load}\right) P_{Load} \Leftrightarrow E_k^T = H^T P_{Load} \quad (1.6)$$

ultimately, E_k^T and H^T are the system variables that translate the ability of an EPS to oppose frequency variations [27]. Accordingly, these quantities relate with the rate of change of frequency, $RoCoF$. The initial value (infinitesimal) of $\frac{df}{dt}$ after a frequency disruption, denoted as 0^+, is represented by the tangent line in Figure 1.4 (assuming a power loss event), and it can be computed as:

$$\lim_{t \to 0^+} RoCoF = \frac{f_n \left(P_G\left(t\right) - P_{Load}\left(t\right)\right)}{2H^T P_{Load}\left(t\right)} \quad (1.7)$$

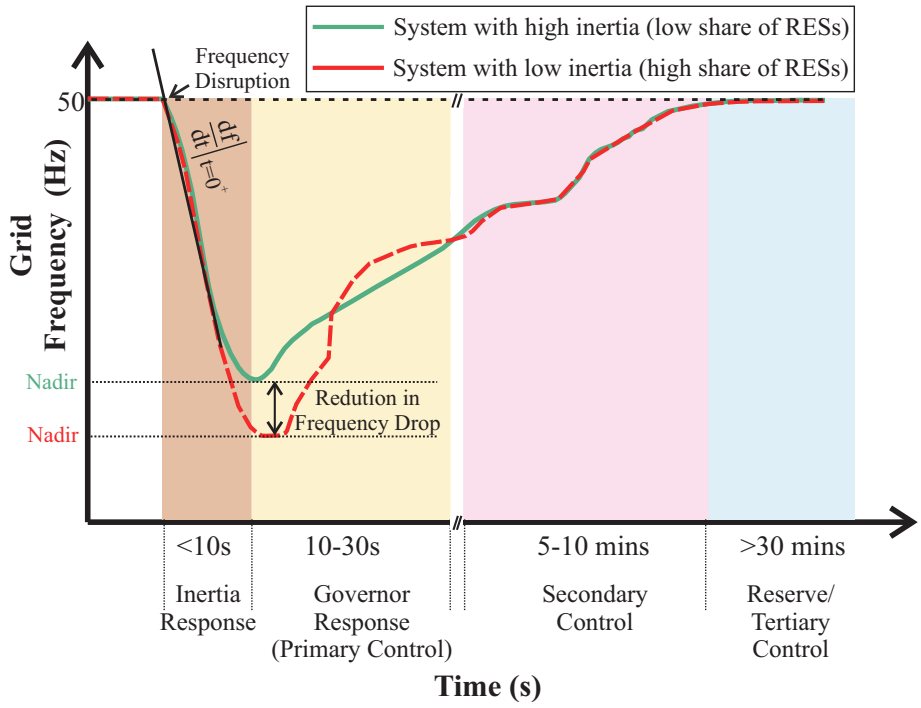


Figure 1.4: Typical EPS frequency response during a under-frequency contingency: control time-frame.

The limit given in equation 1.7 which is also known as swing equation [28], confirms that the "momentaneous" slope immediately after the frequency event is directly proportional to the power imbalance share (of the total electrical load) and inversely proportional to the total grid inertia. The same is also illustrated in Figure 1.4, and that is why for the

same frequency disruption (in this case a power shortage), systems permeated with high inertia, i.e., with more traditional synchronous turbine-generators, record a lower drop in frequency, in comparison with systems with a high penetration of renewable energy sources (RESs), where the inertial response is weaker and so the trough in frequency is more pronounced. This minimum frequency point is known as Nadir, and its equivalent to the maximum frequency deviation, $\Delta f_{\max}$. While for an over-frequency contingency (excess power event), the maximum frequency point is known as Zenith. Typical inertia constants range between 2-9s for conventional synchronous generators [25].

Nevertheless, in both cases, as is also highlighted in Figure 1.4, the frequency drop is significant to trigger the primary control, whose purpose is to quickly restore (at least partially) the active power balance. This is the fastest of the three levels of frequency control mechanisms, acting within seconds (usually after 10s [29]), and it relies on a localized (individual) and automatic speed "governor" response, which prompts synchronous generators to use its additional capacity to increase or reduce the output power [30]. Although it is commonly referred, frequency-sensitive (controllable) loads, such as induction motors, can also participate in this control stage [31]. Primary control is therefore a crucial mechanism for the stability of the EPS, and so participation is mandatory and not remunerated for all the dispatchable generation units connected to the primary transmission network (as mandated by the TSOs) [32]. Typically, this response is sufficient to stabilize frequency in a quasi-steady-state (still with a deviation from f_n) [25], until the higher level, centralized and longer period, secondary control (automatic) and then tertiary frequency control (upon TSO request, acting as a replacement reserve) mechanisms are deployed to sustainably bring the frequency back to its target value. Both of these aforementioned frequency control levels constitute what is known as spinning reserve, i.e., unloaded generation capacity, that is rotating in synchronicity with the utility-grid [33]. The deployment of this reserve, the requirements in terms of its amplitude and duration (also known as operating margin), and remuneration scheme are defined by the TSO. These may vary between countries and are dependent on several factors, including system inertia (generation mix composition), market structure and market interconnectedness [34, 35]. In other words, larger synchronous areas are less prone to suffer significant frequency deviations from events like the loss of a single generator or a transmission inter-connector.

The instantaneous variations in the power balance equation require constant frequency coordination (tight control). In addition, it is important to notice, that the presence of a generation surplus, i.e., $P_G > P_{Load}$, frequency increases $(\Delta f < 0)$, while the opposite occurs for generation deficits (outages) $(P_G < P_{Load})$ and so $\Delta f > 0$. The same is highlighted in Figure 1.5, with the different color bands representing different frequency thresholds with different associated responses (coordination). As was the case with the spinning reserve requirements, these thresholds are also system dependent, but considering reference values, the normal operation (dead band) presents a very low tolerance, $|\Delta f| < 0.1\% f_n$, while governor response mechanisms (as described earlier) are activated for ranges in the order of $0.1\% f_n \leq |\Delta f| < 1\% f_n$. Deviations above 1% (0.5 Hz) are

more troubling, given that the governor's response and the reserve capacity are probably insufficient to restore the power balance. Thus, a more abrupt (contingency) response is needed to contain the disruption and prevent material damage.

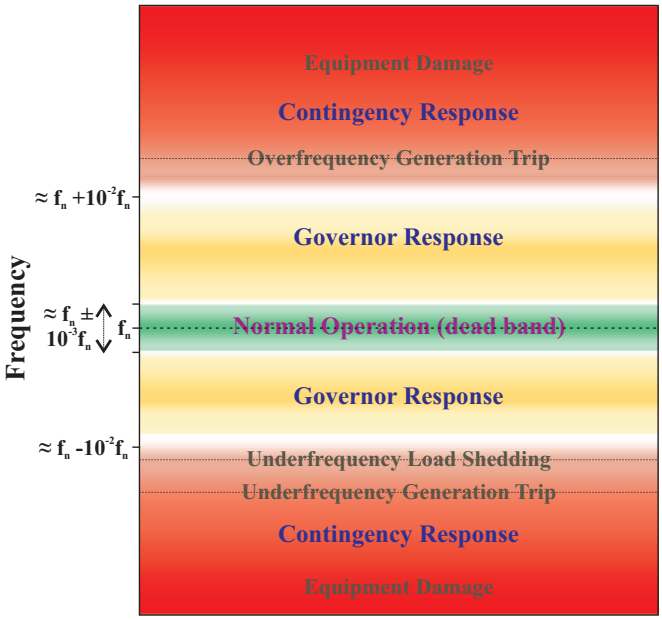


Figure 1.5: Frequency coordination: conceptual view of EPS frequency constraints and main response mechanisms.

Contingency responses are uncommon, but they typically occur in under-frequency situations — due to sudden loss of bulk scheduled generation, unexpected increase in (peak) loads or disconnected transmission lines — requiring the activation of upward operating margins, followed by load shedding mechanisms [28]. Mechanisms that are pre-planned and work by rotating power outages [36], i.e., by disconnecting feeders (opening the circuit breakers) or by reducing load from primary sources [37]. In the case of a more severe contingency, the load shedding is not planned (pre-announced), in what is referred to as a rolling blackout. Subsequently, and for even lower frequencies generation (unit) trip (shut down) can also occur. This last mechanism is key in the case of significant over-frequency events, as is the case of a sudden drop in load (e.g. an energy-intensive industrial plant). In general, over-frequency disruptions are more uncommon than under-frequency ones. For instance, data from Great Britain's system operator, recorded a total of 33 significant frequency events, defined as $|\Delta f| > 0.2$Hz, between January 2018 and October 2019, from which only 5 (15%) were over-frequency disruptions [38].
Analogously to the frequency behavior during power imbalances, voltage levels also vary. In the case of $P_G < P_{Load}$, grid voltage decreases, and when it reaches certain thresholds, it may lead to what is known as a "brownout", an intermediate control stage before blackouts are imposed. During this period, there is a partial reduction in the grid voltage magnitude, which can be accompanied by a selective curtailment of less critical loads [39]. Consequently, appliances may be forced to run slower or even stop and lighting may start to dim. Conversely, when $P_G > P_{Load}$, voltage magnitude tends to rise. Though in both

cases, there is a greater percentage margin for grid voltage fluctuations, with tolerances that go up to ±10%, dependent on the voltage class (for the different transport networks), in contrast, and as discussed, frequency deviations are more critical, with normal operation tolerances between 0.2% and 1% of f_n [40].
Automatic voltage regulators (AVRs), on-load tap changers placed on larger power transformers, and de-energized tap changers on secondary transformers are used to enforce voltage stability. AVRs working principle is based on a permanent voltage error control loop where the generator's excitation is adjusted, for varying loads, acting as a buffer against voltage fluctuations. Additional features include under-frequency compensation and reactive power sharing (voltage droop) between generators running in parallel [41]. As for tap changers, they work by adjusting the transformer winding ratios (normally with coil taps in the primary) to maintain grid voltage within acceptable limits.

1.1.3 Transmission in restructured and interconnected electric power systems

Restructured wholesale electricity markets have assigned the three functional branches of an EPS, i.e., generating, transmitting, and distributing electricity, once reserved to single vertically integrated monopolies, to different entities (actors), operating independently. Transmission System Operators (TSOs) in the EU or Regional Transmission Organizations (RTOs) in the US, are the responsible entities for the bulk delivery of electricity between generation sites and primary distribution networks. In other words, TSOs provide grid access to the electricity market players through capacity auctions (explicit and implicit), ensure the security of supply, safeguard the integrity of the system and oversee the transmission expansion planning [42].
Benefiting from expanded transmission capacity and the harmonization of market coupling regulations, national wholesale markets in the EU have become increasingly interconnected, forming market zones with cross-border electricity trade. As a result, market zones with a pricing and generation capacity advantage—such as countries with significant deployment of renewables—can potentially drive prices down not only within their own national or regional markets but also in neighboring zones, provided sufficient interconnection capacity exists [43]. This phenomenon constitutes what is referred to as "a cross-border driver" [44].
Nevertheless, this cross-border electricity exchange capacity is limited, on scales that vary from country to country. Occasionally, this limitation leads to congestion periods, resulting in different prices for the different market zones (market splitting). This price difference, called price arbitrage (typically hourly values), can lead to sizable financial gains for traders which are also holders of transmission capacity rights. For instance, during the energy crisis in 2022, the yearly average value of one MW of transmission capacity from Germany to its central European neighbors has witnessed a surge of more than five-fold, reaching thousands of euros [45]. Clearly demonstrating why interconnection reinforcements are crucial for a new EPS paradigm, based on a large-scale penetration of "cleaner"

DERs [46].

1.1.4 Electricity markets

When discussing electricity industry restructuring, we refer to a comprehensive transformation driven by economic policies aimed at liberalizing and deregulating natural monopoly industries. This transformation involved profound changes to address system inefficiencies, electricity prices, environmental restrictions, and technological breakthroughs. Looking back, a key outcome was the unbundling of what was previously a vertically integrated structure, transitioning to a horizontal paradigm marked by the influx of liquidity and increased competition [47].

An essential component of this transformation has been the development of auction-based electricity markets, where generation utilities with varying characteristics and at least on paper, importers of electricity from other markets (resellers) compete to serve (supply) load customers, which typically include end-use retailers or guaranteeing suppliers (bulk purchase buying for end-use consumers), large energy-intensive consumers like factories or municipalities and, conversely, exporters of electricity as well [48].

This transformation led to the disaggregation of all-encompassing, monopolistic, state-owned electricity companies, separating generation from transmission and distribution. As a result, these components are now priced separately on customer bills. In many cases, such as in Portugal, transmission and distribution have been privatized, while largely remaining regulated monopoly services. This is only natural, given their role as public utilities, along with factors such as their land footprint, constant expansion needs, Right-of-Way requirements, the necessity of non-discriminatory third-party access to grids and cross-border capacity. Additionally, conflicts of interest, other legal considerations, and inefficiencies arising from building parallel networks — which would contradict the significant economies of scale inherent to such vast infrastructure — further reinforce the need for maintaining such a tight (uni-dimensional) regulatory scheme.

In a brief historical review, Chile is credited with pioneering electricity reform during the 1980s, culminating in the creation of a wholesale electricity market designed to foster competition among existing generation utilities. The underlying thought process was that an open market (low entry barriers) with a leveled playing field would attract capital and, in turn, accelerate investments aimed at gaining a competitive edge (system-wise). This stands in contrast to the more rigid, vertically integrated structure where a single entity controls the entire process, often resulting in a more hesitant approach towards new investments. In its original conception, only large consumers were exposed to unregulated market prices. In contrast, and recognizing the potential volatility of market pricing (a scenario anticipated, especially as market participants were navigating uncharted territories), small consumers — representing the majority of the consumer base but not necessarily the majority of total energy consumption, a crucial consideration for electricity market design and regulatory mechanisms — were safeguarded through a bypass mechanism that largely fixed their energy prices.

Chile's market model, known as "Poolco", operated under a centralized (pool) generation dispatch system, adopting what has become a common feature in virtually all market designs: a two-part pricing scheme. The first component is energy pricing, which in this case reflected the nodal short-term marginal cost, accounting not only for generation-related costs (fuel costs and other operational expenses) but also for transmission constraints which are location-constrained. This approach encourages generation where it is most cost-effective and discourages it in nodes where it may cause grid congestion. The second component is capacity pricing, which remunerates utilities for making their capacities available (or coming online) in the medium to long haul. For instance, during every expected peak demand periods (winter months in the case of Chile), based on criteria such as availability, start-up times, and ramping capabilities (time to reach full load). This mechanism incentivizes energy producers to maintain and invest in additional capacity, which might otherwise not be covered by energy pricing alone (e.g., to recover their fixed costs and support capacity expansion). As a result, it ensures system security even during periods of exceptionally high demand.

In a broader sense, this two-part pricing scheme represents a concerted effort to ensure that, even in a liberalized market, the short-term operational cost-efficiency of meeting supply and demand is balanced with the long-term and capital-intensive grid evolution needed to maintain continuous reliability — an uncompromising trait of any EPS. This point is rather important on a more philosophical note because it highlights that, while the reform introduced "deregulated" energy pricing, it simultaneously resulted in a framework that remains heavily regulated. A transition from a full ownership status or at least operational control to a more indirect yet influential role—acting as a rule-setter and guiding entity(ies). This apparent contradiction is not a fundamental flaw but rather a necessary feature, required to ensure a smooth interplay within a competitive environment.

The truth is, in most western countries, what evolved from the two-part pricing scheme concept has become a landscape with different types of wholesale (systemic-level) electricity markets. Despite varying terminologies, operators, operational scopes, and oversight mechanisms, all these markets share a common goal: the need for a horizontal framework that addresses the time, location, flexibility, and other physical constraints associated with the continuous and dynamic flow of electricity. These constraints — such as demand and, to a greater extent, supply variability (unpredictability-wise), lack of large storage capability, limited transmission capacity, voltage tolerances, start, ramp, and stop limitations, among several others — must be managed to ensure a continuously and virtually instantaneously balanced delivery (a measure of frequency stability as we saw in sub-section 1.1.2). In parallel, this framework needs to encourage competition among participants, system efficiency, and long-term security of supply; therefore, it needs to provide proper risk management tools.

All in all, this is a complex premise but naturally explains why the restructured horizontal picture of the electricity industry requires more than just one market type (exchange), performing different (and complementary) tasks, with distinct price signals that help guide

participants, regulators and policy makers [49]. In simplified terms, and considering the reality in most EU countries, the hierarchical structure comprises four main sub-markets: spot markets, balancing markets, forward markets, and capacity markets. These are illustrated in Figure 1.6, alongside other key sub-markets and mechanisms.

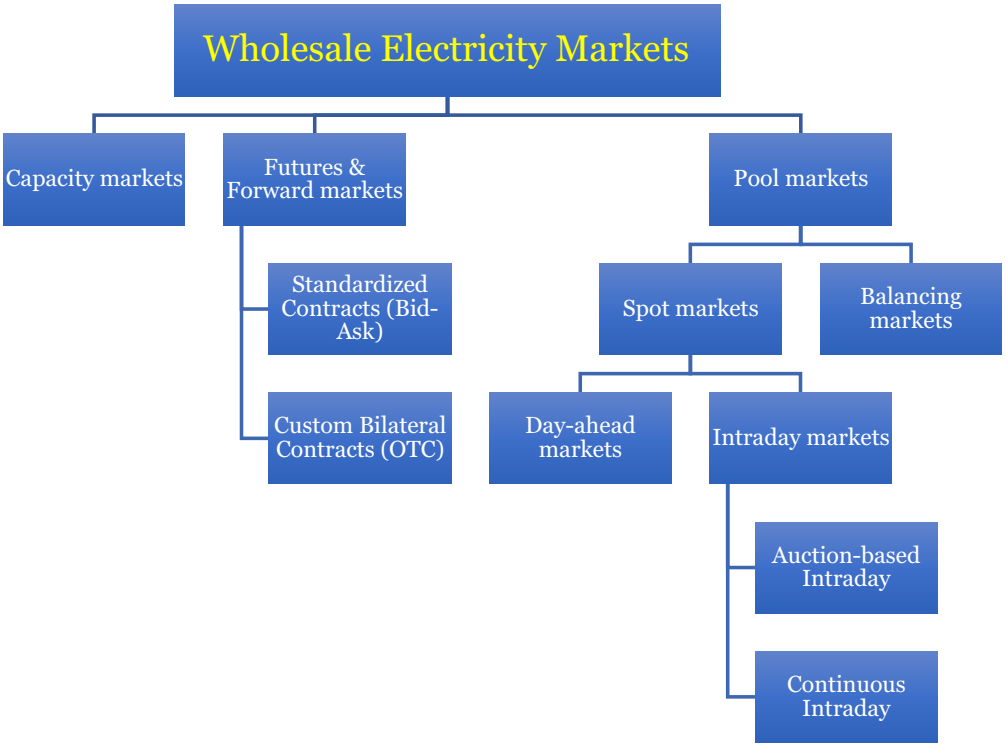


Figure 1.6: Structure (Hierarchy) of most European Wholesale Electricity Markets.

Zooming in on the energy pricing component, it typically encompasses day-ahead, intraday, and balancing markets (markets where there is underlying physical delivery). Operating in a pool type manner, i.e, relying on centralized mechanisms, including marginal pricing, the two first constitute what collectively can be designated as spot markets, because electric energy is traded over short-term or very-short-term horizons, with transactions being settled almost ”on the spot”. The same is largely true for balancing markets as well, however, their focus on real-time system stability, means they are better represented in a separate branch. Despite this, the term ”spot market” is often parsed to imply the day-ahead market.

The day-ahead market, as the name suggests, matches (determines the winning bids and offers) demand and supply for each hour of the next-day delivery. This process relies on forecasted demand and consumers' willingness to pay on one side, and the availability and short-term marginal cost of supply on the other. A unique clearing price (market equilibrium point) is determined at the intersection of the supply and inverse demand curves, based on the last unit of electric energy required to meet demand (known as the marginal technology or supplier), as illustrated in Figure 1.7. Also known as a ”pay-as-clear” auction [50].

Unlike other commodities, load demand exhibits a deeper degree of inelasticity. This

means that almost all consumer offers are positioned to the left (higher) of the clearing price, as electricity cannot typically be deferred or stored by the majority of consumers. The few exceptions include pumped hydro storage stations, certain large consumers, and, more recently, prosumers with flexibility options. The same cannot typically be said of supply offers, which, in the case of reasonably sized EPSs relying on a diverse and over-dimensioned (capacity-wise) generation portfolio, can extend well to the right (higher) of the clearing price (certainly so, when renewables are in the mix). This follows a (hypothetical) merit order, commonly ranked (from cheapest to most expensive technology) as follows: nuclear, wind, solar, hydro, gas, and lastly coal-powered generation [51]. Yet, this is variable. For example, in EPSs like the Iberian Peninsula — and largely due to support measures that we will examine in chapter 2 — (consolidated/mature) renewable sources can auction near their marginal cost, i.e., zero.

Notice that the short-term marginal cost of energy sources that rely on fossil fuels, unlike those that are exempt (renewables), is heavily influenced by carbon emission allowances (which tend to be comparatively costlier in the EU). In microeconomic terms, this indicates that excess supply (surplus) outweighs excess demand (shortage), creating a free-market driving force that encourages suppliers to adjust prices back toward the equilibrium point [52]. However, the extent of this ”force” depends on the specific cost structures and constraints of the auctioning generators — such as marginal costs, carbon permits, and start-up costs — which establish a practical price floor for the various generation technologies.

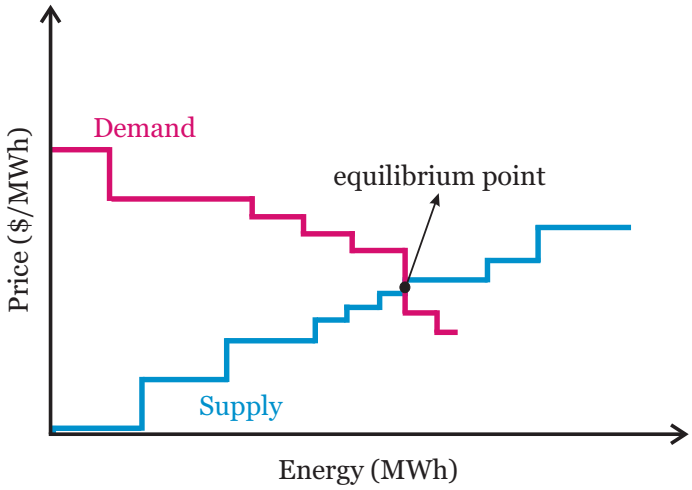


Figure 1.7: Cumulative (aggregated) supply and demand curves derived from consumer bids (magenta) and generator offers (blue), illustrating the typical piece-wise linear clearing method used in the day-ahead market for a specific hour.

In practice, price formation is more intricate due to the interconnectedness between different bidding zones (i.e., not a single zone day-ahead market as is generically illustrated). The existence of these zones implies constraints and losses posed by the available transmission capacity and potential dynamic congestion in the electric power flow between them. This can be seen as a broader application of nodal pricing, known as local marginal

pricing, pioneered as we saw during Chile's electricity industry restructuring and is also common in many US regional markets today, including the well-known PJM Interconnection in the northeastern US. Additionally, the order types (bids and offers) are not all simple hourly bids; they can be complex, such as block orders or linked orders, which are used to accommodate high start-up and stop costs, ramping times, among other things. All this adds another layer of complexity to the matching process. Facing with this added complexity, most European markets rely on the "EUPHEMIA" algorithm, which aims to maximize total welfare [the sum (aggregate) of consumer and producer surpluses (the area enclosed by the demand and supply curves, bounded by the intersection equilibrium point) and congestion rent] while concurrently satisfying the aforementioned constraints.

For all these reasons, day-ahead markets can be considered the primary and most liquid electricity market. A fact that is substantiated by the the overwhelming share of traded electric energy volumes (for example, in Nord Pool, which is one of the main European electricity market exchanges, it accounted for roughly 92% of the total traded volume [51]). Because of this importance, it serves as a reference point for forward markets [50, 53]. Market participants are expected to deliver on their day-ahead commitments in real-time at the firm level. In other words, it is the individual responsibility of each participant, and hefty penalties are imposed when deviations occur. This is understandable given that these deviations will require the activation of other, often time costlier market/system mechanisms to compensate or avoid the potential supply/load mismatch.

This point underscores why, from very early on, regulatory frameworks have primarily evolved around the day-ahead market model rather than other horizons. A 24-hour lead time strikes a balance between having enough lead time to plan generation dispatch, inventory, and maintenance schedules in more traditional generation sites, to have sufficiently accurate demand and weather forecasts (renewable generation), while preserving the ability to respond to more fine deviations and other real-time conditions. This primary electricity market provides a very solid base in terms of the wholesale market needs and consequently, a competitive framework for the expected generation to compete and allocate its resources, while attending generation and transmission constraints.

The remaining bulk volume is usually traded in intraday and balancing markets. As the name indicates, the first, intraday markets consider delivery on the same day, and they can be auction-based (pay-as-clear) as well, but they tend to have a continuous function, meaning trading occurs continuously up to the delivery period. This marketplace, which opens shortly after the day-ahead market, provides an opportunity for market participants to correct/adjust imbalances resulting from inevitable demand and non-dispatchable supply forecasting errors or other unexpected events. These errors become more apparent as the delivery horizon shortens. Common trades include selling short term over-capacities or buying additional short-term load demand needs. However, unlike the day-ahead market that operates on a centralized aggregate auction model (as exemplified in Figure 1.7), continuous intraday market prices are set through a "pay-as-bid" mechanism, akin to the stock exchange, where the latest trade reflects the current market price. In simpler terms,

a direct settlement between two parties (bid has to match an offer). Interestingly, the UK is one of the very few examples of a day-ahead electricity operating on a pay-as-bid basis.

A consequence of the growing share of variable renewables is the increasing volume of orders and energy traded on intraday markets. This trend is evident in Germany, where as of 2018, the intraday volume already can account for more than a sixth of the day-ahead market volume, and there is an apparent shrinking in the pricing disparity between the two spot markets [54].

Balancing markets, on the other hand, are where energy is sold for almost immediate delivery. Their purpose is to continuously address discrepancies between supply and demand, preserving crucial grid frequency stability. This is why they are typically operated by Transmission System Operators (TSOs), who are responsible for scheduling generation based on the day-ahead market outcomes within their control zones and continuously monitoring system imbalances.

Based on these imbalances and the availability of generation flexibility (dispatching up or down) and/or willingness to consume, the most cost-effective orders are selected to balance the system while adhering to transmission constraints. TSOs have a range of balancing options at their disposal, commonly referred to as ancillary services. These include balancing markets as a key part, alongside other options such as voltage control, primary and secondary reserves, and black start capability.

Depending on the severity of any mismatch, TSOs can activate different levels of reserves (including redispatching generators) at various operational time frames, typically ranging from a few seconds to a quarter of an hour under established contracts [51]. Similar to intraday markets, increased generation (supply) variability can heighten the frequency and volume of trades within this market.

This model for balancing markets, which is widely used in European countries, contrasts with the approach followed by US regional electricity markets. In the US, the clearing price for both day-ahead and real-time markets — operating continuously and overlapping (encompassing) with the horizons and function of Europe's balancing and intraday markets — is determined by a single entity, often the Independent System Operator (ISO). Moreover, in the day-ahead market model, introduced during the broad U.S. electricity restructuring that established ISOs and regional TSOs in the 1990s and early 2000s, an interesting and intuitive feature of virtual bids was adopted. These have a hypothetical or exploratory nature, allowing participants to place orders in the day-ahead market without any physical commitment, which are then offset in the real-time market. In other words, speculate on the spread between the day-ahead and real-time market prices. The rationale behind this feature is that it can improve market efficiency by enhancing price formation, i.e., promoting an alignment between price expectations for both markets as more virtual bids are placed. This also reflects a broader conceptual role for real-time markets, which, unlike their European counterparts, are not exclusively tied to continuous balancing operations. Instead, they can also be used to hedge positions and capitalize on price fluctuations. However, this feature may introduce volatility, as excessive "virtual-

ization" of orders could destabilize the market. That is why, in most regional US markets, virtual bidding is subject to stringent credit requirements and transaction fees, which acts as filter and reduces the need for rigid, hard caps on the quantity or volume of virtual bids (which otherwise could distort the original goal of liberalization and confined regulation).

Another trading option that emerged relatively early in the history of electricity market implementation was a framework enabling generation utilities and consumers to contractually agree on quantities and prices beyond the very near-term, pool-based energy window of the spot markets (i.e., covering longer delivery dates). These direct-type agreements, known as non-physical bilateral contracts with different maturities can constitute either a forward or future contracts. These two are often mistaken or used interchangeably, but they are not the same. Forward contracts are privately negotiated and reflect an over-the-counter (OTC) approach, meaning direct settlement without third-party involvement. This results in more tailored agreements but also higher counter-party risk due to lack of oversight. Futures contracts, on the other hand, are standardized with fixed terms and payment schedules, traded on publicly regulated exchanges. This increases transparency and reduces risk, though the downside is less flexibility regarding delivery times or payment structure and the inclusion of exchange trade clearing fees. Power Purchase Agreements (PPAs) are a popular type of bilateral OTC contract, particularly favored by corporations. These agreements enable companies to secure long-term energy supplies from renewable sources, thereby reducing their carbon footprint and obtaining renewable energy certificates. This not only enhances their sustainability credentials but also makes them more appealing to investors. However, in practice, if the renewable energy supplier is unable to deliver for instance due to the variability of renewable sources, the utility may need to find a replacement on spot markets. PPAs can be physical or virtual (financial), but in both cases, the energy received by the buyer does not necessarily have to be the actual renewable energy generation from the supplier due to the nature of the grid (more likely is from a mix of sources). However, the net effect is still ensured, as an equivalent "carbon-free" energy quantity is consumed system-wide.

These primarily financial contracts provide participants with the flexibility to hedge against price volatility in the short-term marginal cost of energy, serving as essential risk management tools in competitive, deregulated pricing environments. They resemble futures contracts in the stock market, where two parties agree (lock) on a price for a stock or commodity today, but the actual settlement occurs at a later date. The contract is then settled based on the difference between the agreed-upon price and the prevailing spot market price. In the context of electricity markets, this means that the agreed-upon bilateral price does not influence the physical delivery of energy but instead determines who compensates whom (purely financial transaction) — and by how much — based on the spread between the actual spot price and the contractual price. Still, the complexity and inherent constraints on how electricity is generated, transported, stored, and consumed mean that, in general, is difficult to anticipate with great certainty, and so liquidity is thin in electricity futures markets. This explains the low trading volumes and reduced number

of market participants [55].

On the other hand, capacity pricing is not directly related to any of present markets. Instead, it usually operates as a separate market (mechanisms) to incentivize generators to maintain available capacity for peak demand periods. These mechanisms exist in many forms and are commonly managed by the TSO, conducting a procurement that can span over a multi-year period. Capacity payments are therefore calculated based on longer-term considerations, such as ensuring security of supply during periods of high consumption (like in Chile's peak case, usually between May and September). Paying gas-fired power plants to remain on standby and ready to ramp up production during potential shortfalls is a common example of a capacity payment.

Overall, we can see that this multi-layered electricity market structure simultaneously serves two inseparable goals: addressing real-time operational needs and ensuring long-term system security. That means, than in essence, the functional time domains of these markets is the best descriptor of their well-defined but different roles, as is depicted in Figure 1.8.

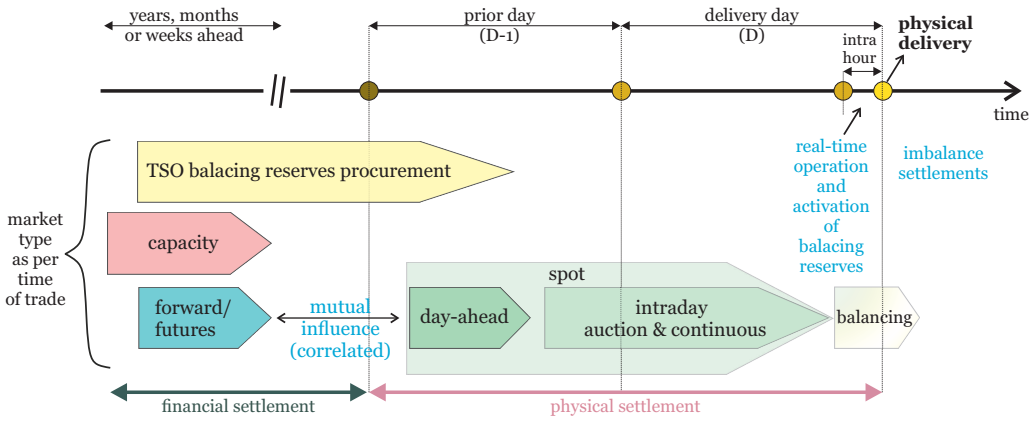


Figure 1.8: Overview of the operational time-frame and settlement type for the different wholesale electricity markets in the EU (adapted from [56–58]).

Returning to the historical review of the transition from a monopolistic framework to a more competitive, private-sector-based electricity market, and with regard to the European continent, the UK and Norway were the first to adopt pool-driven electricity markets in the mid-to-late 1990s [47]. Following their lead, by the mid-to-late 2000s, nearly all European countries liberalized their power generation industries and established their own national electricity markets—or, in many cases, transnational markets. These transnational markets enable multiple countries to participate in shared trading platforms, leveraging resources and interconnections across regions, especially where spatial proximity makes such cooperation more practical. By pooling generation capacities and enhancing cross-border coordination, participating countries achieve mutual benefits, notably: i) a more diverse energy generation portfolio; ii) optimized allocation of generation resources across spatially closer regions; iii) reduced transmission losses via otherwise shorter interconnections (vis-a-vis a longer and reinforced national electricity transport network);

iv) capitalizing on interconnections to transmit renewable energy from surplus regions to deficit regions, thereby reducing overall system costs; v) promotion of the integration of generation technologies with low marginal costs; vi) fostering cost efficiency through enhanced market competition; vii) enhanced grid stability during peak demand or shortages, among others.

One of these multinational electricity markets is the Iberian Electricity Market (MIBEL), which began operation in 2007. Its spot markets include participants from both Portugal and Spain and is managed by the Spanish side of the Iberian Energy Market Operator (OMIE) [59]. The clearing mechanism accounts for possible mismatches between cross-border offers and interconnection capacities, representing the commercially available energy flow between the two bordering countries. Whenever such mismatches occur, the rules dictate that the two market zones — Portugal (PTEL) and Spain (SPEL) — be separated, requiring distinct prices to be determined for each area, in a mechanism known as ”market splitting”.

As for the Iberian forward market, it is also managed by the Iberian Energy Market Operator but is Portuguese-controlled (OMIP). This market enables participants to trade energy derivatives, including futures contracts with maturities ranging from days, weeks, weekends, months, quarters, and years. These contracts involve daily settlements (margins) based on the difference between the transaction price and the daily market quotation (futures). OMIP, which also acts as a clearing house, is responsible for settling the daily margins as well as the contracts on their delivery dates or periods.

Over the years, the regulation of MIBEL has been continuously updated, not only based on operational experience and public consultations but more intensively due to the persistent application of EU directives. These constitute a set of common rules for the internal market organization, encompassing all facets of the energy equation—generation, transmission, distribution, and consumption. With the ambitious goal of achieving an EU-wide integrated energy market (for which cross-border capacity is strongly promoted), the focus lies on producing a substantial share of electricity from renewable energy sources. This is complemented by increased trading (purchasing) of greenhouse gas emission allowances [47] — as we already saw an indirect financial type of deterrent for greenhouse gas emissions mostly associated with conventional generating technologies — supported by a cap-and-trade mechanism where revenues from auctions are reinvested to bolster the ”green” transition. Regarding the expectation of wide renewable capacity integration, the Iberian countries have been at the forefront of this transition, even anticipating some of the timelines defined in those directives. Their annual share of renewable energies already accounts for a significant portion of the energy mix.

However, this progress has also led to increased electricity price volatility, particularly in the day-ahead market, which is the most important market signal [60]. Faced with this uncertainty in a competitive market environment, and within a regulatory framework that heavily discourages both over-estimation (often regarded as worse) and under-estimation through hefty firm-level market penalties, which essentially determine trading corrections

in the generally costlier intra-day markets, it underscores the critical importance for all market participants, as well as ISOs and TSOs, to explore and enhance forecasting models as essential decision-making tools.
To wrap up, after the decentralization and liberalization of functions and roles, electricity and capacity markets, but more clearly spot markets have become a critical component of the electricity industry. They ensure that electricity is traded, priced, and transported efficiently. In the challenging scenario of transforming the generation mix, a broad range of system-wide hurdles and nuanced considerations must be addressed (as the motivation sub-section will delve into) to understand how the market interface can remain faithful to the core principles that initiated this shift in the first place.

1.2 Motivation and Objectives

In light of the necessary shift to a “carbon emissions-free” generation portfolio, a large share of the load demand will be served by a mixture of highly fluctuating and/or variable renewable energy sources. This scenario is no longer as far-fetched as it may have seemed just a few years ago. In fact, several small to medium-sized developed countries with high renewable energy potential, like Spain and Portugal, have been making great strides in this area. These strides can be measured not only by the increasing shares of annual energy served by these clean technologies but also by the ”general” downward effect on electricity market prices.
As argued by Reiche [61], EU countries employed various renewable promotion strategies. Some countries, such as Spain and Portugal, were more aggressive, while others, particularly in Central and Eastern Europe, were less so. These strategies involved many different components, namely: i) different levels of direct or indirect public investments; ii) long-term planning security for investors (predictability); iii) guaranteed remuneration or quota-based schemes (such as feed-in tariffs or renewable purchase obligations); iv) promotion of energy auctions and other favorable market conditions; v) lower grid access (connection) charges; vi) investments in (new) grid injection points and overall grid expansion; vii) punitive levies (disincentive-based taxation) for carbon-emitting technologies; viii) measures to reduce local resistance against renewable projects, etc. Consequently, the success of these strategies in increasing the share of renewables in the energy generation mix varied. However, as is also argued by Reiche, this was also expected, given the disparity of the starting framework in each country. Nevertheless, the commonality of most EU countries’ approach was a demand-oriented promotion of renewables, i.e., overly centered on price guarantees per unit of renewable energy delivered to the grid. In contrast, our transatlantic counterpart, the US, opted mainly for a supply-oriented approach, with targeted subsidies and tax exemption mechanisms, reflecting a milder and less centralized approach [62].
On a more superficial first thought, aligned with the public’s perception, one might think that the EU’s approach is the clear winner, with very few hurdles. Yet, this is not as simple

to conclude and despite being lessly talked, these strides came with a cost, that cannot be simply measured by carbon emission savings or even dollar savings. The analysis is often distorted in the public sphere to just a few percentage elements, which imply that further continuation of the current path does not entail serious challenges in terms of power system operation, control, and planning. Given this predicament, decision-makers should do their utmost to continue promoting the integration of renewables. However, they should take into account the renewables ”malleability” fallacy [63], the ”hidden” environmental costs from mining to installation [64] (e.g., the heavy reliance on critical metals), and the need for affordability to manifest sooner rather than later for public acceptance to increase. For example, industrial electricity prices in the UK, a pioneer in onshore and offshore wind adoption, are on average four times as high as in the US despite the enormous subsidies [65]. Reflecting on the associated transitions lock-ins, on how these can reduce overall system flexibility and ultimately hamper the goal for which these ”near-sighted” policies were designed in the first place. Other concrete examples of these lock-ins are shown in Figure 1.9 based on [63, 66–71].

Given this background, it is useful to conduct a comprehensive assessment of the overarching path that has been followed in Iberia in terms of key energy policy decisions. This includes reassessing some of the premises and goals, highlighting key lock-ins, and rethinking priorities to emphasize top-tier levels of system flexibility — the power system’s ability to maintain the supply-demand balance despite uncertainties, while adapting to regulatory frameworks, emerging technologies, and evolving customer needs — and security of supply. This will require, on the generation side of the matter, a profound understanding of intrinsic characteristics and developments of the generation sources and technologies that have been promoted, questioned, ”frozen”, or even phased out, and those that will be in the future. In the case of renewables, there is no doubt that solar and wind technologies are experiencing significant momentum, surpassing the more established and traditional resources of hydropower and bioenergy (which utilizes well-developed steam turbines). Nonetheless, discussions surrounding the large-scale adoption of renewables must also consider the promising potential of ocean energy [4], particularly in coastal countries like Spain and Portugal.

Coupled with this perspective on an increasing foreseeable variable generation-side, it is necessary to study the developments in terms of storage technologies, and in particular, how broadly they can tackle or even put in question the long standing limitation of electricity to be easily queued or stored in large quantities—a unique trait in comparison with other commodities. On the load side, it will be crucial to not only preview structural changes expected on the demand side (electrification of the transport sector comes to mind) but also in terms of demand-response management, where end-consumers are expected to be more actively engaged.

Figure 1.9: Renewable energy lock-ins (adapted from [63]).

As it stands, if left unaddressed or overlooked, a grid dominated by a large share of highly fluctuating, inverter-based generation (such as solar photovoltaic power sources or more modern type 3 and type 4 wind turbines) increases the likelihood of mismatches between active power and electrical load. This is particularly true until more cost-effective and sizable levels of energy storage solutions are developed and go online. This, in turn, leads to greater frequency deviations, with a detrimentally higher $RoCoF$ (as given by equation 1.7). This occurs not only due to the raised potential for greater active power imbalance (numerator), i.e., energy source variability translates into higher forecasting errors, but also due to the small denominator given the loss (missing) inertia. In fact, low system inertia can lead to much faster frequency excursions for even smaller generation-load inequalities, which consequently may be enough to trigger a generation trip or load-shedding scenario. Other issues include voltage stability disruptions, potential transmission overloads (sudden surges and drops in radial networks force the redistribution of power flows and may overload adjacent lines), and increased wear and possible damage to the grid infrastructure [72].

Additionally, a market overwhelmingly dominated by renewables with merit-based price formation can send misleading price signals due to the typically subsidized nature of these energy sources. This can result in an inaccurate reflection of the true cost of renewable energy integration at both ends of the spectrum. For example, excess generation and saturated interconnection can drive prices to negative values, reflecting curtailment and controllability difficulties. Conversely, during periods of low renewable generation, the market can become overly dependent on the closing technology. For instance, on summer days when wind and hydro power are not available in significant shares, and with an insufficient solar photovoltaic generation profile, prices often rise due to the higher marginal costs and the potential need for a quick generation ramp-up of the (fewer/remaining) thermal units that will determine the equilibrium price and quantity (energy).

Heightened price volatility can be challenging for market stability and planning, which in turn undermines market sentiment, increases financial risk, and may compromise medium to long-term investment decisions. A secondary consequence is that this added uncertainty often leads regulators to intervene in overly strict, single-sided, and/or complex ways, to tackle the root cause problem. However, in doing so, they may inadvertently increase other costs, reduce market efficiency, competitiveness, hinder innovation, and inhibit a more widespread market participation. A real example of the impact of significant price volatility is the retraction of futures market players from their bidding strategies due to the risk of substantial deviations between their projected buy or sell prices and the actual day-ahead market prices. This leads to reduced liquidity, wider bid-ask spreads, and disrupted hedging strategies, creating a cascading effect that may exacerbate and/or prolong the initial price volatility.

Against this backdrop, accurate forecasting models, developed on a solid foundation—including the crucial role of selected data in enhancing the performance of prediction models, from the simplest to the most complex—are essential decision-making tools for

market participants and energy regulatory entities. These models help manage the risks associated with integrating unpredictable energy sources and mitigate the expected price volatility.

Additionally, more frequent imbalances and other related grid disruptions will inevitably require heightened intervention by the TSO and market operator, leading to additional balancing and other type costs, including (deriving from): i) larger trading volumes in the intra-day and continuous markets (where participants adjust their positions to balance supply and demand in real-time), often incurring in higher fees and operational expenses; ii) the costs of activating reserve power plants; iii) the costs of activating and managing energy storage systems; iv) the costs of implementing demand-side response measures, etc.

The AC optimal power flow is a critical problem in the EPS field. Compared to the classic Economic Dispatch problem, which primarily focuses on minimizing generation costs at a given time for a certain load demand, it presents greater optimization complexity. This is because it aims to find the operational variables that minimize single or multiple objectives while fulfilling various generation, network, and other constraints, making it particularly challenging. The aforementioned widespread integration of renewable energy sources and their intrinsic variability brought a renewed interest in this problem. Specifically, there is a focus on optimizing resource utilization, reducing costs, preserving grid stability, and supporting decarbonization goals. To achieve these objectives, key components include energy storage, demand-side management, and more accurate load and generation forecasting.

Given the expected decentralization, conciseness, and affordability of renewable electrical generators and battery storage solutions, hybrid energy systems are consolidating and are closely tied to the concept behind microgrids and the more recent advent of energy communities. The goal of reducing interaction with the main grid is seen as very promising. This approach not only addresses the challenges of ensuring a more sustainable grid but also offers the added benefit, as briefly mentioned in section 1.1, of reducing system losses by directly injecting power at the primary and even secondary distribution levels. In this context, and connecting the dots with electricity price prediction, this thesis also aims to develop hybrid management strategies—another leading research topic in Electric Power Systems. These strategies focus on the optimized operation of a renewables-based portfolio that relies on various energy sources (with different traits) and storage units to balance supply and demand in a more cost-driven manner.

All the points discussed above are deeply intertwined and only reinforce the necessary policy coordination and operational astuteness needed to maintain reliable system operation during such a transformative process. Importantly, it all emerges from the acuteness of the power balance equation considering losses as given by 1.2. This transformative process targets the impacts caused by the EPS's greenhouse gas emissions, which cannot be ignored given its long-term contribution to societal development. In fact, it is leading the charge by tackling the problem head-on, as hinted in this chapter and further elaborated

in the forthcoming chapters.
Nevertheless, it is worth remembering (since it is often underestimated and/or omitted) that this transformation does not start from a clean slate or with the benefit of a restart after a pause. There is an entire infrastructure and chain of actors at different horizontal levels encapsulated in this process. Therefore, serious contemplation through research and experimentation is needed, clearly highlighting challenges and opportunities before such structural changes are advised and in what models/extent for such a vast but finely calibrated system. An outlook illustrated in Figure 1.10 embodies the scope and motivation of my thesis.

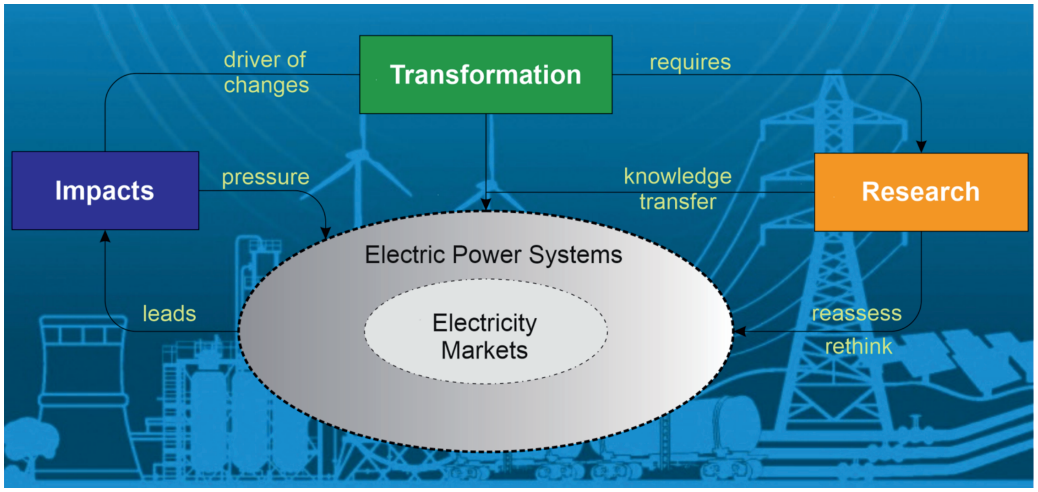


Figure 1.10: Transformation in the Electric Power Systems: Thesis scope.

Having provided the motivations, the specific and explicit objectives of this thesis can be summarized as follows:

- a report of critical energy-related policy aspects surrounding the large-scale penetration of renewables in the Iberian Peninsula (evolution of the generation portfolio);
- a thorough study on the state-of-the-art action vectors that can increase system flexibility;
- a discussion of the most promising action vectors (alternative renewable energy sources, storage solutions, demand-response and others), a policy guideline highlighting propitious flexibility traits, opportunities and also misconstructions;
- a targeted review of the regulatory framework and its impacts, especially from a system's flexibility point of view;
- investigate the impacts of this broad renewable integration on the AC optimal power flow problem;
- identification of key day-ahead electricity price drivers in the Iberian market;
- development of optimized forecasting models for load demand, electricity prices and renewable power generation;

- implementation of a daily management (dispatch-scheme) for hybrid energy system.

To cover the scope of this thesis, a strong theoretical grasp and recapitulation of several topics are essential, including:

- electric power systems and electricity markets fundamentals;
- unconstrained and constrained optimization;
- optimal power flow;
- time-series models, probability distribution and error metrics;
- machine learning fundamentals and neural networks.

1.3 Structure of the thesis

This thesis is organized into eight chapters and five appendices (A–E), which provide supporting material for the core chapters
Throughout this thesis, the necessary theoretical background or historical context will always precede the presentation of key discussions or results, ensuring that the reader benefits from the author's understanding and interpretation of the concepts and the broader EPS "enterprise", which is often deeply intertwined with various topics. This conscious decision, which already began in this introductory chapter, results in a more pleonastic approach to the writing, with the intent of establishing a clear academic foundation. This approach aims to minimize the author's biases and reliance on past and/or flimsy assumptions, providing a grounded point of view of the subjects, that logically leads to the obtained findings and conclusions presented.
Chapter 2 lays the groundwork for addressing the central theme of this thesis, serving as a cornerstone chapter in exploring the dynamics of the large-scale penetration of renewables — the core question of this thesis, as outlined in the motivation section 1.2.
It begins with a study of the macro-evolution of the generation mix in Iberia, highlighting and discussing the thought process, key decisions, the resulting mix, and the remarkable adoption that renewables have already achieved in recent years. This is followed by an in-depth review of some of the most prominent and complementary options and mechanisms addressing both sides of the power demand balance equation, including lesser-known advancements in the conversion efficiencies of dispatchable technologies.
Having established the context and equipped with the necessary theoretical framework, along with a clear understanding of the motivation and research gap outlined in the previous sections, a critical and open-minded discussion is presented on the current approach to large-scale penetration of variable renewables in the Iberian market and its effectiveness — or lack thereof — in addressing the challenge of system flexibility.

In alignment with this, the chapter concludes with a set of energy policy recommendations, specifically tailored to the technical challenges of sustaining large-scale penetration, based on the most promising and carefully evaluated options

Chapter 3 builds upon the discussion of system flexibility, shifting the focus to regulatory aspects. It examines the soaring day-ahead electricity prices in the Iberian market in 2022 as a real-world case study, supported by an analysis of key regulatory documents and system evolution statistics beyond the generation-side infrastructure explored in chapter 2. The discussion offers insightful reflections on the main drivers shaping European energy policy, highlighting in particular some of the shortcomings in the adopted regulatory rationale.

With these two extensive chapters, we conclude the macro-level analysis of key challenges and options surrounding the thesis title: Large-Scale Integration of Variable Renewable Energies in Electric Power Markets, touching on the objectives laid out in section 1.2. This effectively marks the completion of the first part of the thesis.

Building on this foundation, the subsequent chapters draw upon the author's previous research in his master's thesis, shifting focus from power system planning to the operational aspects of electric power systems. While less conceptual in nature, these practical and more micro tasks remain intrinsically tied to the broader macro transformations discussed earlier—a connection that will be revisited when presenting case studies and interpreting key results.

The discussion centers on two fundamental tasks: forecasting and optimal power flow—both of which have direct implications for dispatch operations, thereby justifying the inclusion of Optimization Models in the thesis title.

Nevertheless, before introducing any specific case study or methodology, and in alignment with the comprehensive approach taken in chapter 1, Section 1.1, the next two chapters — along with chapter 6 on Optimal Power Flow — serve as a preface to any problem-solving tasks, including the proposed methodology and case study. Specifically, they provide a broad theoretical overview, establishing the essential foundations required to contextualize the methodological choices made in designing and fine-tuning the adopted approaches. These considerations are later detailed in the chapters that present comparative results.

Accordingly, chapter 4 presents essential concepts and formulations in optimization theory, with a particular focus on the broad taxonomy of optimization methods, the necessity of global-type approaches — especially metaheuristics — and other key aspects/subjects such as objective function properties, gradient information, constrained optimization, penalty methods and multi-objective optimization. Additionally, to consolidate the understanding, the chapter concludes with the implementation of a surrogate-assisted optimization based on a metaheuristic algorithm.

In turn, chapter 5 is dedicated to developing a deep understanding of key concepts surrounding the use of machine learning tools, particularly those designed for regression tasks, such as artificial neural networks. Adopting a deductive approach, the chapter aims

to navigate the complexity of these black-box models.
Beginning with the neuron as the fundamental processing block, the discussion progresses to the training of entire models, emphasizing the importance of regularization techniques and the common principles underlying various architectures. All while touching upon historical milestones that have shaped the field.
Chapter 6 presents a theoretical overview of a fundamental optimization problem in EPSs theory — the Optimal Power Flow problem. This chapter reinforces the technical complexities and mathematical foundations involved in load flow analysis, a crucial aspect in both energy policy formulation and operational decision-making. Additionally, it examines how various optimization objectives, such as cost minimization, voltage stability enhancement, and grid security constraints, shape the OPF framework. Lastly, it explores how metaheuristic approaches, including swarm-based intelligence algorithms, can be adapted to effectively solve OPF problems.
The penultimate chapter, chapter 7, building on all previous introductory knowledge and considerations, provides a targeted empirical examination of three key time-series in the EPSs context: electricity load (demand), day-ahead electricity prices, and a promising renewable and variable generation source. It includes a thorough state-of-the-art review, addressing both classical statistical forecasting models and contemporary machine learning-based approaches, including various types of artificial neural networks, deep learning architectures, and ensemble methods.
This chapter conducts tailored forecasting studies, not only exploring the intricacies and hyperparameter optimization of non-linear machine learning models versus classic statistical time-series forecasting techniques, but also evaluating how hybrid methodologies can enhance predictive accuracy. Furthermore, the proposed methodologies emphasize feature selection strategies, ensuring the most representative labeled data is used for optimal forecasting performance. Special attention is given to error metrics and validation procedures, such as RMSE, MAPE, and cross-validation frameworks, to rigorously assess model effectiveness in real-world time-series applications.
This thesis concludes in chapter 8 by outlining the final inferences drawn from the entirety of this body of work, including some research directions that have opened new avenues for further studies.

Chapter 2

Large-scale penetration of renewables in the Iberian power system

The seriousness of the global climate crisis makes the path towards the decarbonization of societies irreversible. In response to this call for action, the entire power system community has been actively researching, proposing, and implementing measures to minimize the share of emissions. Instigated by ambitious European Union (EU) legislation, increasing public awareness of the problem, and the competitiveness of PV and onshore wind technologies, the past two decades have seen several EU countries paving the way towards large-scale penetration of renewables, such as Portugal and Spain, where the share of renewables has grown by leaps and bounds. As this process has unfolded, the large scope of challenges posed by integrating these two technologies requires further debate. These include technical and conceptual aspects of the infrastructure, such as an increasing number of grid connections, lack of inertia, power balancing issues derived from the stochastic nature of renewables, and issues of market participation and fairness. Notwithstanding, this has also "opened the door" to a broad range of solutions that can effectively tackle the lack of system flexibility. Among them, the Iberian Peninsula is a privileged spot to invest in promising renewable energy sources with attractive energy densities to complement the existing portfolio, namely ocean, offshore wind, and hydrogen technologies. Hence, this chapter analyzes and discusses this transformative process, promotes a better understanding of its challenges, and provides guided policy suggestions to move forward while preserving the security and resilience of the system.

2.1 Introduction

The unceasing emissions of greenhouse gases (GHGs), particularly CO_2, are a leading cause of human-induced global warming. Moreover, more recent scientific assessments paint a very "dark picture", raising the alarm about the inevitable and dreadful consequences for human life if concerted and forceful actions are not pursued [73]. As expected, the electric power system (EPS) sector, with its indispensable role in daily life, has a fair share of responsibility in terms of carbon emissions. In response to this call for action, the entire EPS industry has been at the forefront of these efforts. For instance, Gielen et al. [74] provided an impressive estimate of the sheer scale of the task at hand; that is, for energy-related CO_2 emissions to comply with the Paris agreement targets, the share of renewable energy in the global power sector would need to increase from roughly 25% in 2015 to 85% in 2050, which in turn implies a very costly financial burden.

In this regard, the ambitious set of national and community policies adopted in several EU countries, particularly the more soundly "Renewable Energy Roadmap" (Directive

2009/28/EC [75]) and "Energy Roadmap 2050" [76], have successfully promoted the integration of large quantities of renewable energy sources (RESs) and the gradual phase-out of coal-fired power plants. Hence, Fekete et al. [77] described these as prime examples of effective mitigation policies implemented by major emitting economies. Worldwide, the urgency of the problem has also been recognized. For the last couple of years, the added renewable capacity has accounted for the majority of newly installed power generation capacity and constantly surpasses the prior year's record [78, 79]. The expanding demand for RESs has prompted research and production investments, which, in turn, have led to economies of scale that have resulted in rapidly declining costs (maturation) of renewable technologies, particularly wind and solar power [80]. In most situations, this is the least expensive option for adding marginal capacity to the system [78]; this has helped accelerate the power sector's transformation towards a more CO_2-free energy mix.

However, the path towards decarbonizing the power sector is demanding and sinuous, as it entails a series of challenges that impact the entire power system infrastructure. In particular, the old concentrated and vertical paradigm, where, most of the time, one or very few companies control almost every aspect involving electricity, is no longer true. Indeed, in the current scheme, the opposite occurs because of more fragmented, liberalized, and market-centered functioning. Thus, the intricacy of the interdependencies and coordination between all stages that comprise the generation, transmission, distribution, commercialization, and end-use energy sub-sectors should not be underestimated [81]. Consequently, a more attentive and proactive supervisory framework is crucial for harmonizing decarbonization goals with a modern power system's stability and security principles. In this regard, numerous studies have been conducted to study the viability of this high penetration level of RESs, and a consensus is that a multilevel approach can make this scenario feasible [82].

Among the wide range of challenges posed by the large-scale integration of solar PV and wind renewable energies, their intrinsically variable nature is the most problematic. Differing (fundamentally) from the conventional dispatchable technologies in the following ways [83–85]: i) variable renewable energy (VRE) output is variable; ii) VRE output is unpredictable; iii) VRE generators are often modular and smaller in size; iv) VRE generators tend to have low short-run costs; v) VREs power densities are considerably lower than thermal-fired generation, particularly when compared with natural gas and nuclear power plants; vi) VRE generators are location-constrained (to effectively explore its potential); vii) Solar PV produce a DC output, requiring costlier inverters; viii) Most common type of wind generators are non-synchronous (three-phase asynchronous generators); ix) VRE sources reduce country's energy dependence, but without the proper investments may increase its technological dependence. The same applies to the PV panel market, where China dominates worldwide production [86].

The state-of-the-art reveals that a large focus is centered on projected scenarios of renewables in the system, simulating the required levels of RESs [87], with inferences about the overall cost, the existence of surpluses, and the need for interconnection capacity [88, 89].

However, energy security has often been undervalued. Recognizing this, the study presented in [90] combined life cycle assessment and energy systems modeling to study the energy security of the Spanish generation mix; interestingly, they found that the objectives are not incompatible, and higher energy security targets can help fulfill climate change goals. Nevertheless, considering that bordering countries share a common electricity market, there is a gap in the literature regarding the issue of flexibility, particularly in addressing the role of dispatchable thermal capacity while detailing important policy decisions in the two countries.

Considering that the Iberian case is often pointed out as a role model for a rapid adoption of VREs, with review works mostly centered on the policy support mechanisms and its relative success (as proved by the overall figures), it is necessary to extend this consensus further, not only to understand how and when additional strides can be made in terms of a "greener" electric grid in Iberia, but also to formalize some of the lessons of this ambitious transition period, hence, allowing an effective transference of knowledge for other countries/EPSs (in earlier stages of this transition). In view of this, an all-embracing study is conducted, that without any reservations, looks at both policy and engineering aspects, this last, translating crucial and practical issues surrounding system flexibility, constitutes another trademark of this work. To encapsulate, it is essential to conduct an holistic analysis of the progress made during the past decade in implementing this "green agenda", while addressing the issues caused by such different traits of VREs versus conventional dispatchable generation, evaluate their current and future impacts, and conceive a feasible approach towards a more flexible and robust Iberian power system, based on an extensive literature review, that does not neglect the energy security pillar.

2.2 The power energy mix in Iberia: characterization and trends in installed capacity and annual energy mix

Along with other EU countries, Spain and Portugal are among the forerunners in embracing the task of adding significant amounts of renewable energy to decarbonize electricity generation. Hence, validating the integration of RES as a prime component of any effective mitigation policy, to be implemented by major emitting economies [77]. As Figure 2.1 illustrates, the renewable installed capacity in Peninsular Spain has been growing quickly. Data from Red Eléctrica de España (REE) [91] shows that the ratio of renewable installed capacity to the total installed capacity, $renewablescap_{ratio} = \frac{renewables_{installedcap}}{total_{installedcap}}$, surpassed 60% in 2021 when in 2008, the ratio was close to 42%. Over these 13 years, Spain's installed wind, solar thermal, and solar PV capacities recorded growths of approximately 2-, 25-, and 200-fold, respectively. Moreover, since 2019, the wind installed capacity ($\sim$27 GW in 2020) has surpassed the (dispatchable) combined cycle capacity (24.5 GW) as the largest installed technology.

In addition to the outstanding progress in renewable integration, examining the available dispatchable installed capacity is relevant. A clear contrast emerges from the recorded

growth seen with renewables, where the plummeting share of non-renewables installed capacity is also explained by the 70% and 15% declines in coal-fired generation and other thermal non-renewable/cogeneration, respectively.

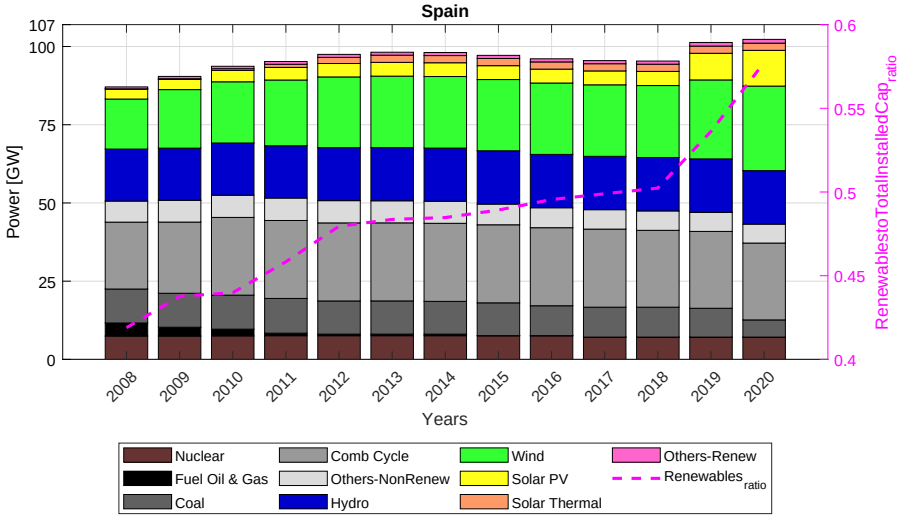


Figure 2.1: Peninsular Spain: Evolution of the installed capacity in GW. Bar plot of the main generation technologies share (y-right) and the $renewablescap_{ratio}$ (dashed pink y-left).

According to REE, the fruits of this targeted effort are steadily gaining momentum in Spain. In 2021, renewables will achieve their best record, producing nearly 47% of all generations in Peninsular Spain [9.9% Year-over-Year (YoY) gain], as highlighted in Figure 2.2. In Portugal, the weight of renewables in the annual generation mix is normally even higher, as shown later. The annual generation mix of 2021 is compared with that of 2016 (five years prior) in Figure 2.2, underlining the steadiness of the nuclear energy share, a steep decline in coal generation, a significant growth in solar PV generation, and a smaller decline in terms of the cycle gas turbine (CCGT) generation, which in 2019 represented a share of 1/5. Additionally, wind power has become the main source of electricity generation, accounting for approximately 24% of the total yearly generation, constituting another noteworthy occurrence, given that nuclear energy has been the main power generation source in Peninsular Spain for many years.
Moving on to the Portuguese Peninsula capacity, data from the Portuguese Directorate-General for Energy and Geology (DGEG) [92], shown in Figure 2.3, reveal a steady rise in the $renewablescap_{ratio}$, exceeding 65% in 2021, representing a growth of 15% compared with 2010. This results from an ambitious energy transition strategy that started with an undifferentiated long-term feed-in tariff (FiT) scheme to all RESs established in Decree Law 189/88. A jump-start was then followed through the years by a series of other national policies in coordination with the evolving European targets, including the "National Energy Strategy", approved in 2005, which, together with electricity market liberalization, fostered an environment conducive to the rapid adoption of RESs [93]. Portugal has con-

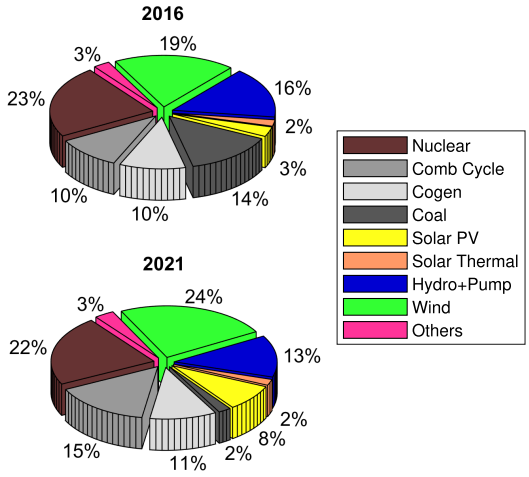


Figure 2.2: Peninsular Spain: Annual generation mix share by technology: 2016 vs 2021 (comparison).

tinued its past trend of investing in hydropower capacity while adding onshore wind generation capacity has gained great momentum. Subsequently, and in a similar way to what was observed in Spain, in the same 13-year span, despite the late start, solar PV was the technology with the greatest increase (17-fold), accounting for roughly 1.1 GW of renewable installed capacity. At the same time, wind generation has grown by almost 80% since 2008, reaching 5.5 GW, thus comfortably surpassing natural gas (the largest fossil fuel-based dispatchable generation), but still below the hydropower installed capacity, that together with mini-hydro surpasses the value of 7 GW. Paired with this rapid integration of renewables, the sinking share of non-renewables installed capacity is also explained by the 70% and 64% declines in the "Others-Non Renewables" share (see Figure 2.3), which include the closure of fuel-oil, diesel and propane plants, while most of the coal capacity remained online up until 2021, the capacity factor of coal plants in Portugal has followed the accentuated declining trend seen in Spain over the last couple of years; this is highlighted in Figure 2.4, where the share of coal generation in the annual mix bottomed out at 1% in 2. In contrast, in 2016, it represented a share of approximately 26%.

As expected from the installed capacity, in these five years, the two main renewable technologies (wind and hydro) consolidated their positions as the largest (single) contributors to the Portuguese annual mix. With wind generation taking first place in 2021, with a share of 28% (~13 TWh), hydropower generation (including mini hydro) is only one pair, given that the annual pumping-storage energy is added. By virtue of a more decentralized policy, that for the first decade and a half of this century was mainly centered on residential installations, the maturing solar PV generation has been slowly but steadily gaining ground to achieve a larger share of the mix. Doubling its energy generation in these five years, corresponding to a 4-fold increase in its annual share. In the case of biomass generation, in the same period, the growth has been more gradual in both nominal and percentage share terms. However, this consolidates its role in the renewables

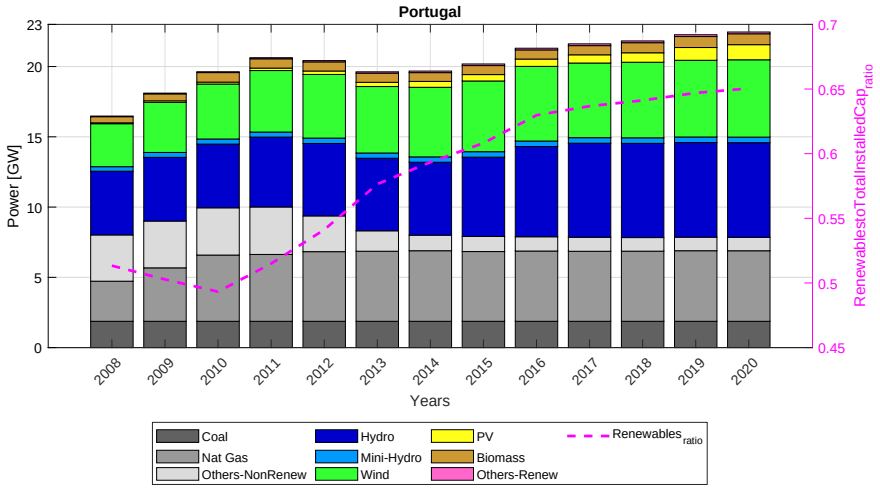


Figure 2.3: Peninsular Portugal: Evolution of the installed capacity in GW. Bar plot of the main generation technologies share (y-right) and the $renewablescap_{ratio}$ (dashed pink y-left).

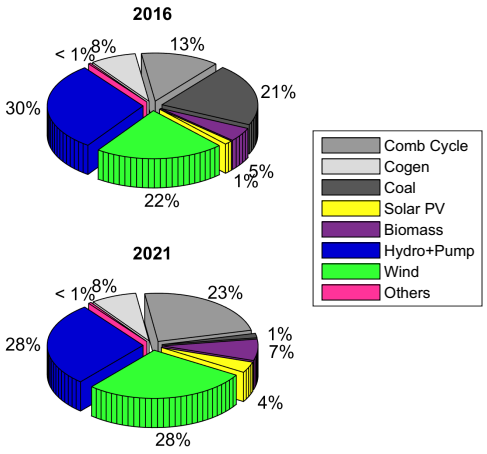


Figure 2.4: Peninsular Portugal: Annual generation mix share by technology: 2016 vs 2021 (comparison).

share of the mix (7%). Hence, virtually matches the fossil fuel-based cogeneration share of 8% (mainly natural gas), which, in practice, remained the same in this half-decade (despite representing a lower nominal value). To compensate for the significant decline in the coal and other high-pollutant thermal technologies, CCGT share has rised to nearly a quarter (∼11 TWh) in 2021 while in 2016, it represented a share south of 1/7 (∼7 TWh).

The following subsections offer a deep dive into the relevant dispatchable capacity for both countries, underlining some of the policy changes and current events that shape the future of the energy mix in Iberia.

2.2.1 An insight into the thermal dispatchable capacity in Peninsular Spain: policies, regulation, and market environment

Following the closure of the most environmentally harmful fuels, oil-fired generation (which was already at residual levels during the 2010s). The specific case of coal-fired generation disassembly in Spain is very interesting, given that until 2018, it was recurrently responsible for annual shares of more than one-eighth of the total annual energy mix. While accordingly to the REE, in the subsequent two years, the annual share dropped to reduced levels of 4.9% and 2%, respectively. Thus constituting a good contemporary example of how often political/regulatory decisions regarding the energy sector can fasten the required change towards a "greener" generation portfolio, but at the same time, they must benefit from a larger consensus, especially when it involves the security of supply at tolerable prices.

Even though coal generation was authorized in Spain until 2030, a harsh regulatory environment, including hefty EU carbon permits, costly investments in filtering, and carbon capture technologies, required to continuously comply with GHGs emission limits, has resulted in higher marginal costs for coal-fired plants. In contrast, natural gas, a less-polluting energy source, exhibited increased price competitiveness in 2019 and 2020; this, coupled with the lack of proper standby payment mechanisms, has accentuated these plants' poor economic market viability compared to CCGT plants and (often remunerated) RESs; this is ultimately proven by the residual number of (annual) operating hours and the smallest share (slice) in the 2021 pie chart shown in Figure 2.2. Therefore, it was not a surprise to see big market players, like Iberdrola, Endesa, EDP España, Naturgy, and Viesgo, with their commitments to the green transition, to decide to abandon this high-intensity GHG emission generation source.

This has culminated in the operational shutdown of seven coal thermal power plants in 2020 alone, i.e., Meirama (Coruña), Narcea (Asturias), La Robla and Compostilla II (León), Andorra (Teruel), Puente Nuevo (Córdoba) and Velilla (Palencia), which is certain one of the biggest, if not the biggest, halving of dispatchable installed capacity in the system (over 4600 MW of gross installed capacity). Full dismantling of these units may take up to five years. While the coal plants of As Pontes (Coruña), Carboneras (Almería), Lada (Asturias) and Los Barrios (Cádiz), were effectively halted at the time (2020), still waiting for a closure authorization from the Spanish ministry for the ecological transition [94–96], but still available to produce; this meant that in the second half of 2020 and for most of 2021, only three coalfired power plants were still effectively online in Peninsular Spain, namely Carboneras/Litoral (Almería) from Endesa, Aboño 2 and Soto de Ribera 3 (Asturias) both from EDP, given their backup supply role in a region with large industry, as well as the considerable investments (14.7M€) to meet revised environmental standards and improve its efficiency [97]. Recognizing the specificities of the Balearic Islands EPS, the Es Murterar (Mallorca) coal-fired plant is expected to be the last plant to be shut down in Spain. After nearly 40 years of service, in December 2021, Endesa decided to close its Carboneras coal plant, and the dismantling process involved the

installation of roughly 1200 MW of PV capacity to replace 1159 MW of conventional thermal capacity [98]. Further, closures are expected to occur in upcoming years; however, as of Q4 2022, the four coal plants listed in Table 2.1 are available for power generation in Peninsular Spain [99–102].

Table 2.1: Coal-fired generation: Gross installed capacity in Peninsular Spain (available in 2022).

Power Plant	Technology (Manufacturer)	Gross Installed Capacity [MWe]
Aboño 2 (Asturias)	mono-bloc unit 4-cylinder turbine 19 kV (Brown Boveri)	556
Soto de Ribera 3 (Asturias)	mono-bloc unit 2-cylinder turbine 20 kV (Westinghouse)	361
Los Barrios (Cádiz)	mono-bloc unit 20 kV (General Electric)	568
As Pontes (Coruña)	Four single-bloc units 18 kV (Mitsubishi/ Westinghouse)	1468 (369+2*366+367)
	Total Coal	2953

However, an unexpected chain of events that took place in the last two quarters of 2021 and early 2022, including the easing of COVID-19 restrictions, fiscal and monetary stimulus, a steady recovery of the demand side, worldwide supply chain disruptions, a slow ramp-up of energy production to pre-pandemic levels, and, as of late February 2022, the war in Ukraine, have all contributed to an unseen (in decades) inflationary environment in Europe. Energy-related commodities, particularly oil, natural gas, and electricity, are among the chief catalysts for the skyrocketing prices in Europe. The extraordinary rise in EU wholesale natural gas prices in 2021 resulted from a perfect storm of supply limitations and a considerable YoY rebound in demand [103]. For instance, the European Commission reported that in the fourth quarter of 2021, TTF spot prices increased, recording high levels of volatility, temporarily reaching an unprecedented value of 183 €/MWh, to then settle around 70 €/MWh at the end of Q4 [104]. This pressure affecting gas prices was further exacerbated at the beginning of 2022, with the war in Ukraine leading to a rebound in prices to approximately 100 €/MWh [105]. Given the prominence of natural gas in terms of the dispatchable electricity generation portfolio in the EU, this adverse context marked a turning point in the relative competitiveness of natural gas. It triggered a resurgence of coal and lignite generation in the old continent, which, according to the quarterly report on European electricity markets (Q4 2021) [106], increased by 21% YoY, representing a growth of 68 TWh, despite increasing coal and carbon prices. In contrast, less GHG-intensive gas generation decreased by 16 TWh, translating to a 3% decline. The sanctions applied to Russia and the extreme dependence of some countries on its gas exports, particularly in Central Europe, including Germany, have left them virtually no other short-term alternative to resort to coal to face harsh winter months.

The same trend was also true for the Spanish electricity market, since despite its share in the annual mix steadying around 2% in 2020 and 2021, the recovery in the demand side and the smaller pool of available coal-fired capacity post 2020, i.e., after the closure of seven of these plants, means that in nominal terms the continuous decline in coal-fired

generation witnessed since 2017, when the annual value stood around 42 TWh, up until 2020, when it reached a minimum of 4.7 TWh (∼9-fold drop), was interrupted in 2021, given the ∼142 GWh increase in coal generation in a YoY basis. In line with the cited European reports, this (annual) resurgence of coal generation in Spain is essentially explained by several bottlenecks involving fossil fuels, prompting record prices. The Iberian electricity market is no exception, with day-ahead prices regularly exceeding the 180 €/MWh mark, thus bringing coal back to a profitable window even in the current scenario. The 2021 inflection regarding coal generation was mostly explained by the significant YoY monthly percentage variations in the last three months of 2021 (Q4), where gains of approximately 125 %, 72 %, and 225% were recorded, according to REE. Faced with this prospect in the latter part of 2021, power companies reverted their policies and decided to resupply their coal stock. For instance, Endesa decided to restock 80 thousand tons of coal for its As Pontes plant (the largest in Spain) [107]. Early data from REE show that this recovery trend of coal generation will spill over to 2022. Still, regardless of this more favorable environment for coal, the Spanish government has ruled out an extension of older and more polluting thermal units, thus reinforcing its commitment to the large-scale penetration of renewables [108].

Following the evolution of coal-fired power generation, analyzing aspects of other and more modern thermal units in Peninsular Spain is important. In this regard, several dozen conventional cycle units (gradually abandoned) and, more importantly, combined cycle units with installed capacities close to 25 GW have gained crucial importance in ensuring the stability of energy systems dominated by variable and intermittent renewables. By means of an additional/second (Rankine) cycle and the respective steam turbine, these units are more efficient than conventional units in salvaging the waste heat produced by the first (Brayton) cycle (gas turbine+boiler), exceeding the 55% threshold with lower CO_2 emissions, requiring less cooling water, and operating with a greater amplitude of output partial loads [80, 109, 110].

With the commencement of a nuclear program in the 60s, it was not until the mid-80s, when most of the current active nuclear power plants began to operate, that nuclear energy became an integral power source for the Spanish generation portfolio. Comprised by seven reactors, spread over five different sites, totaling ∼7398.7 MW (Table 2.2) of gross installed capacity, it constitutes a vital baseload power source [111]. This source is responsible for generating yearly figures of approximately 55 to 60 TWh of electric energy, representing by far the highest capacity factor of any of the available technologies in the mix (between 84.8 and 92.5%), operating north of 91% of the hours during a full year (the highest amount in the EPS) [112]. In addition, for many years, nuclear energy has not only been responsible for a sizable share of emission-free generation but also provided a well-bounded and relatively inexpensive energy source, given its low marginal costs.

The planning made in terms of the location, grid connection and sizing of the five different power plants is also very important, and in general the chosen locations, presented in Table 2.2, highlight a proximity towards large population centers, with a solid installed ca-

pacity in one hand. While on the other, still leaving sufficient "room" for the exploration of other technologies, without overburdening the transmission lines. For example, the Cofrentes nuclear plant has sufficient capacity to serve the entire residential load in the Valencian community, perfectly fulfilling the role of a base-load type of dispatchable generation. Moreover, after the decommission of the Santa María de Garoña nuclear power plant, the Cofrentes nuclear power plant is the exception in terms of reactor architecture, with a boiling water reactor (BWR). The remainder are equipped with the most common type of light-water nuclear reactor, a pressurized water reactor (PWR). PWRs differ from the BWRs architecture in that the secondary coolant loop is separate from the pressurized primary loop of water (outside the reactor core), which drives the flow of steam to the turbine [113]. Nevertheless, the aging of generation II nuclear plants, with most approaching the forty-year mark, is raising awareness about the lifespan and safety of the Spanish nuclear infrastructure.

Table 2.2: Nuclear power generation: gross installed capacity in Peninsular Spain (active in 2022).

Power Plant	Fuel (Reactor type)	Gross Installed Capacity [MWe]
Almaraz I and II (Cáceres)	Nuclear (PWR)	2093.9 (1049.4+1044.5)
Trillo (Guadalajara)	Nuclear (PWR)	1066.0
Ascó I and II (Tarragona)	Nuclear (PWR)	2059.7 (1032.5+1027.2)
Vandellós II (Tarragona)	Nuclear (PWR)	1087.1
Cofrentes (Valencia)	Nuclear (BWR)	1092.0
Total Nuclear		7398.7

2.2.2 An insight into the thermal dispatchable capacity in Peninsular Portugal: policies, regulation, and market environment

Moving on the Portuguese side, its smaller scale, favorable climate conditions, inherited long-term bet on hydropower capacity, and aggressive RES adoption policy have translated into a faster push to abandon heavily intensive GHG. A push started in the early 2010s, with the deactivation of once system-relevant fuel oil-fired plants, namely in Carregado and Setubal (the largest power station until 1985, $\sim$1 GW). The abandonment of coal generation was the next target of Portuguese government enforcement, with 2023 as the original final date to phase out this technology. The closure of the two remaining coal power plants, Sines and Pego, was actually anticipated for 2021. The first was particularly relevant because the Sines power plant was the largest thermoelectric facility in Portugal, with a gross installed capacity of 1256 MW, which was responsible for a substantial part of the Portuguese load during the 90s. Still, according to the owner, the EDP group, the company's strategic alignment with decarbonization, and the slim profitability margin of the four-unit power station have driven this decision [114]. In the case of Pego's power station, the property of some of the biggest power players in Europe, Endesa and Engie, effectively retired its two identical 314 MW generation units in November 2021 with the end of the existing power purchase (backup) contract that expired at the end of that month.

Hence, it marks a historical event in the EPS at the end of coal-fired power generation.

From the remaining non-renewable installed capacity in Peninsular Portugal, natural gas reigns as the primary fuel source, accounting for close to 64% of the total non-renewable capacity (7850 MW), particularly with more modern CCGT plants, mostly owned by EDP and Trustenergy, that total around 4000 MW of installed capacity. The main thermal generation plants of CCGT are listed in Table 2.3.

Table 2.3: Main CCGT thermal generation: gross installed capacity in Peninsular Portugal (available in 2022).

Power Plant	Commissioning Year	Fuel/ # Groups	Gross Installed Capacity [MWe]
Tapada do Outeiro (Gondomar)	1999	Natural Gas/ 3	990
Lares (Figueira da Foz)	2009	Natural Gas/ 2	885
Pego (Abrantes)	2010	Natural Gas/ 2	830
Ribatejo (Carregado)	2004	Natural Gas/ 3	1176
Total CCGT		3881	

In the matter of cogeneration, a more efficient and cost-effective way of simultaneously generating electricity and heat, natural gas is also the main fuel source; however, unlike other thermal capacities, it is scaled down and spread across the country. Commonly, these are connected to specific industries that require or self-produce heat or sewage treatment plants. For instance, the Fisigen cogeneration plant (Barreiro), commissioned in 2010, with its two groups, uses gas turbines to generate 24 MWe and 121 MW of thermal capacity. Considering the stable capacity that is expected to remain online for at least two decades, natural gas-fired generation plays and will continue to play a crucial role for system flexibility in the Portuguese mix, which means it can respond to variations in hydroelectric and wind production (the two largest renewable sources). However, in the same manner as highlighted in the Spanish case, unprecedented commodity costs meant that the competitiveness of this technology would be seriously compromised in 2021 and 2022; this harms the economy because electricity prices are a major driver of inflation [115], and is particularly relevant in countries where energy poverty metrics are above the EU average [116].

Another substantial difference from Spain is that Portugal has no nuclear plants. In contrast, it has relatively more hydropower capacity than Spain, with a relevant pumped storage capacity, which is crucial for better electricity production management [117].

On a side note, it is important to note that, despite being less commonly mentioned, hydropower, like virtually all renewable sources, exhibits some variability. It depends on water availability, which has intra-annual and inter-annual seasonal variations depending on weather patterns and climate conditions. However, it is much more predictable and controllable than other renewable sources. Typically, large hydropower plants come with sizable reservoirs (especially when equipped with pumped storage capability) and can be considered to some extent a dispatchable energy source. In contrast, small hydropower plants (run-of-river) act more like VREs, given their almost exclusive dependence on river

flow and its tributaries, coupled with a virtually negligible short-term reservoir capacity, resulting in greater output fluctuations.

2.3 A comprehensive strategy for the future grid: options to increase system flexibility

Most VRE characteristics pose significant practical issues regarding power flow, grid stability, energy quality, system balance, and market price formation [118]. Contrary to public discourse, in countries such as Spain and Portugal, even in the current scenario, where VREs already represent a respectful share of the mix, operational issues and market distortions are good evidence of the negative impacts of intermittency on the grid [119,120]. The aging of current renewable farms–their repowering–is another question that policymakers will face in the coming decades. Consequently, it is another good reason why, it is crucial to examine an all-around approach of the available options to reinforce the system flexibility, while preserving the security of supply (the subject of section 2.3. To highlight this fact, a recent study has shown that by 2025, nearly half of the wind installed capacity (∼10 GW) in Spain, is reaching 20 years of age, hence, coming to the end of its operational lifetime, a situation that is not much different in Portugal, and other EU countries [121]. Moreover, when assessing how the repowering of wind energy in Spain takes place without the previous FiT scheme, it was found that an old wind farm with greater wind resources (than those available in vacant sites) and repowering with new wind turbine technologies is still viable in the long term, while a recommendation was made to analyze both technical and retributive requirements to avoid market distortions, as occurred in the past [122].
To address these issues, it is necessary to increase the flexibility of the power system as a whole; that is, to increase the ability of the system to cope with the numerous challenges posed by an energy mix with a high share of renewables. Typically, this process is a not a "one size fits all" or a single-based type of approach; on the contrary, it needs to be tailored to the needs of each power system and is substantiated on several action vectors, as represented in the numbered blocks in Figure 2.5. The first two action vectors target flexible supply side options, involving the diversification of the renewable portfolio with different and complementary technologies (subsection 2.3.3), which help mitigate the degrees of intermittence/seasonality of the current portfolio, while simultaneously preserving, upgrading, or integrating a relevant portion of thermal dispatchable generation (subsection 2.3.5). The following action vector targets flexible demand-side options, compromising the wider adoption of demand-side management schemes for individual consumers and community-level schemes (subsection 2.3.4). Reinforcing interconnections and other grid upgrades can ensure supply security of supply (subsection 2.3.2). Finally, the maturation of a wider range of energy storage options (subsection 2.3.1) is a game changer, both on a smaller scale when thinking of prosumers and for market participants with VRE generations, which, by better managing the dispatch of the flows to the storage or the grid, can provide system flexibility.

Policy support measures must be revisited to advance these action vectors. As mentioned earlier, FiTs are crucial for the faster integration of renewables in Iberia [123]. However, the prior (generous) subsidized remuneration regimes for injected power were tightened after several years. In Spain, a more cost-driven fixed payment remuneration scheme (Royal Decree 9/2013) was enacted that replaced the old and more advantageous regime (Royal Decree 661/2007) for both past (retroactively) and future RESs settlements [124]. As mentioned in section 2.2, in the Portuguese case, a legal scheme supporting the adoption of renewables has been in force since 1988, and has been amended a few times throughout the years, including in 2005, when FiTs were reduced, but had no impact on producers operating under past assumptions. Historically, this meant that renewable producers other than large hydro plants (mostly wind farms) and cogeneration were labeled special regime production (SRP) [125], benefiting from guaranteed purchases and stable remuneration (above wholesale market electricity prices) for 15 years or up to the first 33 GWh produced per MW installed [126]. Faced with a debt crisis, things changed in 2012, and one of the impositions was to halt the growing tariff debt; thus, the FiT support scheme was abandoned [127], except for new (occasional) capacity auctions or grants that had yet to expire (Decree 215/2012).

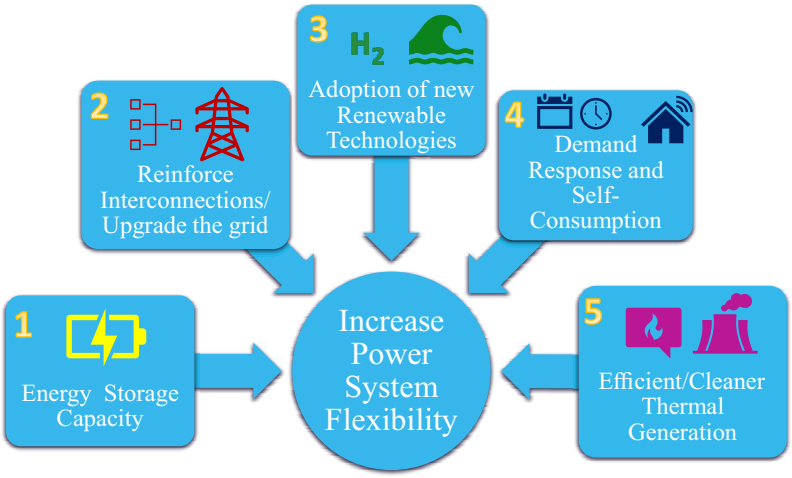


Figure 2.5: Main vectors of action to increase power system flexibility.

The levelized cost of electricity (LCOE) is often considered a sufficient measure to assess the overall competitiveness of different generating technologies; it represents the average revenue per unit of electricity generated, which is required to recover the costs of building and operating a generating plant. Additionally, for storage technologies, a similar metric, the levelized cost of storage (LCOS), the average revenue per unit of discharged electricity, is required to recover the most important capital costs.

In its "Annual Energy Outlook 2022" report, the U.S. Energy Information Administration (EIA) estimated these costs for a variety of technologies, considering a 30-year cost recovery period, a nominal after-tax weighted average cost of capital with reference values from 2021, and assuming capacity factors at the high end of the spectrum [128]. CCGT with its flexibility is evaluated in baseload operation (high-capacity factor), whereas a combus-

tion turbine was considered for peak load cycles (as were the batteries). The main results are compiled in Table 2.4, where the relevant technologies, that is, with existing or potential applications in Iberia, are characterized as either dispatchable, VRE, which means that they are, to different extents, dependent on resource availability, and, lastly, storage capacity, which are typically high-energy-density units.

Table 2.4: Estimated unweighted LCOE and LCOS energy resources entering service in 2027 [128].

Plant Technology	Type	Capacity factor (%)	LCOE or LCOS [€/MWh]
Hydroelectric	Renewable/Dispatchable	54	64.27
Solar (standalone)	VRE	29	36.49
Wind (onshore)	VRE	41	40.23
Wind (offshore)	VRE	44	136.51
Biomass	Renewable/Dispatchable	83	90.17
Ultra-supercritical coal	Thermal/Baseload	85	82.61
Combined cycle	Thermal/Baseload	87	39.94
Advanced nuclear	Thermal/Baseload	90	88.24
Batteries	Storage capacity/Peak-load	10	128.55
Combustion Turbine	Thermal/Peak-load	10	117.86

As expected, the estimated LCOE costs reflect the maturity achieved by solar PV and onshore wind technologies, the lowest among the RESs. Regarding dispatchable technologies, the CCGT-fired generation, in which fuel costs represent the largest share of the LCOE [129], appeared to be the most competitive. Based on 2021 data, this result is partially tied to the pandemic's impacts on commodity markets. Indeed, unexpected geopolitical events, such as the current sanctions applied to Russia, have hindered natural gas price competitiveness, which, in these studies, may tilt the balance in favor of other technologies. In fact, the LCOEs of all fossil fuel-based energy sources and, to a lesser degree, biomass plants are sensitive to fuel prices [130]. Modern nuclear plants may require longer recovery periods for a proper comparison. Battery storage and combustion turbines, with their low-capacity factors for peak load following, relatively high capital costs in the case of the first, and high fuel costs in the case of the second, have the highest levelized costs. They are only surpassed by the offshore wind generation, which is been steadily decreasing its LCOE, and accordingly to Kost et al., these are situated in a lower range (72.3-121.3€/MWhh) for the German market [131].

Conversely, the sharp increase in CO_2 certificate prices in the EU and the assumed fuel costs and capacity factors (more favorable in the US environment) explain the significantly higher LCOE for the thermal energy sources in this report versus the EIA estimates. Though, importantly it confirms the overall trend, i.e., CCGT is the most competitive among the thermal dispatchable technologies (77.9-130.6 €/MWh) and solar PV (utility scale) and onshore wind are the best ranked sources in terms of LCOE. To this extent, when deciding upon flexibility options, LCOE values are a good analytical piece of the bigger puzzle.

2.3.1 Energy storage systems

The ambitious EU plan to reduce GHG emissions by at least 80% by 2050 means an aggressive push for integrating non-dispatchable generation will continue. Integrating energy storage systems (ESSs) is a critical component of future grids to avoid compromising the reliability and cost-effective operation of grids [132, 133]. ESS adoption is expected to increase exponentially, reaching a capacity of approximately 850 GWh by 2040 [134]. It provides the desired flexibility by absorbing excess power during off-peak load periods, which allows it to shave off-peak loads or fill valleys [135]. Consequently, the benefits of ESSs include output smoothing, time shifting, frequency regulation, oscillation damping, mitigation of transmission congestion, reserve role, and even a decrease in LCOE when the storage duration is sufficient [119, 136].

The same is valid both at a grid-scale level, where **pumped hydroelectric storage** (PHS) is the only viable option for queuing and storing large amounts of excess generation for a long time [137], and at a decentralized level, where prosumers want to maximize their self-consumption. As mentioned, both countries have tightened their remuneration schemes, including micro- and mini-generations. In the Portuguese case, with the introduction of Decree-Law 153/2014, which replaced all the prior regimes moving forward, it established ”generation units for self-consumption” (known as UPACs in Portuguese) as a central framework to incentivize the adoption of small RES and storage solutions (in a way that balances demand and production as much as possible/intended). It does so with more reasonable market-based settlements for the injected power surplus [138].

In this day and age, a vast selection of ESS technologies has been and continues to be developed. These vary in terms of working principle, efficiency, maturity, scalability, feasibility, and scope of application. The most relevant in terms of the potential for supporting wide VRE integration is shown in Figure 2.6. ESSs are mostly characterized by their capacity (also known as energy rating) in [Wh] or [Ah] as a measure of the total energy they can store, and their power capacity (or rating) as a measure of the (rated) amount of power they can deliver in [W]. Additional considerations include the storage period (discharge time at rated power), cycle life, energy and power densities, i.e., the value for these variables per unit of mass or volume, among others.

As of 2017, PHS, a mechanical technology based on the gravitational potential energy between sequential reservoirs, is the most mature ESS, accounting for the overwhelming majority of the global energy storage capacity, with 153 GW of installed peak power capacity from a total of 159 GW [135]; this is explained by the diluted capital costs, greatest energy and power rating of any ESSs (the largest PHS stations surpass 2,500 MW of installed capacity, with most of these facilities located in China, Brazil and the United States.), long storage period, and infrastructure lifespan [132, 139]. Usually, a PHS plant is characterized by its power capacity in the pumping (upstream) mode, mainly dependent on the head elevation and flow. Besides, as seen earlier, both Portugal and Spain extensively use this solution, which is the most tailored solution for bulk system energy storage and consequently, power demand management (enabling time-shifting excess en-

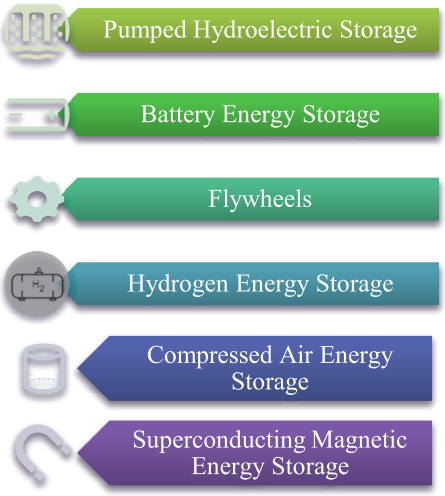


Figure 2.6: Most adequate energy storage technologies for VRE integration support.

ergy, grid stability and a reserve role), which can partially offset the SRP's variability. Continuing the trend, Iberdrola announced the opening of its 1.1 GW Tâmega giga battery (three hydropower plants), a huge investment resulting in one of Europe's largest pumped storage plants [140]. Even before these new additions, the PHS capacity was already unmatched by any other energy storage system in Iberia and will remain so for the foreseeable future. Its load-balancing role is therefore pivotal in minimizing the intermittency of renewables and simultaneously reducing the use of expensive peak/reserve thermal plants, as highlighted in Figure 2.7.

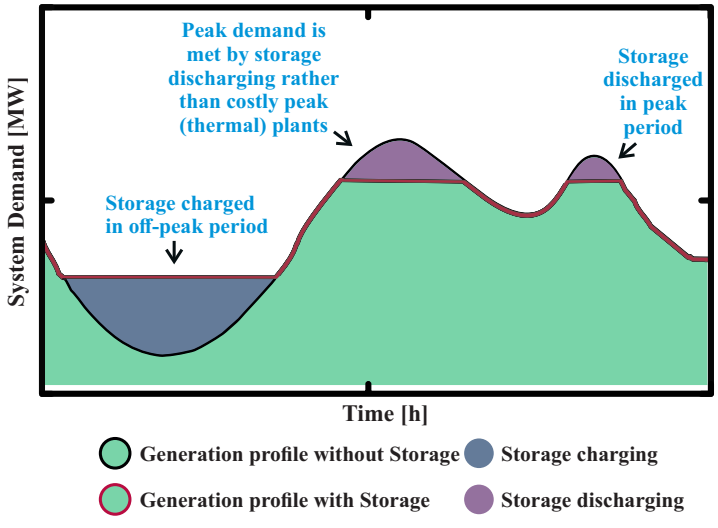


Figure 2.7: Effect of large-scale penetration of ESSs: breakdown of electricity generation.

Flywheel energy storage systems are another mechanical type of ESS technology that stores energy as a spinning mass. The flywheel slows down, converting kinetic energy to electricity, whereas for charging, the opposite occurs, and the spinning mass increases its velocity. The fast response time, high efficiency, and high power density (W/kg) indicate

that it is particularly tailored for frequency balance. However, capital costs remain too high to expect substantial adoption [119].

Battery energy-storage systems, which are primarily based on electrochemical processes, are another well-known solution. Its modularity, rapid response, and mid-range energy density (Wh/kg) render it a viable technical solution for achieving many of the (ESS) benefits, as described earlier. However, capital costs, degradation, and challenging operational conditions can tilt the pendulum in favor of other technologies. The most common choices are nickel, cadmium, and lithium ions (Li-ions). Freshly inaugurated in 2022, the floating PV solar farm in Alqueva, Portugal, the largest in Europe besides the 5 MW of installed generation capacity, will also have a storage capacity of approximately 2 MWh, guaranteed by a Li-ion battery bank [141]. Indeed, the cost of small-scale Li-ion batteries has fallen by over 60% since late 2014, which, coupled with some financial support, means that close to 40% of small-scale solar PV systems in Germany have been installed with some type of battery system in the last few years [142]. Therefore, this type of ESS is becoming increasingly mature. Flow batteries are a promising grid-scale storage solution owing to their increased energy rating, scalability, long cycle life, and reduced environmental burden.

As one the sunniest regions in the EU, with an yearly average of sunshine hours above 2800h ($\sim$7.6h/day) in its southern half [143], with a fast-growing PV generation capacity, battery ESSs are therefore well-positioned, as a way to mitigate the variable nature of this RES (effectively absorbing some of the energy surplus from the peak period production in what is essentially a bell curve pattern).

High power density storage solutions, such as supercapacitors, differ in their energy storage mechanism. Unlike conventional battery energy-storage systems, which rely on chemical reactions, supercapacitors store energy electrostatically. This allows them to provide quick bursts of electric energy, enabling much faster discharge and charge cycles. This trait, combined with a long life cycle (number of cycles to end-of-life), makes them a promising candidate for inertia emulation and grid stability [144].

Like PHS and flywheels, **compressed air energy storage** is also an old technology with a mechanical/thermal principle. Electric energy is converted from and to thermal energy through the expansion and compression of air stored in reservoirs, sharing some similarities with conventional gas turbines. It has very good properties in terms of capacity, storage period (only behind PHS), and a very low impact on the environment. The geological structure required for storage (reservoirs) hampers easy and inexpensive integration.

Hydrogen plays a key role in decarbonizing the energy mix, by assuming the role of a ”green” energy carrier that can be used to store, transport, and deliver energy produced from RESs. In this regard, hydrogen energy storage, is a chemical type of storage where electrical power is converted into hydrogen (through electrolysis, gasifying biomass, or other fossil fuels) for storage in underground formations or steel containers. Subsequently, the stored hydrogen can be transported to where power is needed, and used as fuel in fuel cells or adequate combustion engines to produce electricity.

Publicly, it is seen as a great option by Iberian policy makers. Therefore, both countries have heavily betted on developing this infrastructure and technology in their EU backed recovery and resilience plans. Nevertheless, high capital costs and low round-trip efficiency mean a long way to go before a wider adoption at the grid level occurs [145].
A more novel storage technology is the **superconducting magnetic energy storage**, a purely electrical system, which as the name suggests, makes use of a superconducting coil, where a DC current passes, creating a powerful magnetic field, which when electricity is needed can be converted back to AC electric power (the only conversion), resulting in efficiencies over 90%; this implies that it has tailored properties to mitigate variability in renewables. Still, the involved temperatures and the need for a costly refrigeration system are the main drawbacks that translate into a very high initial cost (as was the case with many ESS). Still, it is not so unexpected in this case, given how undeveloped the technology is [119].
The **electric vehicle** (EV) is another frequently mentioned form of storage flexibility connected to load shifting. Impacting the EPS in a number of ways that will be the subject of a separate subsection. Furthermore, an ESS involving thermal processes is also a viable option, and is explored in greater detail in subsection 2.3.5.4.

2.3.2 Interconnections

Although the future grid will be shaped by more disaggregated and decentralized generation, the rapid incorporation of home and industrial PV-type systems, with or without storage solutions, is a trend. Which, in and of itself poses well-studied issues at the distribution level, including voltage unbalances, deteriorated power quality, potential for reverse power flow and others [146]; this does not mean continuous examination of the transmission level is less important. In contrast, the ability to reliably transport large amounts of electrical energy, using a complex system of typical overhead wiring, under high and extra-high voltages (>100 kV), from power plants/farms to the consumption centers, over long distances, is not given. Particularly, a profound transformation occurs at the end of generation and consumption. Requires reassessment regarding capacity usage and congestion, power flows, losses and energy quality, redundancies, infrastructure lifespan, and expansion.
Therefore, a strong transmission infrastructure is a source of system flexibility and reliability that provides a wide geographical network of connections, enabling access to a diversified generation portfolio, fostering competition between market players through expanded market access, promoting the flow of low-cost electricity to high-price markets [147], and reducing the chances of energy curtailment in substantial generation surplus occurrences (impacting a single node) [148]. Thus, it plays a crucial role in achieving better intra- and inter-market integration in a foreseeing scenario where very high levels of renewable generation are expected.
Regarding interconnections – the capacity to allow cross-border energy exchanges – the Iberian Peninsula still lags far behind its European peers. Considering the specificity that

in the Iberian market (established region), only Spain borders other non-market countries, the reality shows a weak interconnection ratio; that is, it falls short of having the recommended EU 10% threshold of the installed generation capacity in cross-border transport capacity; this (electricity) import/export capacity limitation is the reason why Iberia is considered an "energy island", which was vital for the European Commission to allow the Iberian countries to limit (by decree) the gas prices for electricity generation, as of Q2 2022.

As such, to fulfill the ultimate goal of the EU policy makers, which is to achieve the still elusive "Internal Electricity Market in Europe", is necessary to strengthen the interconnections with France, which for a long period was not seen as a priority, especially by the French part [149]. According to the REE, the aim is to reach at least 4,000 MW of interconnection capacity, as such, and despite the geographical challenges, it is essential to continue to invest in transmission infrastructure, with overhead/underground upgrades in the Pyrenees, and with new projects using high-voltage direct current (HVDC) lines, that are shown to improve the transport of RES and the stability of the grid, with reasonable costs when compared to its AC counterparts, a practice that is been widely implemented in China [150]. This commitment includes the projected Spain-France submarine interconnection through the Bay of Biscay using a dual-link HVDC with a two $\times$ 1000 MW capacity. A similar submarine interconnection between Portugal and Morocco was suggested in [151].

While, in the case of the Portuguese-Spanish interconnection, which is made up of ten AC lines, as showed in Figure 2.8 [152], the situation is a little different, since it already satisfies the 10% interconnection ratio.

However, considering that the last two major infrastructure changes were the 400 kV lines in the international Douro region (Lagoaça-Aldeadávila) and interconnection between the Algarve and Andalusia (Tavira-Puebla de Guzman), it is essential to continue to strengthen the interconnections, especially with the long planned second interconnection between Minho and Galicia (in 2021 the commitment was reinforced) and with a second line in the Alqueva-Brovales connection, as highlighted by the dashed line in Figure 2.8. This reduces stability problems during contingencies associated with the maximum phase shift between substations [153].

2.3.3 Other renewable energy sources

The previously discussed shortcomings of the current renewable generation portfolio in Iberia, which is almost limited to onshore wind and solar PV generation technologies, include significant intermittency, a frequent mismatch of the peak production with the peak loads, strong seasonality, a limited window of operation for PV systems, an already high-level technological maturity, geographical constraints in terms of future wind farm locations, and a less-mentioned challenge, which, as pointed, is the upcoming (sizable) task of repowering the existing renewable generation.

By promoting/integrating different renewable technologies into the system, while tak-

Figure 2.8: Interconnections in Iberia: installed and planned locations, voltage levels and total capacity.

ing advantage of different natural resources, in different countries, in a more connected continental grid, is a great approach to tackle status-quo limitations [154]. Accordingly, the most suitable alternative RESs for the Iberian portfolio were selected based not only on the large availability of the endogenous resources (required), being environmentally friendly, the potential capacity for a system-wide grid, that is, for this study, the focus is not on small-scale RES deployments, and on current and projected installments in Iberia. The selected emerging technologies [155] include hydrogen energy systems, ocean energy, concentrated solar power, biomass gasification technology, offshore wind farms, and novel floating wind turbines, as represented in Figure 2.9.

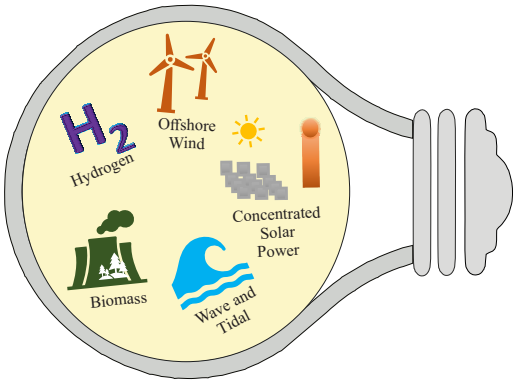


Figure 2.9: Promising RES to complement the existing generation portfolio in Iberia.

2.3.3.1 Hydrogen energy systems

Hydrogen is an energy carrier that can be produced from various sources; however, the vast majority comes from fossil fuels. However, in the past few years, a change in the tide has been noted, and hydrogen technology has become one of the most promising technologies [156]. Proof of that is that global demand for hydrogen is rising fast, and several countries have been deeply invested in implementing support policies for hydrogen deployment [157]. Moreover, it has also been shown that hydrogen production systems can facilitate widespread intermittent renewable deployment [158].

In addition to the important fact that hydrogen can be produced locally, thus reducing the energy dependence of external suppliers, it also has a high specific energy density in comparison with normal fuels, which makes it suitable for transport and industrial purposes, where it complements or even replaces the conventional combustion engine. Hydrogen can also produce electricity through a fuel cell [159]. However, to take advantage, of the "clean-burning fuel" label, it is necessary to power the hydrogen production (electrolysis) with renewables [160]. As the International Energy Agency (IEA) recognizes, this should be a priority for countries with more favorable natural conditions, including the Iberian Peninsula.

Following this trend, the Spanish government has not only launched its ambitious roadmap for hydrogen. Still, it has also positioned the country among the world leaders in hydrogen deployment, with a great focus on its integration with renewables (green H_2). Signs of this commitment, include the presentation of the largest green hydrogen plant for industrial use in Europe by Iberdrola, the commitment from Endesa to install a green hydrogen complex in its coal plant "As Pontes". The inauguration of the first industrial green hydrogen plant in Lloseta, in the Balearic Islands, a first of its kind in Iberia, coupled with a 15 MW PV farm. Portugal follows the same trend, and Sines is the epicenter for this transformation; the former Sines coal complex is being transformed to receive two 100 MW electrolyzers.

2.3.3.2 Wave and tidal energy

With far greater energy density than wind and solar power, ocean energy is an RES with a theoretical global potential of 32 PWh per annum, sufficient to meet the global electricity demand. This figure is much more modest for tidal energy, with estimates of 3 TW of potential power capacity [161]. Europe is the leader in wave and tidal energy, with steadily growing capacities of approximately 12 and 28 MW, respectively. However, it is still far from fulfilling the objective of deploying 100 GW by 2050 [162].

Numerous studies have investigated all aspects of harnessing this source to take advantage of this vast kinetic energy potential, particularly in developing more robust wave energy converters to extract the most out-of-wave energy density (kW per meter of crest length) [163].

Tides result from gravitational pull from the moon and sun, and tidal conversion oper-

ates similarly to how turbines harness wind power. However, water is more powerful because it is denser than air. Wave energy is extracted from wave motion (i.e., the interaction between winds and the water surface). In other words, linear devices can harness mechanical energy, exploring constant oscillation, or even rotating devices. A substantial difference between these RESs is that unlike wind-dependent waves, tides are more predictable and stable, resulting in a steadier (less variable) source of electricity.
The proximity of the power grid to ocean resources in Portugal and Spain represents an important advantage in building necessary infrastructure, which would otherwise be even more costly. Furthermore, both countries have maritime areas located in the upper quarters in terms of both wave and tidal energy potential [164]. Firming its commitment to diversifying the generation portfolio, Spain set plans to deliver up to 60 MW of marine energy by 2030. As of the last years, an agreement was signed to commission the first wave energy power plant in Spain (port Adriano) with an expected capacity of 2 MW. A crucial first step is to verify the feasibility of these sources.

2.3.3.3 Biomass

Biomass is a renewable organic material, and biomass energy can be defined simply as "energy from living things", and is one of the oldest energy sources [165]. Wood, agricultural residue, sewage, and garbage constitute different forms of biomass. Biomass can be transformed into several byproducts through different processes. The most common is direct combustion, which can heat buildings or industrial heat processes or produce electricity through steam turbines. Recapturing some thermal energy and converting it to electric energy in a secondary stage is known as co-generation. As discussed in section 2.2, represents a solid share of the annual generation mix and is expected to reach double digits by the end of the decade. Moreover, there is some seasonality in its collection and variation in type, moisture, and ash content, resulting in some output variability. Nevertheless, the possibility of storage and feedstock regulation (input modulation) means that, unlike wind and solar PV, it is essentially a dispatchable renewable energy source.
Accordingly, serious efforts have been made in the EU to adopt this technology, which has grown by a 3-fold factor from 2005 to 2017 when it reached 32 GW of installed capacity [166]. In the case of Spain alone, the biomass related installed capacity has grown about 10% in the past decade, reaching close to 900 MW plus around 600 MW of renewable waste (50% of the generation from solid urban waste [91], e.g., municipal waste incinerators), thus, totaling $\sim$1.5 GW.
The widespread use of biomass for heat and electricity can significantly contribute to decarbonization. In addition, traditional coal plants can be refurbished to burn biomass for electricity generation, which is an advantage when countries such as Portugal, Spain, and Portugal are decommissioning these plants [165]. Its dispatchable nature results in a high-capacity factor (shown in Table 2.4), complementary to the VRE-based generation mix. Notwithstanding, its relatively high LCOE, derived from its shortcomings in terms of lower energy density and high moisture content, net conversion efficiencies, land require-

ments, fuel availability, and storage [167–171], can question its (present-day) suitability as a great baseload technology for the long-term replacement of traditional highly centralized fossil-fueled power generation sources [172, 173].
Over the last few years, advanced biomass gasification technologies have been capturing the attention of the power industry, given their ability after a set of thermochemical processes to produce a more homogenous byproduct from solid biomass, syngas (high-temperature gaseous fuel), which can then be fed to a boiler that in turn produces high-pressure steam to drive the turbine and produce electricity [174].

2.3.3.4 Offshore wind

Offshore wind energy takes advantage of the higher and more constant wind speeds, generating 20–40% more energy than onshore turbines [175] in relatively shadowed waters and away from the coastline, with no barriers or visual or acoustic impacts. The absence of spatial limitations also implies that turbine sizes can be larger, resulting in higher capacities. However, this comes with a higher capital cost than its more mature onshore counterparts.
Wind turbine and power plant upsizing is an effective way to reduce the LCOE, and the evolution of wind turbine technology, where there are already commercial solutions with 10 MW of rated power and rotor diameters exceeding 160 m, is a testament to successful progress [176]. The wind turbine landscape has historically been dominated by high-speed asynchronous (induction) generators with a gearbox. However, especially in the harsher offshore environment, with stringent maintainability and reliability requirements, low-speed direct-drive permanent magnet synchronous generators are gaining traction due to their higher efficiency and power density. The absence of a gearbox reduces the number of moving parts, thereby decreasing mechanical losses and maintenance needs [177, 178].
An additional encouraging development is the advent of floating offshore wind energy, which, as the name suggests, does not require a fixed structure, offering an advantage in spatial diversification, assembly, and construction costs. More importantly, it opens the door for turbine deployment in deeper waters (depths up to 700m), which increases the wind potential, but with downsize of increasing the transmission costs and associated losses [179, 180].
Regardless of these breakthroughs, the numbers show that offshore supplies less than 0.5% of global power generation, so accelerating its integration is necessary [162], aligning with the offshore wind strategy as defined in the "European Green Deal". Spain, which already ranks in the top five countries with the most installed wind power capacity, yet in nominal terms, is still far behind countries like Germany in terms of offshore wind power, with only approximately 50 MW of installed capacity; a figure that in the Portuguese side should be close to 80 MW [181]. However, things will change, since the Spanish offshore wind roadmap targets 1–3 GW of offshore wind power by 2030. Under the correct framework, this can lead to relevant downward price pressure in the Iberian electricity mar-

ket [179]. Moreover, even with more modern floating offshore wind farms, Spain has some of the best conditions, translating to one of the lowest LCOE. In Portugal, EDP, has been pioneering floating offshore wind generation, in a project called "WindFloat Atlantic",'with a projected capacity of 25 MW.

2.3.3.5 Concentrating solar power

Concentrating solar power (CSP) is a well-studied RES that relies on solar irradiation channeled through mirrors into a receiver that converts solar energy into heat for electricity generation. Due to a series of technological breakthroughs, particularly with new advances in solar thermal storage solutions such as molten salt technology, CSP technology has re-emerged as a capable and adjustable green energy, thus improving power system flexibility [182], and with the potential to be a very price-competitive technology in high potential regions [183]. Another beneficial trait is its lower carbon and water footprints than other energy sources [184].
Coupled with solar plants, molten salt storage is an interesting thermal energy storage system successfully used in Spain. "AndaSol-1" plant was even the first commercial parabolic solar thermal plant in Europe. It works not by exploring waste heat but by concentrating the available solar heat into focal points that warm the molten salts (high boiling points and volumetric heat capacities), which are then stored in an insulating container during off-peak hours. Later, when energy is required, these high-temperature and high-pressure salts are pumped into a steam boiler generator and spun in a turbine that produces electricity.
In the face of these developments, the global cumulative CSP capacity has been growing rapidly, reaching approximately 6.4 GW in 2019, and approximately half of this capacity is being planned or under construction for the upcoming decade. In the case of Iberia, Portugal is behind the curveball in terms of this technology, which has shown great potential for more hot, arid, and isolated regions, such as the southeast of Portugal; however, Spanish regulatory efforts to promote the integration of CSP capacity were fruitful, but it disregarded the required investments to advance CSP technology [185].

2.3.4 Demand response and self-consumption

Demand response (DR) is a term used to describe mechanisms/programs through price signal(s) from retailers, the hourly electricity market, or even a trigger from the grid operator, encouraging the end-user to lower or shift their consumption during peak periods or extraordinary contingencies. In other words, the DR provides an opportunity for consumers to play a significant role in the operation of the electric grid, according to the U.S. Department of Energy (DOE) definition [186]. This, in turn, is very useful in tackling issues ranging from high demand peak-to-average ratios, high generation costs, high GHG emissions, reliability, and congestion in transmission and distribution systems [187,188]. The potential benefits of DR programs are why energy flexibility from demand response

has been a popular research topic for many years [189]. According to the IEA, the last couple of years have brought positive demand-response regulations, ultimately unlocking some of the legal barriers that favor the adoption of DR in several countries. However, to align with the ambitious 2050 targets, it is necessary to fast-track around 500 GW of DR mechanisms involving residential, transport, and industry sectors and hydrogen production [190].

End-user DR actions usually occur in a short-term interval, within hours. Traditional examples of domestic DR actions include shifting laundry appliances, and lighting and charging vehicles. While for the industrial and hospitality sectors include reducing heating, ventilation, and air conditioning (HVAC) output and shutting down specific manufacturing processes, etc.

Various DR mechanisms have been designed to persuade customers to change their demand habits. They can be classified into Price-based and Incentive-based DR [188, 189], as shown in Figure 2.10.

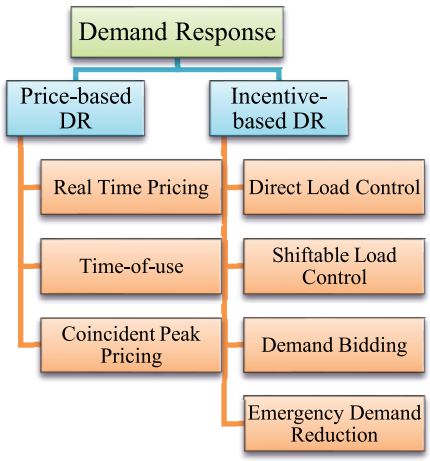


Figure 2.10: Classification: Main DR mechanisms.

The first category, also known as Indirect DR, comprehends time-varying pricing programs that tend to reflect the behavior seen in the energy wholesale market [191]. These can be split into the following subcategories: i) starting with real-time pricing (RTP), where electricity tariffs usually change on an intraday basis (with diurnal price differences), typically on an hourly basis, reflecting fluctuations in wholesale electricity prices. Consumers are usually informed on a day-ahead or hourly basis to act, accordingly; ii) time-of-use (ToU) tariffs are perhaps the most popular among residential consumers, and unlike RTP, prices are preestablished for contractual period that reflects the average daily wholesale electricity price variations; this means that prices change during the day, where typically two or three intervals are considered, that is, off-peak, mid-peak, and peak, which encourages end-users to shift their loads from the peak interval; iii) coincident peak pricing (CPP) is essentially a hybrid version of the previous two, by using averaged price points, with peak periods limited to avoid excessive exposure.

Both bi-hourly and tri-hourly tariffs, which are still very common among residential or

small-business end users in Portugal, are examples of ToU pricing in Iberia. While the regulated market in Spain introduced dynamic tariffs in 2014, the voluntary price for small consumers (PVPC) is an RTP-inspired mechanism with an hourly price that reflects among other things, the costs of production (wholesale market and ancillary services). Nevertheless, according to the household panel of the Spanish regulator (CNMC), consumers with ToU tariffs are much more likely to reschedule high-consumption electrical appliances than their peers with hourly dynamic tariffs such as PVPC [192], which is partially justified by being easier for consumers to adapt their consumption to a few pre-defined intervals than it is to follow market dynamics.

While the second category, which is also known as Direct DR, typically involves bilateral contracts, where large customers from energy-intense industries agree to reduce their demand by a certain quantity for a defined period to receive an incentive, and the non-compliance with the terms of the agreement will imply hefty penalties. This category accounts for over 90% of the DR load reduction [189].

Differences in the several subcategories of Incentive-based DR mechanisms, are mainly in the ability of the energy retailer/utility to directly control loads of the customer, the number of intervals and its extension, the nature of the curtailment need, i.e., a contingency or not, and in the specific case of the demand bidding mechanism, it targets the energy-intensive industries, that are also players in the day-ahead market, and are paid to alter its market bidding, thus curbing the pressure in the demand curve.

Several studies have been conducted on the application of DR mechanisms in Iberia. For instance, the research presented in [193], have analyzed the economic and environmentally benefits of DR electricity program in Spain. The main findings are that the generation sector is the biggest beneficiary of a large-scale implementation of a DR program in the long term. In contrast, the benefit for residential consumers is modest, and that a further integration of "smart grid" concepts to make the gains bigger. Behind-the-meter applications in which consumers manage their bills, reduce peak demand charges, and increase self-consumption from residential PV installations [142].

Similar inferences are drawn for the Portuguese case, where, despite the lack of a well-defined legal framework and public incentives, there are several important pilots involving the control of HVAC loads, load shifting in industrial loads, and a sizable smart grid project, InovGrid, whereby concepts such as self-consumption, load shifting, storage, and DR are integrated [194]. In another study, the role of DR in industrial and commercial buildings in Portugal was deemed vital in the path towards a "net zero energy" balance [195].

2.3.5 More efficient/cleaner thermal generation

Thermal power generation has maintained a steady presence in electrical power systems over the years, given its ability to ramp up generation when needed and provide inertia to the system with its large rotating generators. Therefore, even in an upcoming scenario where renewables constitute the overwhelming majority of the generation portfolio, cre-

ating curtailment and ramping challenges [196], it is essential to understand the critical role of thermal generation and its associated flexibility to deliver fairly quickly, additional active power, and inertia to the grid, which in turn ensures its stability and supply security [197].

Despite the consensus on the role of thermal generation in the decades to come [196], the truth is that, from the perspective of an investor/power company, the existing paradigm, similar to that observed in Iberia, with reduced cash flows, hefty charges, a distorted public perception about the impact and dependence on these plants, squeezed annual capacity factors, and lack of incentives/backup payments, means that there are few reasons to invest in new capacity of this type [198].

As such, it is essential to not only reiterate their importance, but, more importantly, to pinpoint the progress of these technologies, both in terms of a slimmer carbon footprint and in operational terms, particularly given the greater efficiency levels, amplitude of power output, and ability to work as a source of heat. Depending on the country's needs, investment availability, and external paradigm, an all-around approach is possible, as shown in Figure 2.11, which highlights the major options available that simultaneously fulfill the quest for added flexibility while mitigating a large portion of the harmful effects of previous iterations of these thermal technologies.

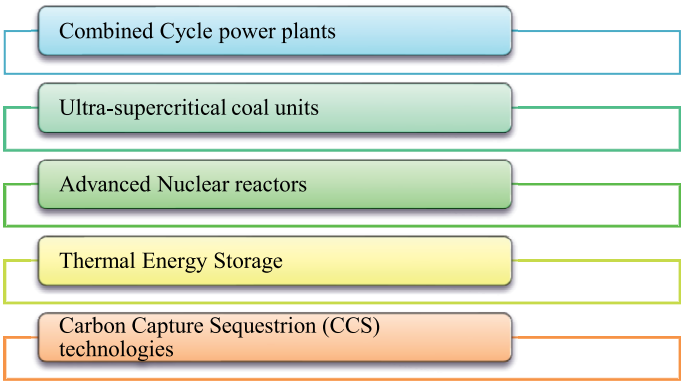


Figure 2.11: Main modern solutions (technologies) for a "greener" and more efficient thermal generation.

As the profile of thermal generation changes from a conventional base-load source to a (responsive) supply side variability accommodation (peak power), an increased number of startups (cycling) and fast load transients are required. Conditions clashed with the original high-capacity factors (few cold starts) for which they were originally designed. Consequently, this reduces thermal efficiency, slightly increases emissions, and adds thermal fatigue and component stress for coal and older gas-fired power plants [199,200]. The following subsections explore the progress and features of the main thermal technologies.

2.3.5.1 Combined cycle gas-fired generation

As seen in the annual generation mix analysis (section 2.2), both Portugal and Spain increasingly depend on gas-fired generation to ensure the security of supply during periods of low renewable production; this underscores a global trend that saw the gas used for electricity generation double between 1992 and 2016, reaching approximately 1500 billion m^3. Strengthening the notion, that among policymakers, natural gas was and is seen as an important "bridging" source to help decarbonize the power systems [201]. Validating this point, the EIA estimates that between 2022 and 2025, 27.3 GW the new natural gas-fired capacity is scheduled to come online in the US alone, resulting in a 6% increase in capacity (compared to August 2021) [202].

Its relatively low carbon intensity, only approximately 56% and 71% of the coal and oil carbon output, respectively [201], renders it an attractive bridging dispatchable high-capacity energy source while phasing out dirtier fossil-fuel-based generation. In addition, as mentioned earlier, the higher production flexibility of these technologies, particularly in CCGT plants where the reutilization of exhaust gases drives efficiencies of up to 60%, is in contrast to the typical 35% of coal-fired generation [203].

Regarding sizing, there is also some variability depending on the application, and the U.S. DOE reports in its "Combined Heat and Power Technology Fact Sheet Series", very few single gas turbines exceed the 300 MW mark. At the same time, they can have outputs as low as 30 kW (microturbines). However, it is also stated that for combined heat and power purposes, the 5 MW and above is where the threshold for greater profitability resides [204]. Nevertheless, in CCGT plants, several single units are commonly installed, forming different groups known as power blocks. According to the EIA, with technological breakthroughs, the average size of these power blocks has increased significantly since 2015, reaching an average of 820 MW in 2017 [205]. In addition, in terms of configuration, combined cycle (CC) units feed the exhaust gases from the gas turbine(s) to steam via a heat recovery steam generator, which then drives the steam turbine, that is, they combine the features of both the Rankine and Brayton cycles to achieve greater efficiencies [206]. The EIA reports that two combustion turbines (2×1) with one steam turbine are the predominant configuration among the existing CC power plants in the US; this is reasonable, because it normally surpasses the output power of a (1x1) configuration, which is the second most common configuration [207].

Owing to their higher efficiency, relatively lower carbon emissions in comparison with coal-fired utilities, smaller land requirements, quicker construction timelines, enhanced operational amplitude with faster ramping rates (advancements on start-up processes and improvements on control systems like Load Commutating Inverters and Adjustable Speed Drives), and wide range of sizing options, modern gas-fired plants present good maturity levels, as evidenced by the LCOE in Table 2.4. As such, they are advocated by experts as one of the best technologies for balancing fluctuating wind and solar PV. Therefore, it plays a vital complementary role for generation portfolios, such as those in Iberia, particularly in the absence of large-scale storage technologies and effective demand-side man-

agement [208].

2.3.5.2 Coal-fired generation

Notwithstanding the efforts in Iberia, and more broadly in the west, to phase out coal generation, the single largest source of global carbon emissions, the truth is that a large share of worldwide electricity production results from coal combustion. According to the most recent data from IEA, coal-fired generation is set to reach an all-time high in 2021, reaching approximately 36% of the global power mix, with China and India leading much of this growth [209].

Given the unbearable natural gas prices due to the war in Ukraine, the rising trend in electricity derived from coal sources is expected to exacerbate by 2022, including in the EU. Even with a smaller dependence on natural gas, in comparison with Central and Eastern European countries, the contamination of global markets has also severely impacted countries like Spain and Portugal, where around the beginning of Q2 2022, a special temporary agreement was reached with the European Commission to limit the price of gas used for generation. Still, the unstable situation, which was already seen in 2021 (to a lesser extent), has forced Spanish authorities to halt some of the expected closures of coal plants (again, as seen in subsection 2.2.1); a serious public debate took place regarding the closure of both Pego and Sines coal units, and whether, given the current framework, it was a premature decision or not; this is a valid discussion, when by several accounts, the flexibility levels are insufficient to renounce coal units use in the Iberian system.

This scenario is a hard reminder of how many exceptional times can change more grounded assumptions. In this case, it compromises the competitiveness of gas-fired generation, that is, the marginal costs. Therefore, it is necessary to provisionally re-establish coal. For instance, reports from Germany suggest that a policy turnaround is likely, meaning a delay in the scheduled decommissioning of some facilities and that some retired coal-fired thermal power plants may return online [210].

Before discussing the evolution and adequacy of coal generation units, it is vital to understand that the relatively low calorific value of different types of coal, for example, well below the natural gas value, is a limitation. Moving on to state-of-the-art coal-fired generation technologies, the observed guiding principle is to address both these plants' environmental burden and efficiency [211]. As such, the implementation and development of ultra-supercritical (ULTRSC) units, as well as the use of advanced capture and sequestration (mainly carbon) systems, are seen as leading solutions for future cleaner coal generation [212]. On the grounds that the second subject will be the topic of a separate subsection 2.3.5.5, the focus will be on the first.

First, it is important to briefly contextualize progress in the design and technology of coal power plants. These are often categorized in terms of the pressure and temperature conditions under which the fluid, commonly water, operates below or above its critical point (liquid-vapor coexistence). Accordingly, Figure 2.12 briefly contextualizes the advances in coal-fired generation, particularly regarding the most important thermal characteristics.

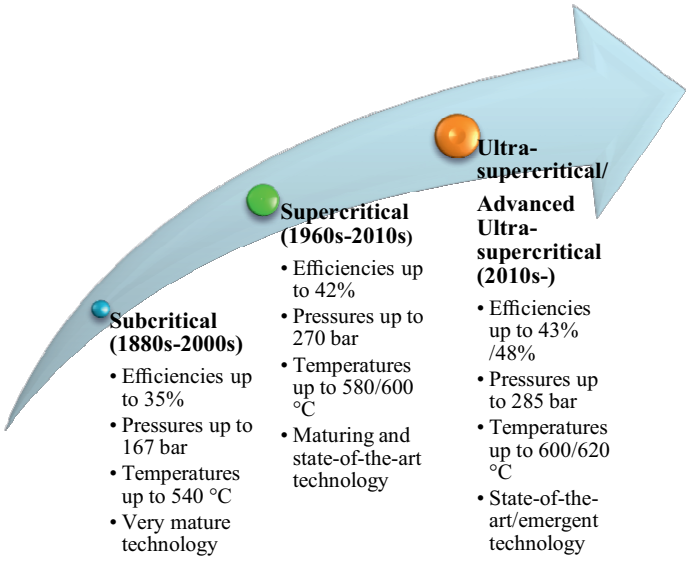


Figure 2.12: Evolution of coal power plants technology.

The initial plants were made of subcritical units (SUBCR), achieving thermal efficiencies of approximately 35%. In contrast, the efficiency of supercritical plants (SUPERCR), which were very common in the last quarter of the past century, ranged from 38% to 42% (in the more modern versions). An improvement directly tied to the substantial jump in the (boiler) pressure (∼+50%). In the last few years, ULTRSC power plants employ more optimized designs and advanced metal alloys, allowing even more extreme conditions, with superheat and reheat steam temperatures reaching up to 620 °C. Finally, Advanced ultra-supercritical power plants (A-ULTRSC), which are expected to come online during the next decade, will likely approach the 50% efficiency threshold given the significant rise in pressure, well into the 300s bar, and temperature, around 700 °C [213–215]. For example, an innovative design involving the integration of hard coal or lignite gasification technology with a CCGT, known as an integrated gasification combined cycle (IGCC), is considered a promising solution to curtail the pollution of common coal units [216], given their higher carbon conversion efficiency, when compared with traditional pulverized coal plants, and the need for less water.
In a comparative table presented in [217], the operational parameters for these three main types of coal plants with a 500 MW net output capacity are benchmarked. Concerning the parameters of the ULTRSC with SUPERCR and SUBCR plants, sizable reductions of approximately 11% and 21% were observed in coal use and CO_2 emissions, respectively. Thus, there is clear progress towards a less intensive and more flexible GHG energy source that diversifies the thermal technologies in the portfolio and fuel sources. When capital costs decrease, this is a viable technology to at least consider for a reserve role.

2.3.5.3 Nuclear powered generation

As discussed in subsection 2.2.1, the Spanish generation portfolio relies on nuclear power as a stable and considerable energy source to meet constant demand. However, with an

aging nuclear infrastructure, Spain, where the newest plant dates to 1988, with most reactors surpassing the forty-year mark, faces a common problem across the west. For instance, half of the currently operating reactors in the US are more than 40 years old [218]. Similarly, in Europe, most plants were built during the 70s and 80s with a typical projected lifespan of 40 years [219]. Even if reactor lives were extended by a few decades, many countries could still face a tough nuclear power crunch between 2030 and 2050. In practical terms, from the estimated total of 101.9 GWe of nuclear net capacity (2022 approximate figures), the majority will come offline during this period, while the commissioning of new nuclear capacity lags by a significant margin. According to the World Nuclear Association, without a change, this will lose almost all the EU's nuclear generation capacity by the mid-2050s [220]. Therefore, the aging nuclear power infrastructure constitutes another challenge for power systems simultaneously pushing for decarbonization and integration of variable renewables. Hence, such a high-capacity factor, dispatchable, and alternative to CO_2-emitting technologies, which account for nearly 10% of global electrical energy generation [221], is deemed vital for achieving ambitious goals while preserving system flexibility.

To address this issue, an increasingly common trend is to use advanced industrial plant life management practices that closely monitor (and potentially replace) the key components of a nuclear plant, including the reactor core, cabling, and external concrete structure, to extend the lifespan of these plants, as recommended by the International Atomic Energy Agency (IAEA). The U.S. Light Water Reactor Sustainability Program is a good example of how the implementation of these measures can foster the long-term operation of the nation's nuclear infrastructure. Regarding the addition of new capacity, rough estimates from the European Atomic Forum (FORATOM), in 2015 proposed the commissioning of 100 new nuclear power reactors between 2025 and 2045, totaling more than 122 GWe [220] to (only) offset the loss of existing capacity in Europe alone. In this regard, the European Commission, "Energy roadmap 2050" targets, seem unfeasible, without an urgent proper (significant) transnational plan for the construction of modern nuclear plants and given that its largest economy planned to phase out nuclear production by 2022. Despite the lower comparative capital cost, extending the life of existing plants rather than investing in new nuclear capacity is still not cheap, and it may not solve all security problems.

The circling "National Energy and Climate Plan 2021-2030", that foresees that Spain is on track to complete the nuclear power phase-out by 2035 [222], seems unsuitable in such unstable scenario, with high penetration of variable renewables without sufficient diversification and lack of storage capacity, and when considering the only alternative source, gas-fired plants, is tremendously affected by record commodity prices. Thus, for these reasons, it is a decision that will certainly negatively impact system flexibility. Besides, the policy makers idea to lessen the dependence on nuclear energy may not contribute to the CO_2 net-zero goal, and according to the IEA, "Spain 2021: Energy Policy Review", despite the progresses in ramping up renewables in its electricity mix, the future trajectory of

the power mix warrants careful consideration to ensure a smooth transition, i.e., without compromising the security of supply and the overall market competitiveness. The same idea is reinforced in a report in [223], where through the analysis of several scenarios, the concluding remarks indicate that the ramp-down of nuclear energy in the US would compromise the goal of curtailing emissions, leading to an outcome with an overdependence on natural gas sources, as the current growth of renewables and its intrinsic variability cannot keep up with the loss of nuclear generation. Moreover, it is also mentioned that the distorted market framework in some states, where subsidized renewables bid at negative prices, threatens the economic viability of several nuclear power plants. This lesson was also seen in countries such as Germany and Iberia (on a reduced scale).
As of 2022, and after weighing in on many of these factors, the European Commission and the EU parliament have ruled out in favor of labeling nuclear plants as sustainable "green" investments, when subject to stringent safety and environmental requirements. Therefore, a very important step towards recognizing the importance of a diversified dispatchable portfolio is becoming more environmentally friendly. Beyond the general division between fast and thermal reactors, the most popular categorization of nuclear thermal reactors is the generation stage, as shown in Figure 2.13.

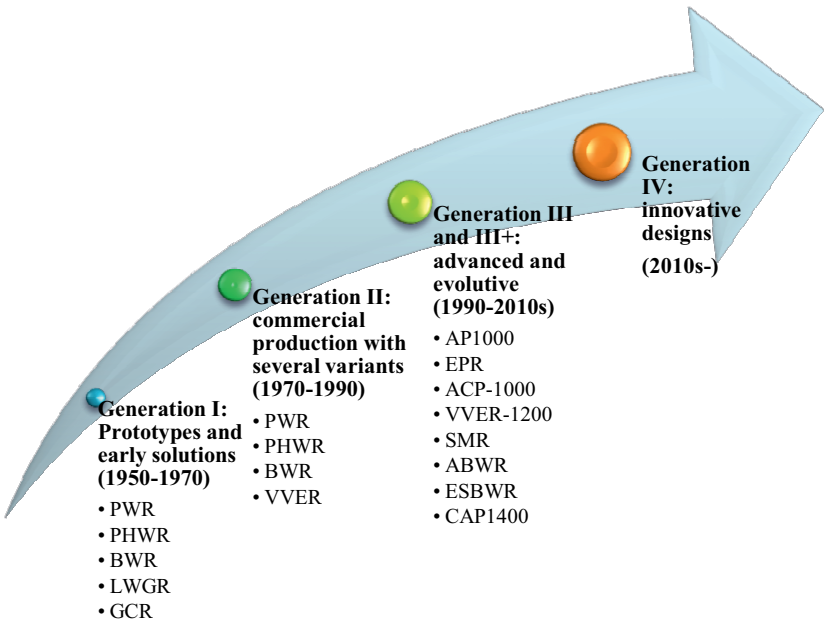


Figure 2.13: Evolution of nuclear reactor technologies.

Generation II type reactors constitute most installed units, with PWR as the preferred design (67% of the total operating nuclear reactors as of January 2020). Nonetheless, PWR design has evolved over the years. Some noteworthy more modern PWR generation III and III + type variants include the US Westinghouse AP1000 2-loop design, Russian VVER-1200, and the European Pressurized Reactor (EPR) in France, and Chinese ACP-1000 and CAP1400 [224].
The EU has fallen behind some of the other leading economies in terms of integrating some modern nuclear capacity designs. For example, the EPS of the neighboring country

of France, is heavily reliant on nuclear energy, which, as part of the necessary nuclear "fleet" renewal, is expected to only have its first generation III+ reactor ("Flamanville 3"), fully operational, in mid-2023, after successive delays. This EPR design will bring an additional estimated capacity of 1570 MWe to the Flamanville nuclear power plant, where two generation II PWR reactors have already been installed. As a result of the energy crisis in Europe, the French government was forced to partially change course regarding the prior intentions to substantially move out of nuclear power, and according to news outlets, the instruction to the state-controlled EDF-Electricité de France was to accelerate the construction of new generation III+ nuclear reactors.

By virtue of net output capacities per reactor between 1000 MWe and 2000 MWe, new nuclear power units should not only be considered to replace some of the aging active plants in Spain, but also be brought back to the discussion table in a country like Portugal. Even with a large penetration of renewables, preserving the old high-capacity factor baseload power paradigm is important. By analyzing several past electricity demand curves for the Portuguese system, a thermal base of around 2–2.5 GWe has been consistently crucial; as such, a modern nuclear plant with an extended lifespan can prove not only to be financially competitive, but also to help reduce the carbon footprint of other thermal technologies. In addition, the ability to adapt the architecture of these plants to supply heat and hydrogen production for non-electrical applications is a secondary advantage for this energy source.

2.3.5.4 Thermal energy storage

Thermal power plants and industries often release significant amounts of hot gas(es) into the atmosphere. However, this is not inevitable, because thermal energy storage (TES) solutions have been developed for many applications, resulting in a fast-paced increase in cumulative capacity. Broadly, thermal energy can be stored in two main forms: sensible and latent heat. Owing to the use of phase change materials, the last form is considered a better solution, given its higher storage density for the same volume of material. As of lately, thermochemical heat storage solutions have also been gaining adoption [225], and the same was underlined when concentrated solar power technology is discussed in subsection 2.3.3.5.

TES solutions can be very useful not only for obvious cogeneration facilities, but also for public buildings such as hospitals, where HVAC systems represent a large portion of the energy consumption to help reduce peak loads, which maximizes the match with the output of solar PV panels.

The main waste heat energy sources are presented in Figure 2.14, representing the activities with the highest waste heat recovery potential. In most cases, these industries effectively use only approximately 40% of the input energy, which means that a significant margin of otherwise lost energy can be subsequently used for either electric energy conversion or secondary heating applications.

A very promising TES solution is the E2S power thermal energy storage technology. Its working principle is to convert clean electric energy from RESs into high-pressure heat

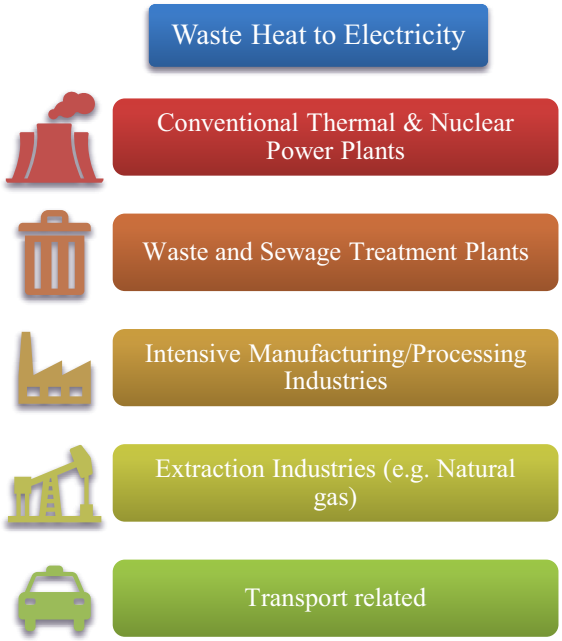


Figure 2.14: From waste heat energy sources to electricity.

(boiler), store it, and, when necessary, feed this superheated steam to existing turbine generators to generate electricity [226]. By repurposing old or even retired thermal power plants (validated in coal power plants), E2S power's technology can store surplus energy from VREs and deploy the necessary system flexibility during peak periods. A pilot with 250 kWh of thermal energy was established by the India power corporation [227]; this gives second life to otherwise abandoned thermal power assets, which is another advantage for the environment.

2.3.5.5 Carbon capture and storage, and other filtration technologies

Carbon capture and sequestration (CCS), as the name suggests, consists of the process of separating the produced high-temperature CO_2 streams, compressing and transferring them by pipeline or ship to a storage location, and later injecting them deep beneath several layers of ground (2-4 km, passing the impermeable seal), that is, the sequestration process of the unwanted byproducts of fuel combustion [228]. Typically, geological sequestration is preferred in depleted oil and gas reservoirs or deep saline formations; this is illustrated in Figure 2.15.

Chemical conversion is possible even with terrestrial sequestration. Furthermore, since power plants account for one-third of total CO_2 emissions, it is easy to realize the potential of CCS technologies to significantly reduce GHG emissions, thus fulfilling the goal of adapting existing or future thermal plants to a near-zero emission footprint [229].

Nonetheless, the high costs of the associated infrastructure and, to some extent, the lack of policy support mean that the adoption of CCS technologies is still moving slowly. Recent studies [230, 231] have highlighted several technical/economic challenges and opportu-

nities posed to modern gas-fired plant equipped CCS systems, namely, a degraded LCOE, which is higher than the addition of renewables; this suggests a straightforward path for its application in coal and fuel oil plants. Approximately 40 million tons of CO_2 per year are captured and sequestered from all sources. Even though reputed agencies like the IEA, in its "Energy Technology Perspectives 2020 series",designated it as a key vector to the decarbonization policy, with the major advantage of being retrofitted to the existing troublesome industries.

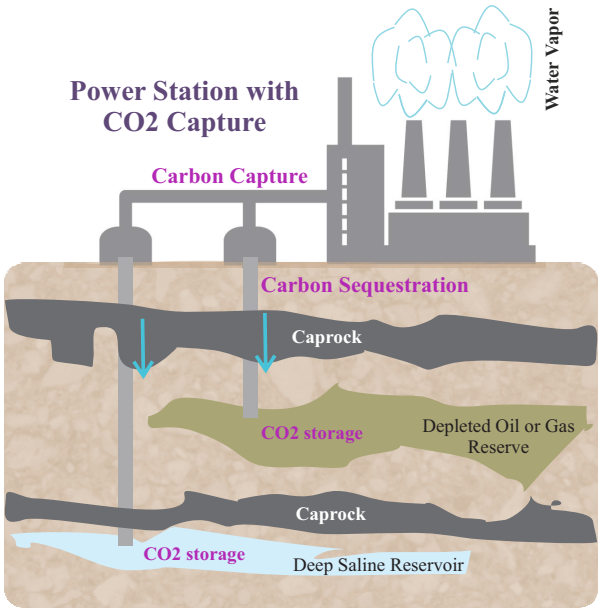


Figure 2.15: CSS modalities for thermal power plants.

Even more ambitious is the net emission reduction of atmospheric CO_2, which is the goal of projects such as bioenergy with CCS via the combined effects of photosynthesis and CO_2 capture and sequestration [232].

Sun et al. [233] revealed that after a few assessments supporting CCS development in Spain, three pilot projects were developed between 2006 and 2014. Namely, Compostilla OXYCFB300, Elcogas, and La Pereda. Nonetheless, high costs and uncertainty regarding its application scope have meant that progress has stalled despite an estimated potential of up to 68.7 Mt CO_2 per year ($\sim$21% of Spanish emissions). Given its lack of alignment with the EU strategy for CCS, the Spanish EU Council presidency was questioned after a suggestion to limit efforts to industries that cannot fully cut emissions, thus excluding electricity generation [234]. On the Portuguese side, the situation is similar with no grant requests to engage in CCS. According to the DGEG, the current and past projects have been limited in confirming the feasibility of these technologies [235].

Other well-known processes/solutions for curbing the GHG emissions of conventional thermal plants include generalized and improved coal washing practices, flue gas desulfurization, new generation fabric filters and electrostatic precipitators, low-NOx burners, among others.

2.4 Discussion

The previous sections have highlighted: i) the path that is being followed in Iberia; ii) critical aspects of policy decisions surrounding thermal dispatchable capacity and targeted large-scale penetration of renewables; iii) the main action vectors regarding system flexibility, with a thorough review of the main technologies, weighing in on their suitability/feasibility, maturity levels, breakthroughs, and implementation instances in the two countries. As such, the purpose of this section is to detail the shortcomings of the current approach concerning the generation side, assess the challenge posed by electrification (mainly transportation) regarding the demand side, and conclude with an all-encompassing view that allows decision makers to think through all the elements and avoid popular but often short-sighted policies that can be hurtful for long-term objectives.

2.4.1 The leading problem with the current RES integration approach

The variability/intermittency of renewables, such as wind and solar PV technologies, is far from a minor operational issue. The hourly capacity factor for these technologies is assessed using data from the REE [91], that is, the ratio of measured onshore wind and PV generation to the total installed wind and solar PV capacity, for a solid 5-year period (between 2017 and 2021) in Peninsular Spain. The results were then analyzed, and six different classes of capacity factors were formed and compiled by season, as shown in Figure 2.16. The bar plots reveal the magnitude of the issue in question, that is, even with increasing (significant) installed capacity, the net dispatch of hourly capacity for these technologies is reduced, with almost all the occurrences falling below the 55% capacity factor class.

Wind generation has a mean capacity factor of 24% for the entire period, translated by higher bars (see the first subplot in Figure 2.16) for the class of capacity factors between 10% and 25%; this means that this class had the highest relative frequencies across seasons, followed by the class with capacity factors between 25% and 40%. Additionally, the large degree of seasonality of this power source is also perfectly visible, given the disparity in the relative frequency for capacity factors above 40% for winter seasons approximately 1/4 of the time; during the summer, they are very sporadic, with this percentage falling to only approximately 2.6% of the hours. As expected, the mean value for solar PV across all seasons in the same five-year span was even lower, at approximately 18%. Which is easily explained, by an average of $\sim$7.6h/day [143] of effective solar irradiance; this means that even in the sunniest parts of the Iberian Peninsula, for 2/3 of the day PV farms cannot generate electricity for two-thirds of the day. However, unlike wind generation, there were more occurrences in the capacity factor class between 55% and 70%, particularly in summer, when it represented a relative frequency of 0.183, in contrast to the behavior observed during winter; this underlines an even lower capacity factor and strong seasonality for this technology. This fact alone justifies why, until a set of diversification options is implemented, which will take time on a system-wide scale, the role of thermal generation

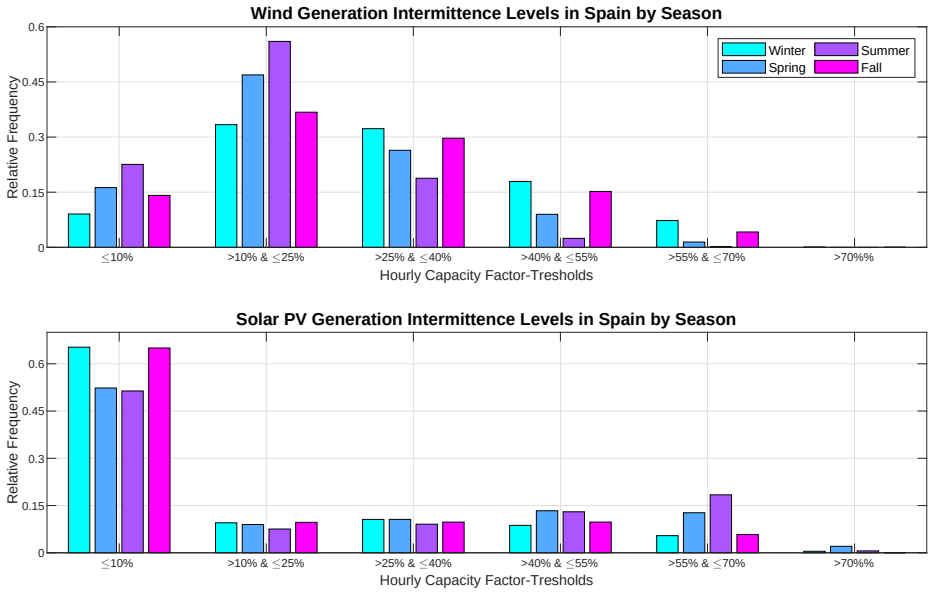


Figure 2.16: Wind and PV generation (hourly) power capacity factors by seasons of the year (2017-2021).

should be valued rather than downplayed in the short term.

Moreover, as hinted in section 1.2, there is a heightened chance for power imbalances as a consequence of source variability and intermittence. Notice that we are referring to a greater magnitude in these imbalances, given that there is always some residual difference even when served by a fully dispatchable generation mix, precisely due to load uncertainty, transmission losses and other dynamic factors. Coupled with the pressured decommissioning of synchronous generation (dispatchable but higher-emission thermal units) that inherently provide inertia to the system (passive support). An added factor that increases the likelihood of more frequent frequency disruptions (less latitude before resorting to more stringent active measures).

Concerning grid frequency fluctuations, a study conducted on data from Europe, the U.S., and Japan revealed two interesting facts [236]. Beyond the expected renewable generation feed-in variation, mostly due to instantaneous changes in wind speed and irradiance (both due to very dynamic atmospheric conditions), frequency fluctuations were particularly correlated with the 15-minute time frame in which most bulk power adjustments (different dispatch combinations) occur in the European intra-day electricity markets. These fluctuations were, in many instances — especially when comparing the UK and other European countries to the U.S. (which has a much smaller share of wind and PV generation capacity) — more significant than those caused by renewable feed-in. Additionally, as discussed in section 1.1.4, real-time markets in the U.S. operate slightly differently from their European intra-day and intra-hour counterparts. The second surprise concerned the statistical distribution itself. Grid fluctuations did not follow the expected Gaussian distribution (symmetrical around a mean) but instead tended to be more leptokurtic, much

like Lévy-stable distributions, i.e., with heavier tails (a non-negligible probability around the more extreme fluctuation values).

Circumstances that help highlight another significant difference between the traits of the dominant generating technologies in the classic versus the future power generation mix. A view of the main differences between synchronous generators and inverter-based generation systems is presented in Figure 2.17. The compiled list makes use of several adverbs or graduations of the adjectives to underline that these are useful but general traits. For instance, regarding the ramping rates, as we saw in sub-section 2.3.5.1, more recent CCGT-fired generation can actually be much more responsive. The operational lifespan of nuclear and hydroelectric utilities can extend beyond many decades, even reaching the century mark, in comparison with other thermal utilities that last on average up to 30 to 40 years. The capital costs cannot be generalized and it is important to normalize them by output power. According to the latest EIA reports, both simple and combined cycle turbines can be more competitive than some of the more established renewables, which is not a reality widely applicable to many European counterparts. However, energy analysis should focus on the potential and adequacy of the source rather than solely the capital costs. For example, wave and tidal energy conversion technologies have high capital costs due to their infancy and associated R&D expenses, but these should not be discarded or left untested in favor of cheaper options. Similarly, caution is needed when considering marginal costs, particularly for PHS hydro units, as the ”water value” is difficult to quantify — especially in the context of a fully market-integrated system. Precisely due to the intra-annual seasonality associated with rainfall, as discussed in sub-section 2.2.2, the value of water can fluctuate significantly throughout the year, which in turn can lead to substantial variations in the electricity merit order. For instance, during periods when reservoirs in pumped hydro storage (PHS) plants are full, this source can even compete with financially supported renewables, thereby altering the typical merit order described in sub-section 1.1.4. Similar dynamics can be observed with other dispatchable sources, depending on the specific market context.

Importantly, what Figure 2.17 really underscores — beyond the often-overlooked ’inherited’ limitations of relying exclusively on the currently available purely inverter-based generation portfolio — is the essential role of diversity in enabling the grid to ”absorb” a broad range of complementary beneficial traits. At least until some of the discussed technologies and solutions can fully replicate the predictability, stability, and reserve functions provided by synchronous-based generation, it remains crucial to approach the removal of these sources from the energy mix with great caution.

Countries with an extremely abundant (single) renewable resource, such as Norway — where hydropower generation already serves practically the entire electric load demand — face a different scenario, benefiting from added leeway that allows for better and more deferred planning in this case for sustaining the nearly 100% renewable energy system. Importantly, however, these are exceptions, with only about 10 countries worldwide achieving this, primarily through hydropower generation [4]. This serves as a testament to this

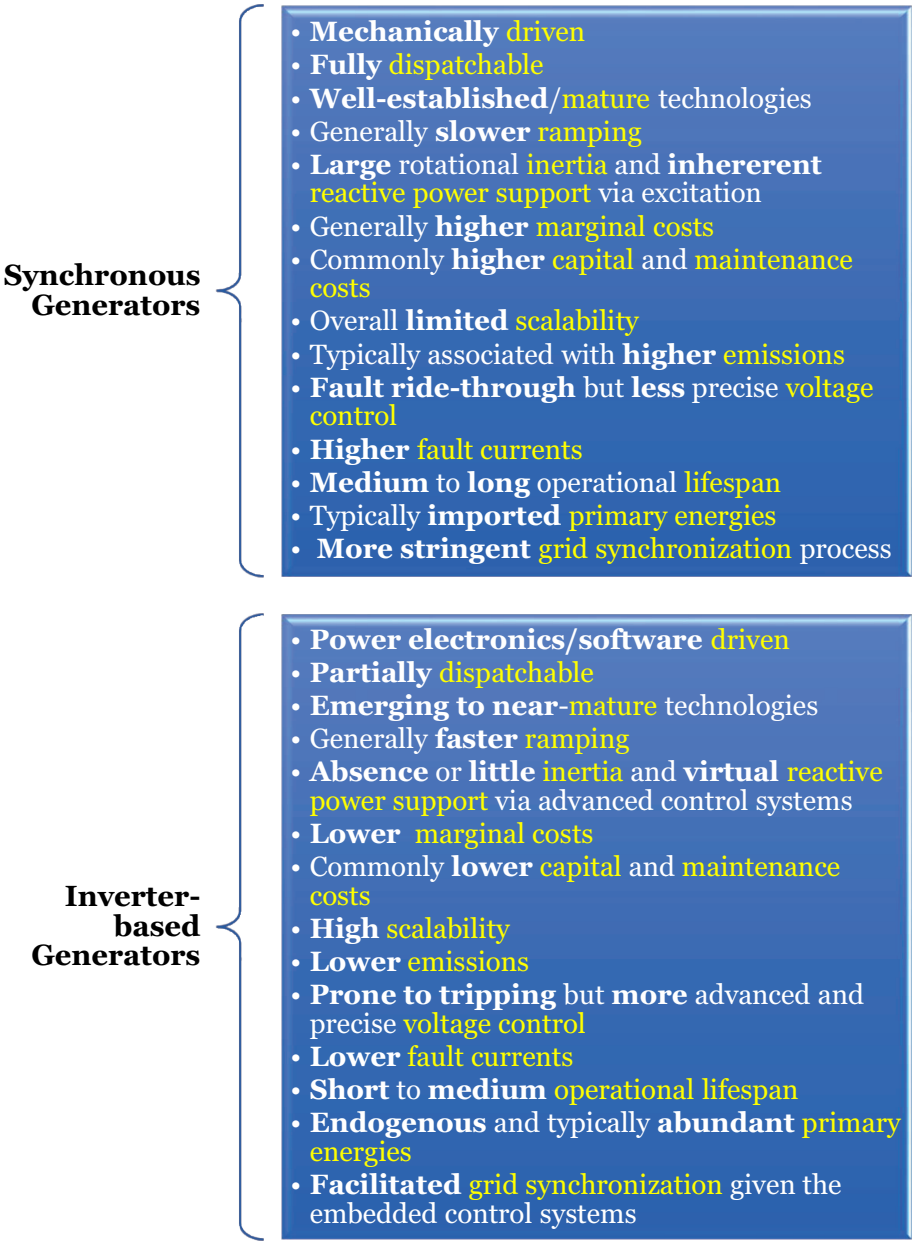


Figure 2.17: Key differences between synchronous generation and inverter-based resources.

source's scalability, provided there is sufficient water availability (key to maximize the head drop and the water flow), and its (still) unique large-scale storage capability.

2.4.2 Challenges posed by electrification

The EV market has experienced exponential growth in the past decade [237] in an overall electrification trend — replacing end-use nonelectric energy demand with electrically powered equivalents to reduce GHG emissions, with transportation as the main target, followed by HVAC and industrial processes [238]. Consequently, electricity demand is expected to rise near the medium term. Thus, with the potential to disrupt the system in

terms of peak demands, not only in terms of amplitude but also concerning the profile itself (altering the peak hours) at the transmission level and voltage regulation issues at the distribution level. This, in turn, cascades to a broad set of issues, namely economic dispatch and generation reserves, real-time electricity markets, congestion and transmission control, energy quality, low hosting capacity (feeders), among others [239, 240].

A study in the US [241] provided valuable clues regarding the impacts of electrification on system demand. For instance, even in its medium scenario, with plausible assumptions in terms of EVs, heat pump and industrial appliances adoption, compound annual growth rates are estimated around 1.2%/year (2016–2050), resulting in a 20% higher electricity demand in 2050 than the reference scenario, with transportation as the overwhelmingly contributor to this incremental growth (∼87%). Despite falling below historical rates between 1982 and 2016, in absolute YoY terms the change (2016–2050) is significant, matching a growth of ∼50-55TWh/year on average terms. Accordingly, the electricity share of end-use energy consumption will rise to 32% in 2050 compared with 23% in the reference scenario. This report also points to a shift in load shapes as another effect of electrification, with pronounced peak demands, especially evening peak loads, and more people charging their EVs after work. For the same medium scenario, the peak national hourly demand was estimated to be approximately 40% higher than in 2016. In addition, the electrification of heating (heat pumps and water heating) results in a greater peak demand during the winter season than during the summer season (which is already reliant on electrically powered air conditioning), thus altering some of the seasonal and diurnal traits of the US load. Up to now, in Iberia, the yearly load peaks (invariably) tend to occur during the winter [242]; still, given the level of energy poverty (as touched upon earlier), there is potential for an uptick in peak demand across seasons.

The increase in evening peak demands, together with a lower demand between the morning ramp and early afternoon, due to an increased adoption of residential PV, has the potential to aggravate a phenomenon known as the duck curve, which in turn may lead to an even lower load factor (ratio of average to peak demand). The same is illustrated in the uncontrolled electrification scenario (particularly for EV charging) in Figure 2.18.

Nevertheless, the uptake of DR mechanisms and storage solutions, including vehicle-to-everything (V2X), flexible (dynamic) EV charging, favorable tariff-based pricing, and residential energy storage solutions, can lessen (shave) the increase in peak demand by diluting it throughout the day, leading to greater load factors, thus deferring the scale of (required) grid reinforcements [243, 244], the same is also highlighted in the EV-based flexibility scenario in Figure 2.18. The same was confirmed in [245] for a real case study based on the Iberian Peninsula in a scenario with a large integration of EVs and a renewable-dominated mix. Proper coordination (control) between EV charging and the system operator makes it possible to avoid wind spillage and reduce operating costs.

Faced with a more challenging scenario, grid capacity planning and system flexibility are the keys to unlocking feasible energy transitions; this includes limiting VRE investments and preserving the role of conventional plants as backup sources [246]. Until these en-

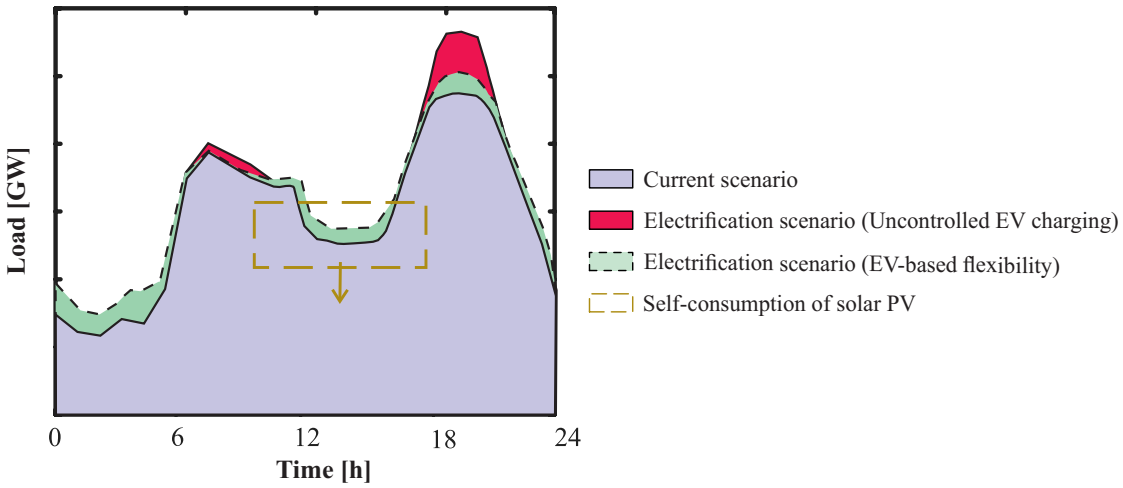


Figure 2.18: Electrification impacts on the system load (scenarios): Current ”duck” load profile (lilac); Uncontrolled electrification with sizable peak load increase (maroon); Electrification plus system flexibility, especially EV-based (mint); PV self-consumption downward effect (olive).

ablers are implemented, caution should be exercised when planning a more drastic addition of EVs [247]. Therefore, the pursuit of system flexibility based on a tailored and diverse array of options is the central theme of this study.

2.4.3 Lessons and policy insights

Moving in this direction, the performed overview revealed several insights for the Iberian case study [248]. First, the technological breakthroughs that translate higher ramping rates, output range, thermal efficiencies, and consequently, less fuel consumption and reduced carbon footprint of modern thermal power plants justify a more open-minded look into these technologies, to help with both a baseload and/or reserve role; therefore, safeguarding the integrity of the system, not only by recognizing the impact of seasonalities and extreme weather conditions on the output of the major RES, that is, onshore wind, hydro, and solar PV, but also its role in providing inertia and frequency control to the grid. Diversifying this thermal portfolio is also advantageous in the face of unpredictable threats that may temporarily hinder the price competitiveness of a certain technology. Following the important EU decision, to label the investments in nuclear and natural gas as “green” [249], its technological maturity, with substantial developments in the reactor’s safety and heat to electricity efficiency, very high-capacity factor, lifespan, predictable LCOE, among others, standout for its uniqueness. As such, following the opinion of credited energy-related organizations, the phase-out of the existing nuclear capacity in Iberia is not consistent with the mid-term goal of decarbonizing the generation portfolio without signing up for the huge costs that other options would bring to the table to perform the same task. Therefore, Spain and Portugal should evaluate small modular reactor technologies to replace some of their existing capacities. In addition, these GHG-emitting technologies can be complemented by carbon capture and other filtering

systems to further reduce the environmental burden. Second, PHS will remain the basis for a system-wide storage capacity source. Still, integrating more mature electrochemical energy storage systems is a trend that should be explored in future PV farms, as is occurring in Portugal in virtually all new floating PV plants, and at the residential level. Third, alternative (including synthetic or virtual) frequency (particularly at the primary control level, i.e., to contain the $RoCoF$ values) and voltage support mechanisms need to be considered and deliberated according to the individual EPS characteristics [250]. These include [251]: i) high power density solutions in lower voltage distribution levels as highlighted in sub-section 2.3.1; ii) deploy stored kinetic energy from the rotating masses in the expanding fleet of wind turbine generators at the expense of a rotor speed decrease and an acceptable mechanical stress (calibrated); iii) PV modules working below the maximum power point to keep a certain reserve margin to provide active power control (an inverter-controlled throttled output according to different frequency levels instead of a more rigid switching off at a more precarious frequency event). This urgency was recognized already in 2012 (with far less levels of penetration) in the "System Stability Ordinance" issued by the German network regulator and may require a more prudent adaption/revision (on the sacrificed margins) given the expected PV prevalence and the problematic peak summer periods with an upside potential for surpluses; iv) new investments in STATCOM devices; v) feasibility studies regarding virtual synchronous generators via DC voltage sources connected to inverters, particularly in a microgrid context, among others. Fourth, despite the success of some classical DR mechanisms and the interesting PVPC price scheme in Spain, which has been modified during the analyzed 5-year period, load inelasticity is not easy to solve. In addition, the potential flexibility gains are more limited and less controllable than other action vectors. Fifth, investing in interconnections with France is critical to break the paradigm of an "energy island", thus promoting greater market integration; this is of particular importance given the clear-cut differences between the French and Iberian mixes, as it can constitute a good way to deal with some low RES generation periods, thus importing more from France, or in a foreseeable scenario of high penetration, avoiding the need for curtailment by exporting. Sixth, the vast Atlantic coast offers opportunities for offshore wind, ocean, and tidal energy sources to be integrated, and a look into solar thermal technology in a region where sun irradiation is high is also promising. Together with the advent of hydrogen energy storage, which can be produced and stored (e.g., during generation surplus periods) and converted to electricity from RES, and in the future, exported to the rest of Europe, as the European Commission emphasized. Seventh, anticipating a future overwhelmingly dominated by variable renewables, which will squeeze the volume of traded energy of dispatchable units, such as gas-fired power plants, in the primary electricity markets (spot), the importance of capacity markets/mechanisms is expected to grow [252]. More often than not, their offers will be to the right (higher) of the market equilibrium point. This scenario indicates a need to start reviewing and preparing a reinforcement of this mechanism. In this way, it ensures that these costlier (short-term marginal costs) and less frequently dispatched units will have access to a steady revenue

stream, and thus remain financially viable even when they are not actively generating electricity. This financial viability is crucial for maintaining their standby readiness, which is essential to justify their continued operation from a shareholder's perspective, and to ensure the security of supply during high-demand situations coupled with low output from renewable sources from the EPS standpoint [253] — at least until sizable storage scalability gains are achieved. Eighth, concurrently with the generation mix transformation, the expected electrification, dominated by a concentrated EV (charging) load, impacts the system demand at the transmission level (mainly in peak demand). Several sensitive issues at the distribution level also arise and require a set of custom-fit options, which entail the flexibility strategy described in section 2.3. Finally, the Iberian system can continue to lead the process to a more decarbonized portfolio by exploring several complementary solutions.

In a market-driven environment, the next generation of incentives and regulations, will be crucial to guide the overall power system towards a strengthened, diversified, and "greener" position in terms of flexibility options, without neglecting the issue of affordability, while avoiding some of the past distortions associated with FiTs.

Chapter 3

Soaring electricity prices in the day-ahead Iberian market: policy insights, regulatory challenges and lack of system flexibility

As a follow-up study, this chapter provides a targeted review of recent major regulatory changes at both European and national levels that have shaped energy policy decisions in the Iberian Peninsula — changes that may have exacerbated market flexibility issues from the perspective of the spot market.

To illustrate this, we begin with an original analysis of the unprecedented energy price spikes of 2021, focusing on the Iberian electricity spot market (MIBEL). Much like historical events serve as foundations for novels, this unexpected yet deeply impactful crisis — one that has significantly concerned consumers, industry, and governments — should serve as a compelling illustrative and historical case, underscoring the risks of inadequate flexibility and the urgent need to prevent similar disruptions in future scenarios.

By emphasizing market volatility, we examine how key regulatory shifts—along with their perception and absorption by stakeholders—have contributed to the current energy landscape. Furthermore, we underscore that neglecting flexibility jeopardizes the fundamental goal of maintaining an affordable and reliable energy system.

3.1 Energy policy insights: a catalyst for large-scale renewable adoption and its trade-offs

To rapidly advance in the integration of renewables, EU countries have been progressively adopting very bullish EU legislation considering a “greener” electricity agenda. The successive legislative packages have not only aimed at perfecting the horizontal harmonization of the electricity industry but more notably have set ambitious emission reduction targets and policy incentives. These, in turn, led to a diffusion of feed-in tariffs and other direct support schemes in the EU [254], as we will discuss in detail below. These, together with the plummeting costs of renewables [255], have been decisive to achieve the purpose of a higher share of clean generation. A great example of this path is Portugal, where for long periods of the day the renewable supply outstrips the national demand [256]. To advance the ongoing green transition, Langarita et al. in [257] has underlined the important role of fiscal measures in Spain.

As many authors point out, this path is not as straightforward as policymakers may hint [258–260]. Some unsolicited consequences have occurred, namely, operational questions, including substantial curtailment periods [261], compromised energy quality and frequency regulation, and electricity market-related problems. Feed-in-tariffs as a form of policy intervention have led to distorted price signals (including negative market prices),

impacting the energy production hierarchy by favoring some energy options to the detriment of others, thus compromising the economic viability of the top options. These market distortions can be linked to certain hikes in market prices, rather than leading to the originally expected downward price pressure [262]. For this reason, despite significant advances in renewables, Germany pays one of the largest electricity bills among all IEA countries. In addition, markets with a strong penetration of variable renewable generation negatively impact the merit order effect [263]. The lack of coordination in energy cooperation mechanisms and cross-border support instruments, among other factors, can further accentuate efficiency losses [264], hurting public acceptance. Consequently, a debate about the future of renewable support instruments should involve discussing whether capacity subsidies can be a better option than the dominating generation-based subsidies [265].

3.1.1 Spiking electricity prices in Iberia: events and timeline

During the late 2010s, the EU electricity markets were generally characterized by relatively stable pool prices. But the unparallel scenario of a worldwide pandemic caused significant disruptions in supply with reflections in the market. In MIBEL, the average wholesale energy prices decreased significantly during the harshest months of 2020, with plunges above 50% [114]. After a tough early 2021, due to the worsening health crisis both in Portugal and Spain, the electricity market prices started to rebound, following a broad recovery of other energy commodities. For instance, since mid- 2020, the daily closing prices of coal and crude oil futures rose gradually. Even before the war in Ukraine, the IMF reported [266] that natural gas spot prices quadrupled to record levels in Europe and Asia, in an unprecedented spiking trend in both duration and global scope.

The generalized spikes in EU wholesale electricity prices in 2021 were due to a convergence of factors: economic recovery in 2021; supply bottlenecks because of constraints in primary energy production and infrastructure; weakening of the euro; increased environmental taxes and licensing costs; freight hikes; and the restraint of producers to expand the supply. The significance of this surge varied between countries, but ultimately it affected virtually all sectors of the economy, and is considered one of the root causes of the considerable inflation levels [267]. A good example of this inflationary environment (the rising commodity energy prices and supply-chain disruptions) is the circular impact on the price maturity of renewables. According to the IEA, the overall investment cost of utility-scale PV and onshore wind plants may well have risen by 25% from 2019 to 2021 and early 2022 [268].

Not surprisingly, this surge in spot prices also occurred in MIBEL, one of the EU's most peripheral markets. Adding to an already challenging framework, coal-fired generation – the cheapest thermal dispatchable alternative to the limited and costly natural gas supply – has been gradually dismantled or heavily curtailed due to hefty CO_2 emission licenses. Furthermore, the inherent intermittence and variability of renewables, particularly in unfavorable meteorological seasons, plays a key role, despite the substantial efforts in terms

of renewables installed capacity. In the Iberian Peninsula specifically, the sheer volume of wind and solar power installed capacity means that abrupt swings in meteorological factors (eolicity and solar irradiance) produce considerably different outcomes in terms of generation mix. Crucially, the existing precipitation-dependent hydropower infrastructure, with its pivotal pumped-storage hydro (PSH) capabilities, can mitigate some variability and provide flexibility by storing during surplus periods and responding in shortage periods. In fact, hydropower and reserve thermal generation capacity is considered the primary source of system flexibility in conventional systems [269].

3.1.2 Study Methodology

The disconnect between guiding policies, public discourse, the speed in which this "green" transformation is required, and the day-to-day operation of intricate and large power systems, often times may lead to unintentional consequences, including a compromised price stability, a series of technical/operational difficulties, among others. These "concealed" hurdles, span through the entire electric power industry. In this context, the study of the impacts on the unique Iberian electricity markets is becoming more and more relevant to address those hurdles. An overview of the employed scientific methodology is shown in Figure 3.1.

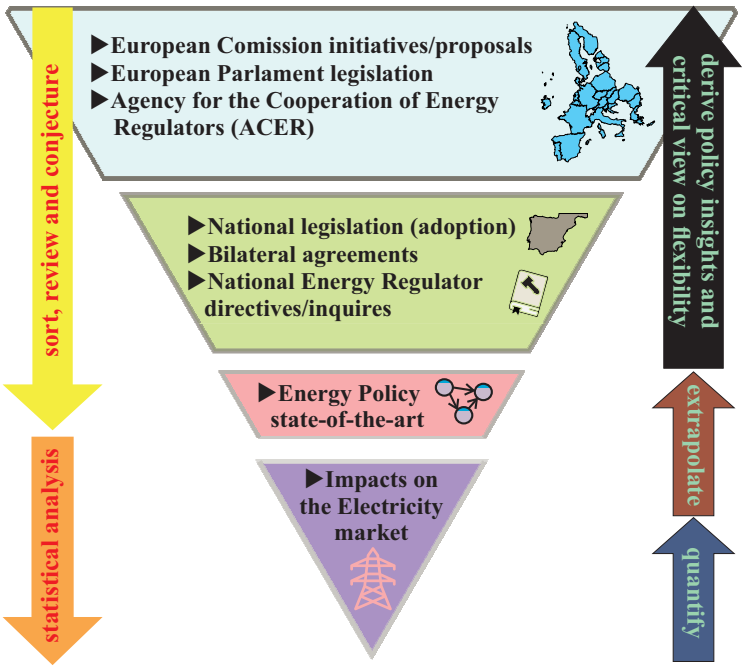


Figure 3.1: Methodology: literature/documentation review and market/grid operator data analysis (downward); extract result inferences and policy insights (upward).

To accomplish a proper (focused) research on the relatively recent spiking electricity prices in Iberia, and its relationship with regulation and policy decisions, we begin by surveying some of the leading EU energy policy legislation, such as the Regulation of the Energy Union and Climate Action, Roadmap 2050, Renewable Energy directives and other spo-

radic/reactive EU measures. With these regulations, EU officials urge the national governments to act in certain directions. And so, the focus in this work is to underscore the subjacent guiding principles/intents, emphasizing the disconnection between some of the energy policy goals. A particular interest is given to the market pricing limits regulation, a competence of the Agency for the Cooperation of Energy Regulators, which has evolved to facilitate market integration.

Much of these legislative goals, are then "absorbed" by the 2030 National Energy-Climate Plans in both Portugal and Spain and are implemented with the help of several intergovernment agencies. Concerning the market regulation changes, the Iberian regulators play a key role, thus, constituting another great source of information for this work. Especially, in their interaction with the market participants, expressed during the public consultation periods.

By presenting a novel description that help us to better understand the decision-making process, which is then evaluated through the lenses of the existing energy policy literature, we will be able to more accurately conjecture about the performed in-depth statistical analysis of the primary system-wide energy related data. Thereby, confirming or refuting prevailing findings, deriving policy insights, as well as drawing new inferences concerning the importance of system flexibility to help mitigate this disconnection (as depicted in the upward processes in Figure 3.1).

3.2 European regulatory environment: an overview of policies and their implications for MIBEL

As one of the oldest and largest (most interconnected) power grids, the European power grid has undergone significant transformations [270], particularly during the last few decades. The governing framework has been continuously reformed, shifting from the old vertically integrated paradigm – where national/regional governments controlled every aspect of electricity (natural monopolies) – to a liberalized framework with horizontal integration. Market efficiency, price competitiveness, and innovation are fostered by clearly splitting the roles in terms of production and by expanding the number of market participants. A broader and more atomized system reinforces the need for attentive, far-sighted, sustainable, and transparent regulatory agencies.

The urgency of climate change and the desire to reduce energy dependency has led EU policymakers, mostly during the last decade, to implement ambitious energy transition policies, impacting the entire system [271]. These included promoting a more variable, disperse, and smaller-scale generation mix, which led to new transmission capacity and operational issues, requiring grid reinforcement to be sorted out. For example a substantial effort is being made to reinforce transnational interconnections. Market bidding and commercialization also became more uncertain due to these changes on the price formation fundamentals: on the supply side, mainly given the increased generation variability, market volatility, added competition; on the demand side, a higher and more dynamic sys-

tem load (i.e., with the emerging electrification of transports and heating), coupled with a (final) consumer more prone to embrace price/demand-response and self-consumption regimes.

One of the most important pieces of EU legislation was regulation 2015/1222 [272], which establishes a "guideline on capacity allocation and congestion management", replacing major sections of European Community Regulation (EC) No 714/2009. It aimed to provide a regulatory environment for a fully functioning and interconnected internal energy market that ensures power flow from low price to high price zones, through single day-ahead and intra-day coupling. New "backbone" statutes are introduced, with a more mature view of the roles and operational scope of the "bridging" market coupling operators (MCOs) with both the (national or regional) electricity market operators (NEMOs) and the transmission system operators (TSOs). In article 7, NEMO tasks are clearly defined, including the determination of maximum and minimum prices. In turn, Article 41 establishes that, after a mandatory period of consultation, NEMOs propose harmonized upper and lower limits for clearing prices to be applied in all bidding zones (single day-ahead coupling).

With this framework, the legislators coordinated rules to facilitate market integration and, at the same time, asserted unassailable objectives, such as: security of energy supply, increasing competitiveness, and ensuring affordable prices. In addition, regulation 2015/1222 reinforced the role of the Agency for the Cooperation of Energy Regulators (ACER) in overviewing the NEMOs compliance with the established MCO functions, and in handling conflicts about terms and methodologies when the national regulatory authorities are unable to do so.

In subsequent years, numerous amendments were made, which further overrode Regulation (EC) No 714/2009. Regulation (EU) 2019/943 recasts the internal market for electricity governing policies (Energy Union). Importantly, it begins by reaffirming the commitment to provide final customers with "safe, secure, sustainable, competitive, and affordable energy". However, it fails to address the conflict between some of these commitments. For instance, sustainability with affordability, or intermittency with security, at least in a short horizon. The thoroughness of the document also offers good insight on some of the targeted policies and challenges facing the Energy Union in the years to come, such as: i) the profound change in the electricity market due to the decarbonization goals; ii) a priority dispatch accommodation of the small renewable power-generating sources; iii) the recognition that state interventions have resulted in an increased distortion of the wholesale electricity market, which may negatively impact investments and cross-border trade; iv) the increasing number of consumers no longer linked to regulated markets, who therefore play a more active role, particularly regarding their demand response; v) the need to strengthen market integration, with added cross-border electricity trade, which implies a coordinated action to modernize interconnections and reinforce their capacity; vi) last but not least, policymakers must recognize that market-based incentives should prioritize investments in flexibility resources, including flexible generation, interconnec-

tion, demand response and energy storage.

The aforementioned conflict between conceptual commitments is not a minor issue. In our opinion, it underscores an inconsistency, or at least an omission, between the narrative of most policymakers and the (present) functioning of electricity markets. The narrative omits the fact that the substantial addition of renewable generation capacity (with null or very low marginal costs) does not necessarily lead to cheaper spot prices and ultimately lowers retail electricity costs. In some situations, quite the opposite occurs, and this is not fully assumed. To a large extent, this is deeply tied to the lack of system flexibility in its different domains, as was recognized in regulation (EU) 2019/943 (section 3.3 will expand on this subject). This conflicting picture is highlighted in Figure 3.2, where the opportunities targeted by policymakers are shown in opposition to the required compromises, accompanied by an overview of the main actors and their interactions to enact this profound transformation.

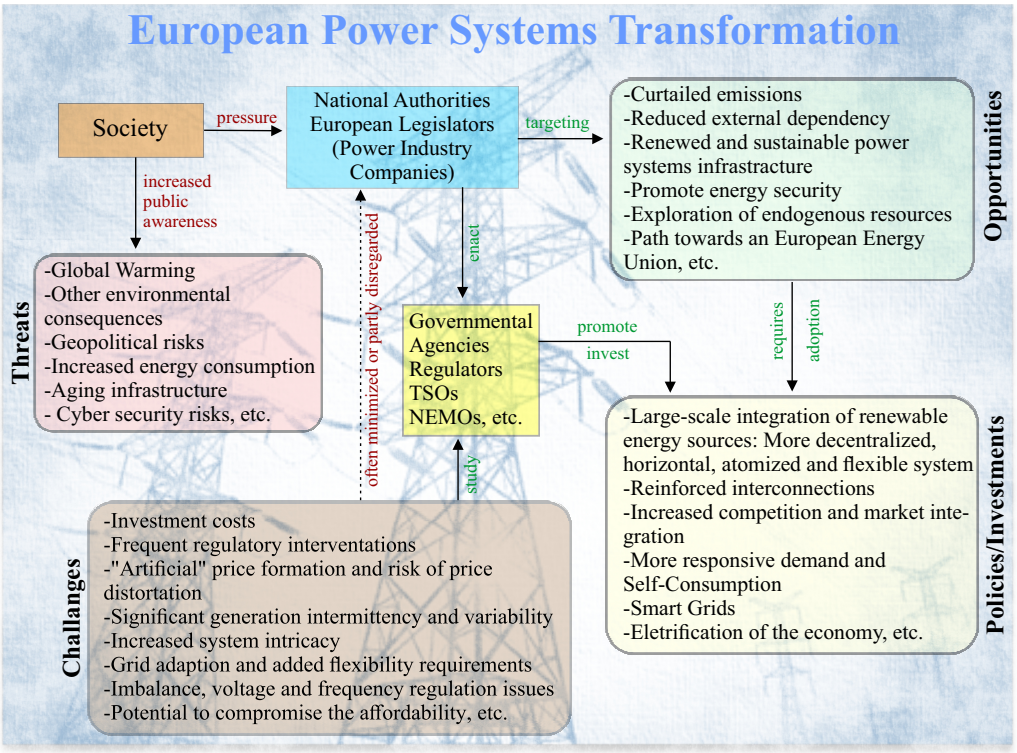


Figure 3.2: European Power Systems Transformation: threats, actors and drivers of change, policies and operational challenges.

To comply with the revised EU regulations, national authorities in Iberia, namely the Spanish National Markets and Competition Commission (CNMC) and the Portuguese Energy Regulator (ERSE), have been discussing the necessary overhauls, encompassing the adoption of common electricity (market) trade/coupling and grid operation rules. Ergo, at the end of 2019, the Spanish agency responsible for the control and supervision of the electricity and natural gas sectors, CNMC, published circular 3/2019 (consolidated legislation). Article 5 empowered the (MIBEL) NEMO, widely recognized as ”OMIE” within

the Iberian market context, with elaborating a proposal that observes EU legislation, after a consultation period with all the intervening parties (including the market participants). The proposal's focus concerns the implementation of new harmonized price limits for the offers (bidding) and clearing prices, in the day-ahead and intra-day markets. These should respect the legislator's goals, as expressed in regulation (EU) 2019/943, specifically in article 10, which reaffirms the principle that a free market should not have price limits. Nonetheless, the article also recognizes the NEMO's authority to apply harmonized price limits, which should be sufficiently wide to not restrict/distort the market.

In this regard, and as mentioned above, ACER's role is adopting a decision concerning the NEMOs' proposal when the regulatory authorities fail to reach an agreement (half-year timeframe) or upon a joint request. Therefore, following a request from all regulatory authorities, ACER released its decision 4/2017, where it proposed the following prices: i) harmonized minimum and maximum day-ahead market clearing prices of -500 €/MWh and 3000 €/MWh, respectively, with an automatic adjustment mechanism to allow a gradual increase if necessary; ii) harmonized minimum and maximum intra-day market clearing prices of -9999 and 9999 €/MWh, respectively. Following the lead of other NEMOs, this clearing price limits proposal was incorporated into the price formation algorithm that establishes the price in both countries.

OMIE's initial public consultation took place in late 2019, when the NEMO suggested a rule for updating the bid/offer price limits. Unsurprisingly, players with a higher market share preferred a solution that equalizes the bid(s) price limits with the clearing prices limits, while large and independent consumers and other applicants preferred a more gradual update rule, with a lower starting point. Having analyzed the responses, in late 2020, OMIE began another public consultation period, with a new proposal. After completing all procedural steps, this culminated in resolution 8362 (May 2021) from CNMC in consonance with ERSE. Among the numerous changes to the operating rules of the day-ahead and intraday electricity markets, the most important was the adoption of the same price limits for the offers/bids as those set for the clearing prices (ACER decision 4/2017). These rules came into force on July 6, 2021.

Constituting a considerable alteration of the market status quo until then, the governing rules considered lower and upper day-ahead bid(s) thresholds of 0 and 180.3 €/MWh, respectively [273]. These radically different limits marked a turning point in MIBEL operation/rules. They implied not only the possibility of negative prices for day-ahead sale bids (requiring an adaption of several accessory rules regarding the settlement of payments and guarantees) but also the upright scale of a 16.6-fold increase in the upper limit. Although negative pricing was not novel, having been observed in several regional US markets, they became more frequent when the US federal government started subsidizing wind generation, with regulations requiring that transmission grids absorb all this electricity [274]. This should serve as a reference for the Iberian market and involved parties, to avoid the unwelcome consequences of a prolonged scenario with these occurrences.

Additionally, the previous upper limit was also tied with price formation in a non-theoretical

sense, since the old regulatory framework required that a large portion of the energy purchase bids be placed exactly at the limit of 180.3 €/MWh [275], i.e., a large percentage of the daily demand (including all the last-resort consumers). So, after weighing in all of this, the occurrence of some sort of secondary effects in the markets was more than predictable.

Portugal is a good example of some of these issues. There was extraordinary progress to achieve a large share of renewable generation (as corroborated by the data analysis in section 3.3), mainly through wind generation, which now accounts for a recurrent annual share above 20%. A generous feed-in tariff (FIT) scheme, i.e., price certainty for each unit of energy produced and injected into the grid, was implemented to promote private investment in renewable capacity and achieve such high figures. To illustrate this, authors in [276] reported that, from 2001 to 2015, the average national FIT was around 103 €/MWh. This type of financial incentive allows renewable generation to enter the market with a "virtual" 0 €/MWh price, providing guaranteed remuneration when producing, and shortening the payback time for the often-disregarded capital costs of renewable technologies.

The mentioned FIT is a substantial value, particularly when it represents around 105% of the bi-hourly active energy tariff (regulated market) in off-peak hours, 0.0978 0.0986 €/kWh (2015), contracted by the common residential consumer (low voltage <=20.7 kVA), as mandated by the Portuguese regulator (ERSE). In a mere comparative exercise, if the FIT were reflected in the clearing price, then the price charged by the Last Resort Supplier (known by the abbreviation "CUR" in Portugal's energy sector) – even when accounting for the contracted power daily fee – would aggravate the already troublesome Tariff Deficit, clearly demonstrating the long-term unfeasibility of such incentive's scheme.

The end of coal-based generation in Portugal is another noteworthy occurrence of 2021, after the closure of its last active power plant, Pego (installed capacity of 628 MW), in mid-November. This closure followed the Energias de Portugal (EDP) group decision (in 2020) to anticipate to January 2021 the closure of the country's largest coal power plant in Sines (1180 MW) [277]. The Portuguese government had announced the Sines plant's expected closure for 2023, i.e., two years after the Pego plant. In the same statement, EDP also announced its commitment to shut down its coal production in Spain, by requesting the closure of its Soto de Ribera 3 plant (346 MW) in 2021, and the conversion of the 342 MW group in the Aboño thermal plant (fuel oil, coal, and surplus gases), while a second group maintains a backup role. To justify its decision, the company cited its decarbonization strategy, the increasing cost of coal production, the escalating price of CO_2 emission licenses, the hefty tax burden, and the relative competitiveness of natural gas. In hindsight, this last argument was extensively hindered given the surging in natural gas market prices during 2021. And in 2022, with the ongoing war in Ukraine, this over-reliance on gas-fired power stations has proven to be a liability. A similar direction was pursued in Spain, where companies have been requesting government authorization to close all the coal plants in the country. But again, the unprecedented scenario of 2021 prompted a

reevaluation of plans.

3.3 Evolution of Iberian power system: the case of the day-ahead electricity prices in MIBEL

To properly investigate the root causes behind the surging day-ahead electricity prices in MIBEL in 2021 and discuss how the revised regulatory framework may have contributed to this surge, one must first analyze (individually) the two main factors that determine the equilibrium price, i.e., electricity demand and supply [278]. The data used for the following analysis were obtained from the Spanish and Portuguese TSOs [279, 280] and from the NEMO [281] webpages.

Annual (accumulated) procurement is a good indicator to gauge the evolution of electricity consumption. Figure 3.3 underscores a relatively stable picture before the pandemic, with annual demands above 260 TWh in the Spanish system and around 50 TWh in the Portuguese system. The severe restrictive measures applied in both countries in 2020 led to sizable consumption dips in some months, with year-on-year (YoY) declines of over 5.5% and 3.3% in the Spanish and Portuguese system load, respectively.

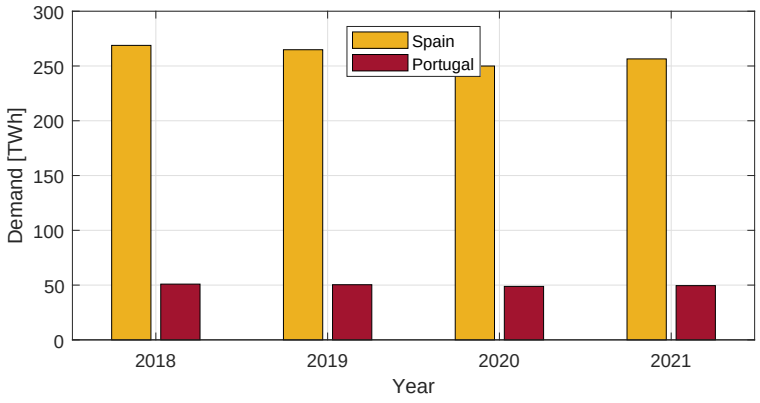


Figure 3.3: Annual Electricity Demand in TWh: the Peninsular Spanish system (yellow) and the Peninsular Portuguese system (red).

The year of 2021 witnessed a rebound from the bottoms of 2020, but the annual levels were still below the pre-pandemic values, particularly in the Spanish case. Thus, 2021 did not show any added “pressure” or profile change in demand (load) per se, in comparison to the previous years, that could explain the extent of the observed price hikes.

Now, moving to the generation component of the market pricing, Figure 3.4 showcases what was already seen in section 2.2 as a annual share evolution per generation technology, and here is portrayed as the balance (ratio) between renewables and non-renewables in the (aggregated/total) annual generation mix (bar chart, left-hand side y-axis). Clearly, the share of renewables grew appreciably in both countries, reaching approximately 47% and 65% in Spain and Portugal, respectively, in 2021 (from previous yearly shares of 40%

and 56% in 2018).

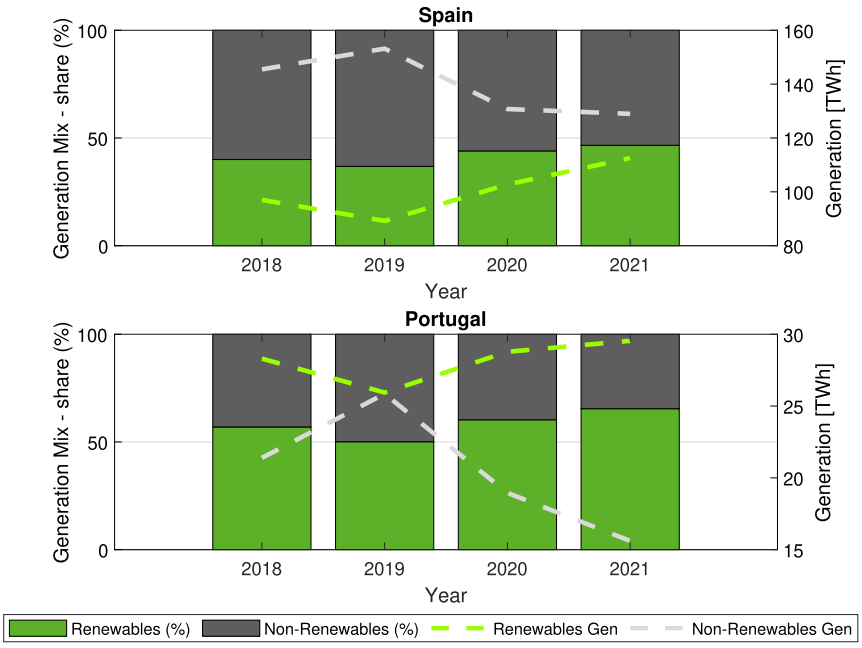


Figure 3.4: Annual Net Generation Mix – Share (%) of Renewables (green bar) and Non-Renewables (dark grey bar) [left axis] vs Cumulative Annual Generation (TWh) of Renewables (lime green dashed line) and Non-Renewables (grey dashed line) [right axis]: data for the Peninsular Spanish and Portuguese systems, respectively.

A scenario that is far from surprising, given the energy policy efforts and regulatory landscape we have been alluding to. This, in turn, has translated into sizable investments in both countries, particularly over the last decade, to significantly increase renewable capacity — as highlighted in Figure 3.5.

In 2021, renewables capacity had reached around 63 GW and 15 GW, in Spain and Portugal, respectively. This represents a 50% plus gain in renewable capacity (in both countries), compared with 2010. In contrast, the non-renewables installed capacity fell close to 24% and 54% in Spain and Portugal, respectively, during the same period. Roughly, this means that for each lost GW of non-renewable power, 1.99 GW and 1.13 GW of renewable capacity were added to the Spanish and Portuguese systems.

Interestingly, the cumulative annual generation of both types of generation, as shown in Figure 3.4 (dashed lines, right-hand side y-axis), tells a different story behind the falling share of non-renewables. In Spain, we see an almost perfect symmetry between the dashed lines, i.e., the 16.5 TWh drop from 2018 to 2021 was almost equaled (since we need to account for a smaller demand in 2021 vs 2018) by a 15.6 TWh increase in the renewable generation (in the same period). But, in Portugal, this symmetry does not hold. Here, even considering a slightly lower demand, the rise of 1.24 TWh in the cumulative annual generation of renewables did not offset the loss of 5.78 TWh in non-renewables capacity, suggesting a surge in terms of energy imports (interconnection use). That leads us to Figure 3.6, showcasing the annual cumulative energy trade for both countries, in terms of

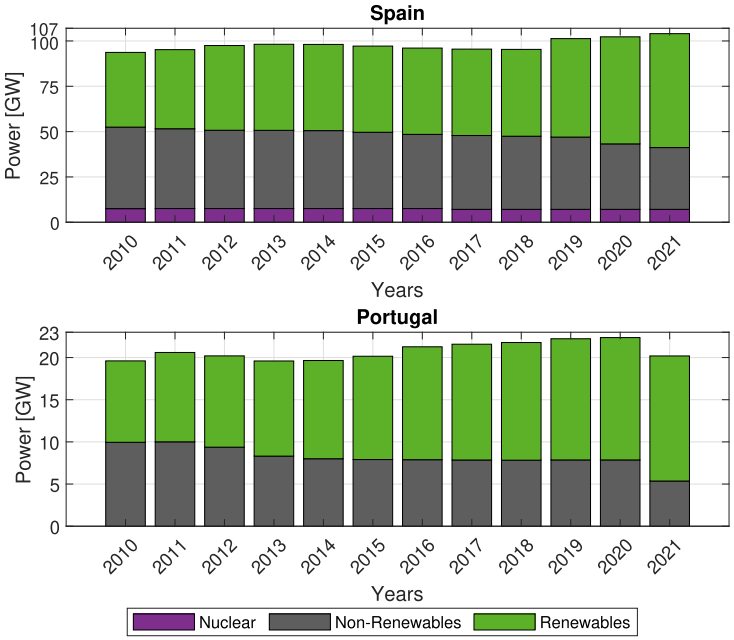


Figure 3.5: Evolution of the generation installed (available) capacity (GW) in the Iberian Peninsula: Nuclear (purple bar), Non-Renewables (dark grey bar), Renewables (green bar).

imports and exports. The two countries exhibit two very different scenarios, even when accounting for the inter-annual variability of renewables. While imports in Spain gradually decreased from the 2018 high to values around 17/18 TWh, the exports (negative signal) steadily increased, reaching a record of 16.5 TWh in 2021. The opposite occurred in Portugal, which not only exported less energy but, more importantly, has become more reliant on imports from Spain. The 3-fold increase in imports (8.92 TWh), in 2021 versus 2018, underscores this point and confirms the cumulative annual generation analysis provided earlier.

This inference is meaningful, and policymakers should realize that energy security needs to be the number one priority of the electric power system. As such, the continuous path of dismantling dispatchable generation sources must be accompanied by an increase in system flexibility, in its three main vectors, i.e., enhanced interconnections, a more complementary generation mix, and investments in localized storage capabilities. Otherwise, this over-dependence on interconnections, with residual levels of system flexibility (including spinning reserve), can potentially lead to more load shedding incidents, which ultimately result in power outages.

To obtain a fuller account of the relationship between some of these price-related variables, the Pearson correlation coefficient, $r \in [-1, 1]$, is calculated, providing a rough estimate of the existing linear dependencies. Making it an appropriate (bivariate) measure of the linear correlation between two different time-series. Correlation coefficients close to the upper and lower limit imply a strong correlation, while values around 0 signal a weak or non-existing correlation.

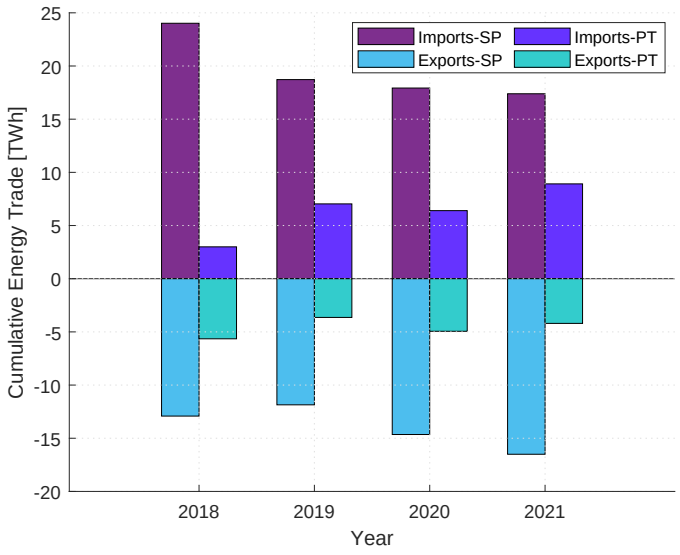


Figure 3.6: Annual Energy Trade in TWh: Peninsular Spain Imports (purple bar) and Exports (light-blue bar) versus Peninsular Portugal Imports (violet bar) and Exports (green bar).

Table 3.1 is therefore introduced to summarize the correlations between day-ahead electricity prices in MIBEL, from 2018 to 2021 (the latter was then divided into pre and post price limits changes), and the previously mentioned variables. These include the renewable generation volume from the three major RES, i.e., wind, hydro and solar PV; the non-renewable generation volume (mainly gas and coal); the pump consumption (storage capability); the respective Peninsular system load, for both countries alike; the cross-border physical balance, reflecting the intra- and inter-market import/export capacity; and a set of commodity-related variables that translate the market sentiment regarding key fuels for thermal technologies. The first was an indexed Iberian gas price (average weighted price of all the trades arranged for the same delivery date), followed by the "API2 Rotterdam Coal Futures" time-series (as an indicator of the medium-term price behavior of coal delivered to Europe), and finally, the Brent oil benchmark, which is seen as a global highly liquid exchange and is not only a reference for pricing petroleum products, but also a crucial (rough) measure of economic activity. The relationship with these variables is crucial to understand market price formation and the fine balance between supply and demand. Contrary to the former cumulative analysis, the correlation coefficients were computed based on the hourly datasets, which required adapting the time-series with different horizons.

In general, the correlation coefficients indicate a relationship of these exogenous variables with electricity prices, for all $r > 0.1$. In addition, seasonality in RES generation and the impacts of the health crisis during 2020 and 2021 accounts for some inter-annual variability, specifically in the Peninsular system load coefficients. Nevertheless, the last two columns also suggest the impact of a new approach to the market. Both Spanish and Portuguese non-renewable generation are the most consistently positively correlated vari-

Table 3.1: Bivariate correlation of related variables with the day-ahead electricity prices time-series.

Variable (scope) / Year	2018	2019	2020	2021
Main Generation: Peninsular Spain				
Renewables	-0.5	-0.26	-0.11	-0.24
Non-Renewables	0.74	0.57	0.64	0.64
Nuclear	0.18	0.19	0.49	0.12
Pump consumption	-0.72	-0.65	-0.59	-0.33
Main Generation: Peninsular Portugal				
Renewables	-0.37	-0.17	0.19	-0.18
Non-Renewables	0.75	0.63	0.75	0.42
Pump consumption	-0.56	-0.5	-0.48	-0.24
Load (Electricity)				
Peninsular Spain	0.4	0.53	0.64	0.11
Peninsular Portugal	0.28	0.46	0.6	0.2
Cross-border balance				
Spain-Portugal	-0.27	-0.2	-0.59	0.02
Spain-France	-0.08	-0.24	0.15	0.08
Other related variables				
Indexed Gas Prices - MIBGAS	0.59	0.62	0.56	0.93
Coal Futures	0.54	0.45	0.51	0.7
Brent Futures	0.22	-0.05	0.52	0.64

ables, with coefficients well above 0.5. On the one hand, this is consistent with market behavior where these thermal and costlier (peaker) plants tend to ultimately define the market clearing price. This also explains the high correlation with commodity variables, particularly the natural gas weighted price in the Iberian market.

An expected result, given the growing majority position of combined cycles plants in the thermal portfolio. On the other hand, non-renewable generation components were far less correlated and tended to show a negative value, revealing a negative downward pressure on spot electricity prices, as observed for the German market [282]. However, when we break this down into different components, the behavior of wind and PV is clearly different from the remainder, given the special regime benefit. Nuclear generation, with its constant output and stabilized marginal costs, presents one of the lowest correlations. Moreover, the pump consumption value is a clear indicator of excess generation and/or low electricity price periods, which explains its negative value and its importance in a system with increasing renewable penetration. Finally, the (net) cross-border physical balance indicated a modest benefit of interconnections to promote the flow of "low" price electricity.

Having analyzed the most relevant variables regarding the electricity price formation, Figure 3.7 depicts the daily average day-ahead electricity price in MIBEL ranging from 2018 to 2021. The plot perfectly illustrates the rise in prices at the beginning of 2021, which increased sharply after the regulatory changes to the day-ahead market price limits, on the 7th of July. The linear trend (approximation) is a good way to quantify the scale of this upswing. The 2018-2020 triennium revealed a slightly negative (-2.68%) slope, supporting the idea of relatively stable price behavior, while most of the downsize pressure

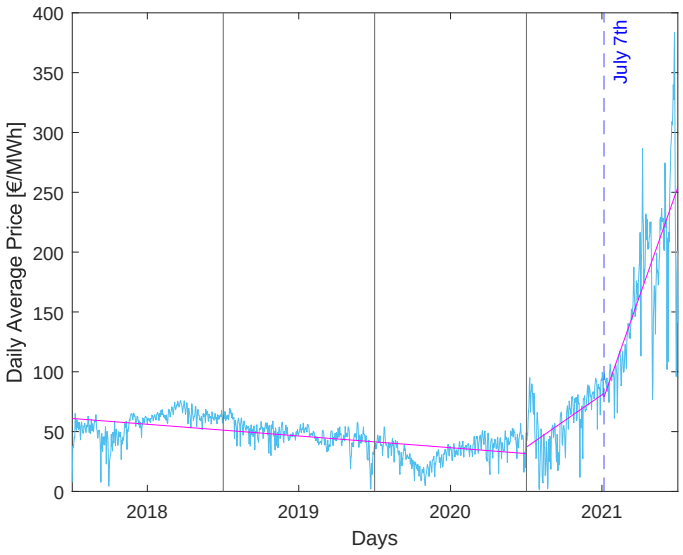


Figure 3.7: Wholesale day-ahead (daily) average price in MIBEL (sky-blue) from 2018 to 2021 (Spain prices in the market-spliting instances) and the linear fitting trends (pink).

was clearly an effect of the COVID-19 pandemic. Yet, early 2021 saw a clear inflection in prices, with a 24% slope from January 1st to July 6th, and then, when the new price regulation entered into force on July 7th, the slope increased fourfold (98%) by the end of the year, which in absolute terms represents an astounding 36-fold increase in the slope, compared with the 2018-2020 period. This meant that annual mean daily averages went from around 47-57 €/MWh to 112 €/MWh in 2021, i.e., doubling the price. The increase is even more notable if we consider the period after July 6th, with the mean daily average reaching 167 €/MWh. The analysis also shows exceptional fluctuation (measured by standard deviations), recording a staggering value of 75 €/MWh in 2021, overwhelming the previous (2018-2020) standard deviations in the range of 11-13 €/MWh, and even exceeding the previous annual mean daily average.

The statistical analysis of the day-ahead electricity prices and the respective main metrics concerning price formation indicate an escalation beyond the newly adopted price limits. To examine if and how market players changed their market bidding approach to the new rules, Figure 3.8 indicates the daily hourly data regarding buying and selling offers. This customized distribution bar plot compares the thirty days before and after the rule change, thus ensuring a close temporal proximity in terms of the other variables affecting prices.

The difference in total number of placed bids, both for buying or selling, in the two selected periods, was inconsequential, in the order of the hundreds of thousands of samples, with no noticeable connection with the new rules. Regarding the hourly buying offers, there was a clear reduction in the relative frequency of the classes up to 110 €/MWh for the post-rules period. In contrast, for the subsequent higher price classes, there was an increase in the relative frequency in this period, with a non-neglectable number of offers being placed in the price range above the previous limit of 180.3 €/MWh, including 2131 offers right

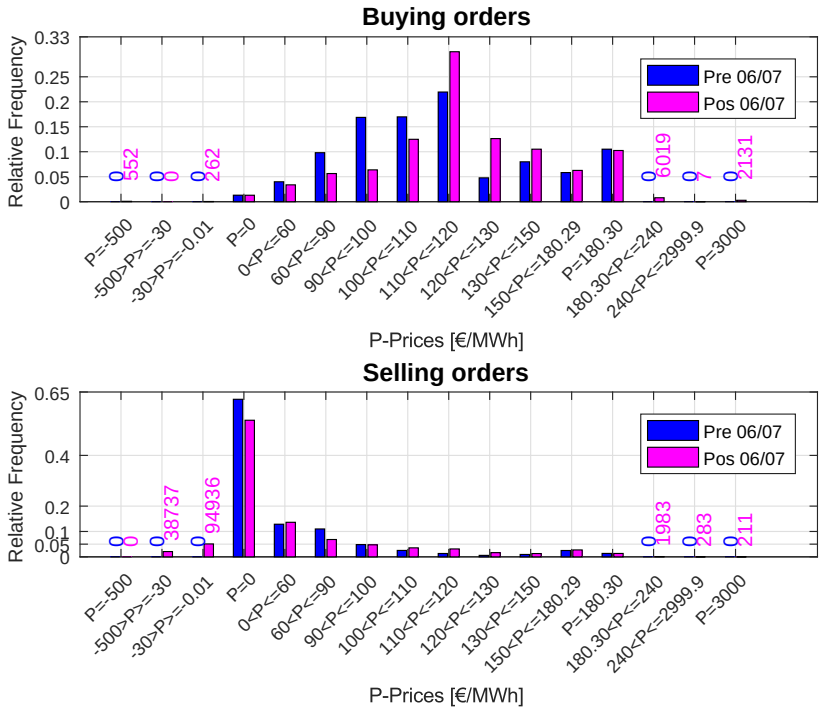


Figure 3.8: Relative frequency analysis of buying and selling day-ahead electricity market (hourly) orders, together with the absolute frequency, above the prices classes defined by the new price ruling: Month prior (07/06/2021-06/07/2021) in blue vs Month after (07/07/2021-05/08/2021) in magenta.

in the upper limit (as highlighted by the absolute frequency values placed above these classes, given the low relative frequency bar values). This explains the 12% increase in the average buying bid between the two periods, while the class of prices ranging from 110 to 120 €/MWh remained the most frequent. In terms of negative values, the number of occurrences was substantially lower. Interestingly, the class of prices exactly at the lower limit was reached very sporadically (552 occurrences). This indicates a more profitable way for buyers with load consumption and storage capabilities, namely PSH plants, to take advantage of high-RES generation periods, coupled with lower demands and a limited interconnections capacity.

A more stable picture emerges in the selling orders bar plot, with only a 2% increase in the average selling bid. The price class equal to 0 €/MWh remained the most prominent (above 50%), highlighting the volume of renewable generation operating in the system and its (largely subsidized) downward effect in the day-ahead price formation. As can also be seen in Figure 3.8, this has the potential to surge during high-RES periods, given the clear transference of selling orders down to the -30 €/MWh class and even further down to -500 €/MWh.

3.4 Discussion

The liberalization and deregulation of the electricity markets have promised to unleash the entrepreneurial potential of the power industry sector, with competitive advantages that should foster the lowest possible retail electricity prices [283]. Despite the sizable progress in transparency and openness, the electricity market's uniqueness, complexity, and multitude of dependencies introduce a set of ever-changing conditions and targets, which in turn have consistently led to a refinement of its predicates.

EU policies have undoubtedly contributed to a large share of renewables in the mix, and consequently, a reduction in CO_2 emissions. However, as we pointed out, decision makers often fail to also highlight the risks and some of the shortcomings. And if it is true that major events, like the COVID-19 pandemic and the war, were unpredictable – international externalities like geopolitical insecurity and its social costs are very difficult to objectify [284], it is also true that the focus of the regulatory changes was to strengthen market integration and the adoption of renewables, with a somewhat "blind" belief that the lower marginal operation costs of these technologies would drive the market prices down, with neglectable exceptions. As often the case, the "perfect" scenario does not exist, and the security of supply in the Iberian electric power system should not be taken for granted.

"Energy security" is not only worthy of discussion but should always earn center stage when considering deep policy implications. The current framework fits the short-term perspective definition of the term, which describes the disruptive impact on the system caused by an unexpected sharp rise in prices and/or supply cuts [285]. Traditionally, "energy security" was a three-dimensional concept, encompassing availability, affordability, and reliability. It has since evolved to also embrace its repercussions, i.e., its associated "footprint", coined as "environmental acceptability" [286], thereby considering the more indirect and harder-to-quantify long-term rippling effects on the earth's ecosystem. This represents a milestone, a positive answer to growing public worries, as already shown in Figure 3.2. From our viewpoint, the case of the European power system industrial complex, the missing linkage in this discussion, is encapsulated by the term "flexibility". This suggests a redefined concept, aligned with these five-dimensions, as illustrated in Figure 3.9.

This is especially true in the context of an electricity market with a large-scale adoption of variable RES, which is why it is important to have a stable policy framework, reflecting a long-term vision that prioritizes system flexibility investments [287]. In the near term, this implies a regulatory environment that supports security reserves or excess capacity at high(er)-capital facilities, which in a purely competitive environment (exclusively driven by the marginal costs) would not be ensured [284]. In fact, larger shares of renewables are directly linked to an (unsolicited) higher volatility in electricity prices [263], although, until 2019, there was a positive leverage effect in the wholesale electricity price in Portugal from the renewables contribution [288]. The same was also validated by the results in Table 3.1. In contrast, a continuation of the existing policy is incompatible with the

Figure 3.9: The evolution of the "energy security" concept in the context of the European power system's transformation.

required payments for non-renewable producers, and the price reduction gains do not cover a similar incentives framework [289]. Several studies indicate that a larger market coupling with France is an essential flexibility option that may help support the current Iberian policy, while driving the prices down [290, 291]. But it is still unclear what EU financial framework may support an expansion of the cross-border electricity trade infrastructure [292].

The historically high energy commodities price levels were the combined result of several demand and supply bottlenecks that squeezed the markets. Together with soaring carbon permits, they directly contributed to the higher electricity prices across the European continent. As such, IEA recognized the need for governments to take emergency measures, without compromising the long-haul commitment towards a cleaner and resilient energy system [293].

The seriousness of the abrupt rise in electricity prices all over the EU exposed some weaknesses that the regulatory changes failed to address/anticipate. This was also valid for the Iberian market. Due to a higher dependence on gas-fired generation, alongside severe weather conditions for the non-dispatchable renewable capacity, the Spanish government was forced to request the reactivation of some coal power plants, and to prolong their lifetime at least until the end of 2022. A paradoxical situation since it came approximately at the same time as the Portuguese ceased their coal generation in Pego. Per this request, Endesa, which is also one of Pego's thermal plant shareholders, brought the country's largest coal plant, As Pontes in Coruña (1400 MW), back online. This decision came after Endesa had already backtracked its decision to shut down operations at the thermal plant of Carboneras (early 2021). EDP Spain joined Endesa's move and also decided to reopen Los Barrios plant (567.5 MW). This move to reopen coal utilities was a common theme in Eu-

rope, and was grounded on the soaring wholesale gas prices, with gas being the preferred thermal dispatchable technology in several EU countries. This scenario of high gas prices led to a shift in short-term marginal costs favoring coal to the detriment of gas-fired power generation, even when accounting for the singular burden of coal plants, in terms of CO2 license costs. Moreover, the European electricity sector recorded considerable negative impacts from the mentioned externalities, in particular renewable companies due to their riskier investment profile [294].

In an unprecedented move, the Iberian countries reached an agreement with the EU commission (first quarter of 2022) for an exceptional price cap mechanism. This was only possible given the "energy island" status of this western European region. Limiting the price of the gas used to produce electricity, helped contain the marginal costs of the combined cycle plants, and consequently lowered their bidding price. This market intervention underscores a lack of alternatives in terms of system flexibility and exposes the problem of intermittency and variability of the renewable policy. The harsh short- to mid-term outlook in terms of gas and coal prices also questions the phase-out strategy of the Spanish nuclear power plants, given its critical baseload role and lack of CO2 emissions, in a period where system flexibility is not increasing. This has relevant policy implications. Unlike the Spanish (substantial) generation portfolio, the Portuguese electric power system does not have an extensive structure to provide a backup with Spain via conventional sources. As such, it has an inner difficulty in handling the intermittence of renewables [295].

Additionally, negative spot market occurrences have increased significantly during the pandemic years (reduced demand), as is the case of the German market [296]. This indicates a further strain on market operations and some of its already affected players. Although the Iberian market was less prone to these occurrences, the regulatory changes and the macro environment also contributed to an increase in these events, particularly in the intra-day sessions, together with a decline in the market-splitting occurrences [114].

Furthermore, the commercialization side of the electricity business was also severely disrupted, with several smaller market players being unable to continue their operation (fulfill their signed commitments) in the liberalized market in the presence of such high prices. This was one of the risks listed by CNMC, when hearing market participants in the first audience, as described in CNMC's resolution 8362. This market volatility erased the already "thin" margins in which some of these players have been operating (before the regulatory changes in terms of bidding price limits). In the face of this problem, ERSE promoted a controlled exit of financially impaired retailers and decided to reallocate their consumers to the regulated market (EDP-Serviço Universal), as announced in the press release concerning the series of exceptional measures to tackle this situation [297]. In Spain, the consequences of escalating prices were also felt, with serious economic impacts. For instance, the Spanish authorities were forced to alter the pricing formation mechanism in the regulated market since, unlike the Portuguese case, the PVPC tariff was much more prone to immediately reflect market variations to the final consumer. There is a real risk of aggravating energy poverty in the Iberian Peninsula, in an environment where demand

has been highly price-elastic lately [298]. In Portugal, the government and ERSE often use the Tariff Deficit to partially absorb the hikes. Lastly, policymakers should focus on growth-enhancing energy policies that promote energy efficiency usage [299].

Chapter 4

Optimization

The minimization (or maximization) of a cost or an objective function is a key task in the vast landscape of power systems engineering real-life problems. As is the case of the OPF problem, as we saw, and many others, including the economic dispatch of power systems, steady-state security analysis, operation of a hybrid energy system, the fine-tuning of certain model hyperparameters, the learning process of neural networks, among others. Often times, depending on the problem specificities certain compromises and simplifications must be made in order to prioritize certain aspects and/or to effectively achieve feasible solutions with a tolerable computational cost. As such, this chapter lays the foundation for optimization by covering a range of topics — from introductory concepts, associated notations, broad classification of algorithms based on various parameters, constrained optimization and penalty methods to the multiplication of metaheuristics.

Concerning expensive and global optimization problems, a growing trend has been the use of evolutionary algorithms assisted by surrogate models. By providing an alternative to the expensive fitness evaluations, effectively limit the computational burden, while the merits of the approximation methods (surrogate building) and the search space mechanisms translate into a good performance. As such, the exploration capabilities of the fairly recently developed Slime Mould Algorithm, a metaheuristic with only one tunning parameter and that does not rely on a personal best information of each swarm agent, is cherrypicked to develop its surrogate assisted counterpart, i.e., surrogate assisted slime mould algorithm (SASMA), with an original database management strategy and surrogate building mechanism. To confirm the effectiveness of the proposed approach and test its versatility, SASMA is tested on a set of well-known benchmark mathematical functions, for 30 and 100 dimensions. The proposed SASMA outperformed many of the selected benchmark algorithms on most of the test functions.

The knowledge and experience gained through the conceptualization, development, and testing of the surrogate-assisted metaheuristic global optimization algorithm extend well beyond the scope of the case study itself. The detailed processes involved — including programming aspects — are crucial for understanding the methodologies presented in the subsequent chapters, particularly those applied to solving optimization problems in forecasting tasks.

4.1 Background

Since ancient times, the concept of optimality was always tied with the development of human civilization. In broad terms it translates the most effective, yet possible, way of conducting a task under a certain set of plainly expressed or implied restrictions. Curi-

ously, unlike other scientific topics, “optimization” was taught without a mathematical formality, which only came much later, namely with the development of calculus in the 17th century. With this regard, the work of Pierre de Fermat, a French mathematician, that proposed a general approach to obtain the local extremum points of a differentiable function, by setting its derivative to zero, is nowadays considered a landmark in the history of optimization. Perfectly illustrating why this concept is deeply intertwined with minimization and maximization operations. This breakthrough was then followed by the work of Euler on the calculus of variations, which was later extended by Langrange, in the 18th century, from which the ”Euler–Lagrange equations” were derived [300]. Later on, and already in the mid-twentieth century, the term ”mathematical programming” was in fact interchangeable with the concept of optimization [301], much due to works on linear programming, namely by Dantzig, the proponent of the simplex method, but also by Karush, Kuhn and Tucker in the development of what is refereed to as ”KKT conditions” (a generalization of the Lagrange multiplier approach, and a framework for the solution of nonlinear programming). These are often presented as the major milestones in the field development [302].

The vast reach and interdisciplinary scope of optimization makes it a vital topic in modern engineering. Nevertheless, by nature, optimization is not “the end, per se” but rather “the means to”. First, it is important to realize that problem modeling precedes the optimization stage itself and it is no less important. It concerns the accurate identification of (involved) variables and constraints, as well as an adequate formulation of the targeted objective [301]. With that, we are then able to carry out the optimization, and recognize whether we have been successful or not in reaching a solution. The decision to impose certain weights, parameters and or targets is assumed by (a) subjective individual(s), where their/its interpretation of the optimization result, of the existing margins, the problem asymmetries, simplifications, and other operational aspects, is key, and it is ultimately based on a perception of “experience”. This ”fuzzy” concept relates to sensitivity analysis, where we aim to refine the solution or at least identify the main drivers for optimization improvements (as measured by the objective function). That is, determine the most impactful problem dimensions (in the solution vector) and modeling characteristics.

In terms of the EPSs, these ubiquitous optimization techniques are key to aid/better inform the decision-making processes of utilities, system operators, energy regulators, policy makers, market players and many other parties, in their quest for a more efficient, reliable, cost-effective, environmentally friendly and regulatory compliant, operation of the different system components, for a wide-range of time-horizons. Long-term planning and strategic decisions (spanning over years), as is the case of regional, federal and transnational multi-year energy and grid infrastructure roadmaps, which are, by nature, transformative, and often linked with sizable investments, planned phase-outs and regulatory changes; medium-term planning (from several weeks to months) such as generation planning, specific rotation (maintenance) schemes, inventory management strategies, namely for thermal generation units, seasonal market positioning in terms of energy

players, among others; and perhaps the more notorious, short-term decisions (hours to several days or even a couple of weeks), including load, electricity prices and power generation forecasting tasks (which can also be of interest for the medium-term horizon), but more notably the dispatch and unit-commitment problems [303], which as we saw in chapter 2, are the backbone optimization tasks, behind the iron-cloud pledges of any EPS, namely the power balance at all times.

Overall, these two (last) recurrent type of optimization problems, involve the bounded minimization of nonlinear, sometimes convex, and often times separable functions, subject to (at least one) linear demand constraint and with gradient methods emerging as the preferred choice to solve them. The dual form of this problem provided a breakthrough in the solution for very large-scale dispatches, where the number of decision variables is sizable. Besides, it is important to notice two things: i) often times demand constraints are relaxed due to several practical reasons; ii) single-objective optimization tends to be "blindsided" to a multiplicity of other correlated variables, which can also be weighted (differently or evenly) depending on the decision maker. These encompass two main issues in optimization, namely, how to handle hard limits placed on the value of independent variables or on the value of relations between those variables; as well as the identification and selection of non-dominated solutions [304]. Leading us, on the one hand, to the topic of constrained optimization (section 4.3) and, on the other, to explore multi-objective optimization theory (section 4.4).

4.2 Classification of Optimization Algorithms

No single optimization algorithm is tailored for all type of problems, and so, some criteria that reflect the inner nature and complexity of the problem, are commonly used to decide upon which algorithm to consider. Namely: i) type of variables; ii) the characteristics of the objective function (e.g. convex or non-convex); iii) the number and type of constraints, among many others [305]. These algorithms often require multiple (objective) function evaluations to ensure convergence, and so they are typically iterative algorithms. For instance, the options to solve a non-linear optimization problem, i.e., a problem where either the objective function, the feasible region (as determined by the constraints), or both, are non-linear [306], are different, from the options for a more straightforward linear problem (where both constraints and the objective function are linear). But more importantly, and it is a key concept surrounding optimization, is the distinction between convex and non-convex problems.

For a convex function (U-shaped for a univariate function), any local minimum is also a global minimum, as illustrated in Figure 4.1 a). Thus, removing the need to search for multiple minima and so one can easily understand why these functions are preferable when targeting minimization problems. Formally, one can define an arbitrary real-valued multivariate function, $f : \Re^n \mapsto \Re$, as convex, if its domain $(\text{dom } f)$ forms a convex set (that is, if, for any two points within this subset of the Euclidean space, the line segment connecting

them, lies entirely within the function domain), and provided that for all $x_1, x_2 \in \text{dom } f$ and $\theta \in [0, 1]$, we have:

$$f\left(\theta x_1 + (1-\theta)\, x_2\right) \leq \theta f\left(x_1\right) + (1-\theta)\, f\left(x_2\right) \quad (4.1)$$

From a geometrical standpoint, if a line segment between any two points on the graph, as shown in Figure 4.1 a), lies either on the curve itself or above it (the epigraph of the function), f is deemed as convex. This particular region of space, forms a half-plane, for univariate functions, while for multivariate functions, establishes a hypersurface. All vector norms (magnitudes of quantities), twice differentiable single variable functions with positive second derivatives, as is the case of quadratic cost functions of generators or the simplistic quadratic power loss $\propto I^2$ (assuming a fixed impedance condition), but also some common loss functions in machine learning theory (Log Loss, L1 regularization, among others) are all noteworthy examples of convex functions, that are often found in EPSs related optimization problems. For a typical constrained minimization problem as formulated in 4.4, the same is considered to be convex if the objective objection is convex, the equality constraints are affine and the inequality constraints form a convex set.

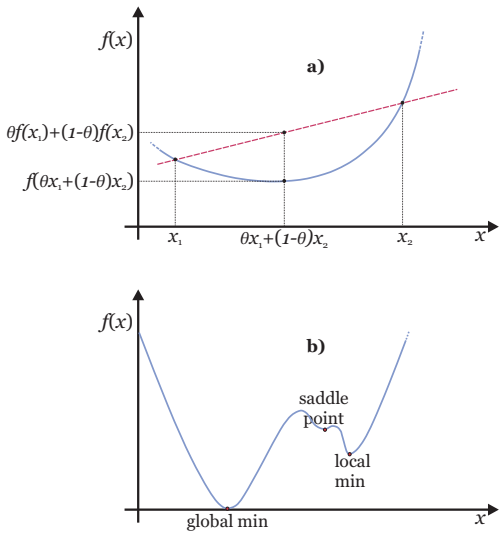


Figure 4.1: Optimization functions: a) convex function; b) non-convex and non-concave function.

Conversely, an arbitrary function, $f : \Re^n \mapsto \Re$, is called concave, if $-f$ is convex. By simply inverting the sign of the inequality given in equation 4.1, we obtain the formal definition of a concave function. As such, these functions have no local maximum that is not global as well, which is a great trait for maximization purposes instead.

In practice, many objective functions are neither convex or concave, and so multiple local minima and maxima may be encountered, with significant degrees of varying curvatures, thus, complicating the task of the optimization algorithms, which in turn may end-up trapped in saddle points (a type of critical point that is neither a local minimum nor a

maximum, but like the others, also has a horizontal or flat tangent plane to the function's surface at this point), as shown in Figure 4.1 b). Interestingly, affine functions, i.e., translated linear functions, are both convex and concave, but not strictly so.

To fully grasp the concept of local and global minimum, the following notation is developed: i) considering $x = (x_1, x_2, ..., x_{n-1}, x_n)$ a n-dimensional feasible solution (vector in tuple form) to the minimization of the objective function, $f(x)$; ii) with $\varepsilon > 0$ being a certain n-dimensional neighborhood of x where every feasible point in the neighboring region is (sufficiently) close to it; iii) x is a local minimum if the condition, $f(x) \leq f(x_{\text{feas}})$ only holds true for the neighboring domain, $x_{\text{feas}} \in [x - \varepsilon, x + \varepsilon]$; iv) x is a global minimum if the same condition, holds for every feasible solution, i.e., over the entire dom f; v) local and global maxima can be defined analogously; vi) hence, the concept of a localized solution is extremely important when dealing with non-linear programming problems, where algorithms (particularly gradient descent based ones) can be near-sighted (mainly due to local information dependence, step-size issues and high dimensionality sensitivity) and so, wrongfully assume saddle points as near optimal solutions, when in fact they constitute sub-optimal ones.

To conduct the minima search with these more intricate functions, often times non-convex functions are "convexified", by approximating the original problem with a simpler convex one, and so leveraging on the well-defined properties of these functions, in order to more easily find its optimal solution. The approximated convex objective function, $g(x)$, is bounded by the original function epigraph, such that $g(x) \leq f(x)$ for all $x \in \text{dom} f$. In turn, this implies that by minimizing $g(x)$, we obtain a lower bound (underestimate) of the original global minimum. Convex relaxation is a widespread technique used across many fields, including in machine learning. For instance, support vector machines (SVMs), which are a powerful tool for classification tasks, often times deal with an original problem that is non-convex due to the complexity of the input data, and so convex relaxation is used to always formulate a convex quadratic problem [307].

To further simplify the problem, and beyond the convex relaxation of the objective function, a common practice is to relax its hard constraints (optimization domain region), while trying to preserve the original bounds. For instance, in unit commitment problems hard constraints like the power balance and ramp rates can be relaxed by incorporating them into the objective function itself, with the respective Lagrange multipliers [308]. Another example is the relaxation of integer constraints to allow for continuous variables (i.e. an easier linear programming problem), and then resort to rounding or branch-and-bound techniques to find an integer solution to the original integer or mixed-integer programming problems [301].

Continuity of the objective function(s), is also important when employing optimization algorithms, given that the function smoothness ensures a predictable behavior in terms of the output for a given small change in the input(s). Discontinuities instead, can cause algorithms to converge to sub-optimal solutions or even fail to converge at all. In formal terms, an arbitrary single variable function $f(x)$ is continuous at $x = a$, if $\lim_{x \to a} f(x) =$

$f(a)$. More broadly, differentiability at a given point, necessarily entails continuity, yet continuity does not invariably guarantee differentiability (a situation that occurs when the approaching left and right-hand limits differ). The smoothness of the function f, is also a key trait, and it reflects the degree of differentiability. In fact, for non-convex functions, one would wrongfully think that in order to find out whether a given point is indeed a local minimum, we would need to inspect all the points in the immediate vicinity. Yet, in the presence of a smooth function (at least twice continuously differentiable), the local minimizer may be confirmed, by examining the function gradient and Hessian (first and second-order conditions as derived from the Taylor's theorem) [301].

So, depending on the solution space being discrete or continuous, a common classification divides optimization algorithms in continuous and combinatorial/discrete [309]. With integer programming problems being an example of the latter, which by nature tend to be more difficult to solve, given the non-convexity exhibited by the feasible sets. A division between deterministic and stochastic algorithms, i.e., algorithms that, on the one hand, ensure that for the same input it always produces the same output (following a certain set of well-defined operations and data assumptions), whilst on the other hand, stochastic algorithms produce different results (often by slight margins) for the same starting input conditions, given the uncertainty associated with both the model assumptions and the inherent probability-based operations (algorithm heuristics). This classification is closely aligned with another, where exact and approximate optimization methods are separated [310]. Where the first tend to be more efficient for smaller problems, while the latter, which includes the subsequently explored metaheuristic algorithms, tends to be more fit to obtain near optimal solutions for large-scale problems [305]. In addition, despite its more qualified scope (a connection with machine learning algorithms), stochastic learning algorithms, which can be split into supervised and unsupervised approaches, can also be included in the generic classification provided in Figure 4.2, as based on [311].

Yet, beyond these traits, there are numerous others that lead to different classifications and sub-classifications of optimization algorithms. For instance, one can establish a traditional distinction between linear and nonlinear optimization problems. In linear optimization, both the objective function and the constraints are linear. In their standard forms, the Simplex method and its variants, as well as interior-point methods — both deterministic in nature — are the leading solvers for these type problems. For non-linear problems, which are often non-convex — as is the case with the optimization problems studied in this thesis — a common sub-classification of interest concerns global versus local search methods.

Regarding local search methods, a prominent category is gradient-based methods, which are typically considered deterministic — especially due to the early development and in cases where the objective function is not overly complex and a full, sufficient batch of data is available. Examples include Newton's method and conjugate gradient methods. However, in large-scale optimization problems, such as fine-tuning machine learning models, local methods are often stochastic. This is because the gradient is approximated using a

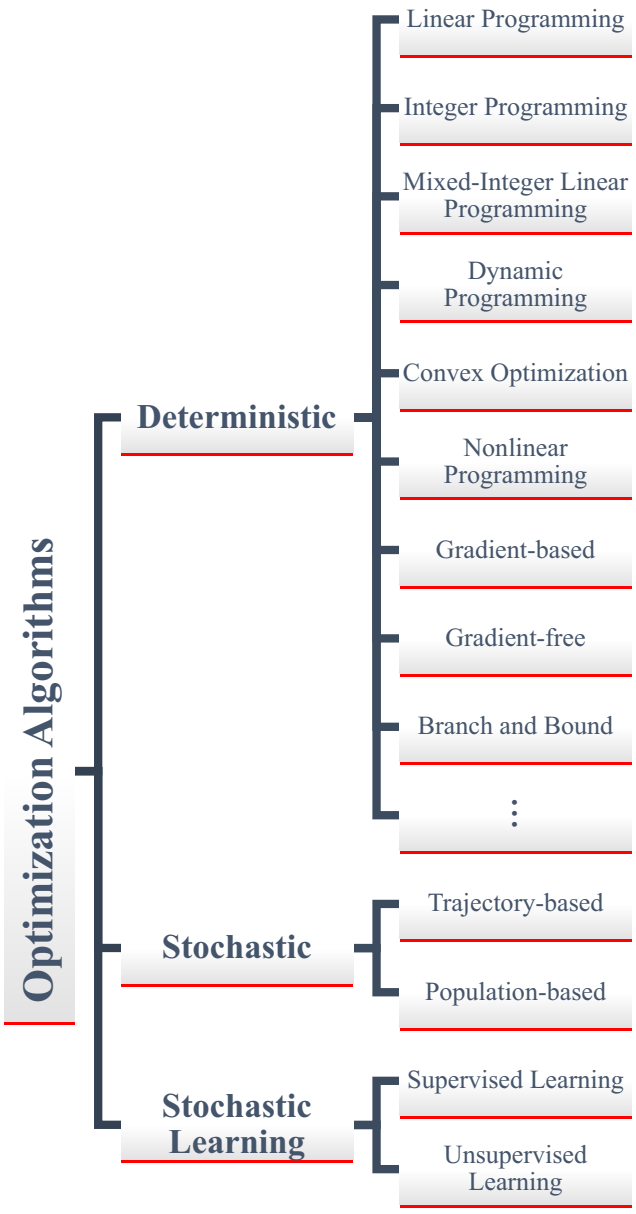


Figure 4.2: Hierarchical classification of optimization algorithms.

subset (often a random mini-batch) of the dataset. Stochastic gradient descent methods are perhaps the most well-known example.

Many local optimization methods are proven to reach — or at least approach with high probability, in the case of stochastic variants—the vicinity of a minimum for strictly convex functions. This success is largely due to how effectively gradient information can be exploited to rapidly descend toward local optima, particularly in scalable and relatively smooth optimization landscapes that exhibit some degree of regularity. Simultaneously, many practical nonlinear and non-convex optimization problems exhibit the property that most local minima are also approximate global minima [312]. Methods from both ends of this spectrum will be further explored in the machine learning theory chapter, specifically in section 5.3.2.1.

Despite this, the reduced effectiveness of local methods in rough or non-smooth environments — where gradients are unreliable or misleading — was a key motivation for the development and widespread use of global search methods. These methods are mostly stochastic and can be either single-solution and trajectory-based, such as Simulated Annealing, or population-based, as is the case with most metaheuristic algorithms (the subject of sub-section 4.2.2).

To conclude this hierarchical and qualitative overview of optimization algorithms, it is important to recognize that this is a very vast field with no universal solution. Instead, it comprises a diverse collection of algorithms, each tailored to specific types of optimization problems [301].

4.2.1 Multivariate optimization conditions and critical points

The analytical identification and classification of critical points, which are potential candidate locations for local minima or maxima, is crucial in optimization theory for finding optimal solutions across various domains. In this thesis, we will focus on its application to neural network training, specifically in the context of gradient descent optimization methods. These methods rely on the first and possibly second-order necessary conditions to navigate the loss surface and find the optimal set of neural network weights that minimize the loss function (error).

Considering a generic multivariate function $f\left(x_1, x_2, \ldots, x_{n_{\text{dims}}}\right)$ where $f : \mathbb{R}^{n_{\text{dims}}} \to \mathbb{R}$, with an n_{dims}-dimensional input (domain) space, the critical points—either local minima, local maxima, or saddle points—can be found using the first-order condition. This involves setting the gradient vector (all partial derivatives) equal to zero:

$$\nabla f = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n}\right) = \mathbf{0}. \quad (4.2)$$

If there is a discontinuity at a given point, it may be impossible to determine one or several partial derivatives, making the gradient vector, ∇f, undefined at that point. Nevertheless, despite the fact that analytic methods cannot be applied, such points are still considered critical points because they represent an abrupt change in the function's behavior.

Moving on, the second-order condition is used to determine the nature of the previously computed critical point(s), namely through the Hessian matrix H (equation 4.3). In here it is useful to think of the Taylor series expansion around this critical point, and so one can understand that for small perturbations ε, there is a dominance of the second-order term, which in its usual asymptotic notation can be denoted as $O(\varepsilon^2) > O(\varepsilon^3)$. That is, the relative growth rates of higher-order terms are much smaller, hence why extrema are primarily determined by the sign of the second-order derivative.

$$H = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}. \quad (4.3)$$

Similarly to the univariate case, where the second-order condition helps us determine the function's curvature (concave down or concave up for negative and positive second derivative values at the critical point, respectively); the diagonal terms of the Hessian matrix represent the curvature of the function along each independent variable axis (how the function curves in the direction of each variable). The off-diagonal terms ("mixed" partial derivatives of the function taken in pairs) describe how the rate of change of the slope in one direction is affected by changes in the other direction, and vice versa (how the curvature changes when two different variables are varied simultaneously).
Having computed the Hessian matrix, the critical point is a local minimum if H is positive definite. Conversely, if H is negative definite, then the function has a local maximum at that point. Additionally, if H is indefinite, the critical point is actually a saddle point (a point where the function curves upwards in one direction and downwards in another). The nature of the Hessian matrix quadratic form (${\mathbf{v_{any}}}^T \, H \, \mathbf{v_{any}}, \forall \; \mathbf{v_{any}} \neq \mathbf{0}$), and a related measure for its stability and convexity traits, the concept of definiteness in linear algebra, assesses both the symmetry and the sign of Hessian matrix eigenvalues. In practice, we check if H is symmetric ($H = H^T$) and determine all the leading principal minors of H (determinants of the leading principal minors (top-left sub-matrices) of all dimensions from 1×1 up to $n \times n$, i.e., all the determinants up until covering the entire matrix). Then one can state:

- H is positive definite: if H is symmetric and all the leading principal minors (determinants) or eigenvalues are positive (${\mathbf{v_{any}}}^T H \mathbf{v_{any}} > 0$);
- H is negative definite: if H is symmetric and the leading principal minors (determinants) alternate in sign (starting with a negative value) or all eigenvalues are negative (${\mathbf{v_{any}}}^T \, H \, \mathbf{v_{any}} < 0$).

For the situations where the Hessian does not provide a conclusive result (indefinite or singular), higher-order partial derivatives (conditions) and higher-order terms in the Taylor series expansion of the function around the critical point can provide additional information to identify the nature of the critical point. For instance, for the simpler univariate function $f(x) = x^3$, faced with an inconclusive second derivative value (condition) at the critical point ($f''(0) = 0$), the positive third derivative ($f'''(0) = 6$) indicates that the critical point is an inflection point (a transition from concave down to concave up curvature). The modulus function, $f(x) = |x|$, is another interesting example where not only the first and second derivatives are undefined at the critical point, but even higher-order derivatives do not help (they are not always applicable). These instances highlight a cardinal

aspect of multivariate optimization, which indicates that when both the first-order and second-order conditions are fulfilled, they are sufficient to identify a local extremum but not necessary. A local extremum can exist even in their absence. So, for non-differentiable functions, the alternative approaches include sub-differential analysis, piecewise definitions, and convexity properties.

4.2.2 Metaheuristics

From the previous sub-sections, the concept of optimization emerges as the process of selecting the best solution from often uncountable alternatives [310]. Global optimization problems are typically characterized by a complex landscape of local minima, making the identification of the global optimum far from trivial. Unlike local optima, which exhibit distinct behavior based on gradients and Hessians, the global optimum does not necessarily have a unique local signature that differentiates it in isolation. Instead, achieving global optimality requires an understanding of the collective behavior of all local solutions across the domain, making it a computationally demanding challenge [313].
Faced with this reality, iterative approaches have been developed over the years to traverse the optimization domain using search heuristics (a set of rules). These methods begin with a set of candidate solutions, often initialized through pseudo-random positioning, forming an evolving algorithmic population with or without memory capabilities, in a trial and error type of optimization procedure.
These techniques, commonly referred to as metaheuristic algorithms, which is a way to signal a ”deeper” or “notably greater level” heuristics, are largely inspired by real, natural or logical phenomena [314]. Many rely on a diverse set of individual behaviors to enhance exploration, ultimately benefiting the broader solution space—a concept closely linked to swarm intelligence processes. This analogy justifies the use of the term ”population,” where solutions act as the search agents of the algorithm.
However, a population of search agents is not a mandatory requirement for an algorithm to be classified as a metaheuristic. Some metaheuristics follow a single-solution search approach, such as the Tabu Search algorithm, as discussed in Section 4.2.
A common feature of these ”meta” heuristics is their reliance on subordinate heuristics—mechanisms that guide the search process to overcome local sub-optimality. Additionally, some approaches employ higher-level global procedures, where the positioning of the entire population influences the search dynamics [315].
A brief search in academic databases reveals that metaheuristic algorithms have risen to prominence in optimization research, particularly due to their effectiveness in addressing non-trivial and non-convex problems. The proliferation of population-based metaheuristic optimization algorithms, a category of stochastic optimization, has been a dominant trend in mathematical and computational optimization over the past few decades [316].
The working principle of these algorithms is to explore subsets of the search space domain, which otherwise is too large to be fully surveyed, using a population of potential solutions rather than a single candidate solution, guided by a series of heuristics that hopefully lead

the search towards an optimal solution, with fewer computational cost than other traditional optimization algorithms, like iterative methods [317]. Furthermore, unlike many traditional methods, metaheuristic algorithms can also handle non-continuous problems. Nevertheless, depending on the problem complexity and the specific algorithm search space mechanisms, namely the corresponding convergence properties, there is no guarantee that the global optimal solution(s) can always be found [311].

In this crowded field, swarm intelligence and evolutionary algorithms (EAs) are among the most popular choices, given their ability to improve several candidate solutions, by working in decentralized manner, mirroring diversified types of cooperative strategies observed in nature [318].

Typically, depending on the origin of the metaheuristic or local search strategy — defined by the nature of the agents and their proxy mechanisms—metaheuristics are often classified into four broad categories [319, 320]: i) Evolutionary — mostly based on genetic operators and differential mechanisms; ii) Swarm-based — inspired by the collective behavior of living organisms (ranging from the most simple bacteria to the most complex animal species); iii) Human-based — driven by common human behaviors and perceptions; iv) Physics-based — modeled after physical or chemical processes and their respective governing equations. This classification is illustrated in the taxonomy tree in Figure 4.3.

In addition, many of these population-based metaheuristics rely on a set of randomly generated variables (within the set of heuristics/strategies) [321]. This stochastic character (randomness) has proven to be effective in the search space task, by avoiding an early concentration of search agents in the same space regions, therefore mitigating the possibilities of being trapped in local optima. From the wide-range of options, particle swarm optimization (PSO) [322], cuckoo search (CS) [323], differential evolution (DE) [324], whale optimization algorithm (WOA) [325], gray wolf optimization algorithm (GWO) [326] are amidst the preferred choices to solve many engineering problems. Another feature of these algorithms is the influence of the different control parameters and the population size itself, which can great influence the global search capabilities [327], and consequently, together with the stopping criteria, determine the convergence rate towards the optimal solution.

The search space is surveyed by the population of these metaheuristic algorithms, by using mechanisms that balance exploration and exploitation of the domain space [328], until the desired stopping criteria is met, and is typically defined by a fixed number of fitness evaluations (FEs). Still, the required number of FEs can be considerable, when the algorithm heuristics actively explore different regions of space to avoid a premature convergence (exploration), before engaging in the exploitation of potential optimal solutions, thus, encompassing a significant computation cost when dealing with expensive optimization problems.

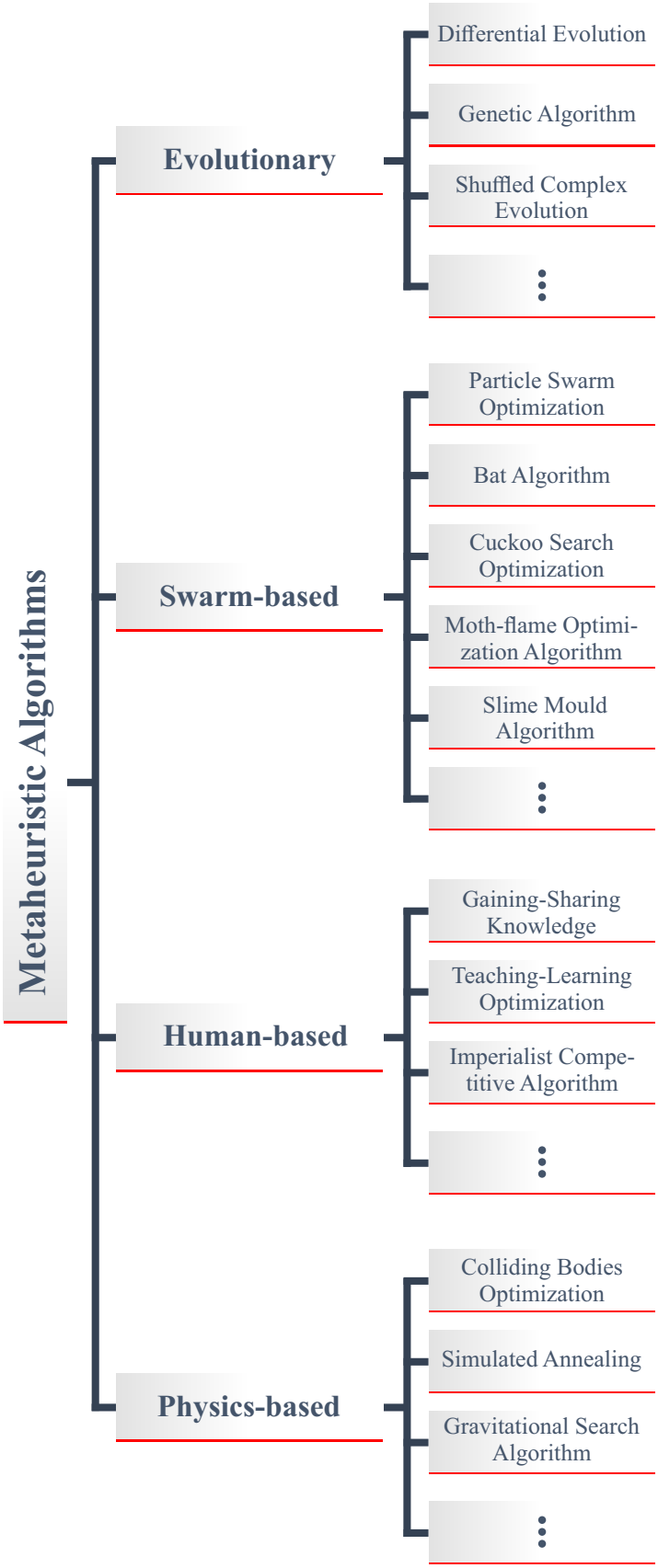


Figure 4.3: Classification of metaheuristic algorithms.

4.3 Constrained problems and the penalty method

Constrained optimization concerns the minimization or maximization of an objective function, $OF(x)$—remembering that $\min OF(x) = -\max -OF(x)$ —with respect to some decision/control variables, x, in the presence of constraints or restrictions on those variables. These constraints represent the search space limits, and they are often individually applied to a certain problem dimension, d, more precisely to its respective variable, x_d, i.e., element-wise bounds, $lb_d \leq x_d \leq ub_d$, where lb_d and ub_d denote those lower and upper limits, respectively, per dimension. Yet, more broadly, optimization problems, often involve equality constraints, $f_i(x) = 0$ — which are hard to cope with for any optimization algorithm including for metaheuristics, given that they decrease the feasible search space (optimum must lie on the intersection of all equality constraints) [329] — and/or inequality constraints, $g_j(x) \leq 0$, that establish linear or non-linear relationships between the different variables (problem dimensions), and so when considering metaheuristics for the optimization, the otherwise sufficient, use of velocity clamping or position update confinement techniques, as suggested in [330], for single variable constraints does not ensure (per se) that a feasible solution is returned.

Considering a generic constrained minimization problem (formulation) P_{co} :

$$P_{co}: \quad \underset{x}{\text{minimize}} \quad OF(x) \qquad \text{s.t.} \quad \begin{cases} f_i(x) = 0, & \forall\, i \in [1, n_{eqc}] \\ g_j(x) \leq 0, & \forall\, j \in [1, n_{ineqc}] \\ x \in \mathbb{R}^{n_{dims}} \end{cases} \tag{4.4}$$

where i and j are finite integer index sets; n_{eqc} and n_{ineqc}, are the upper limits of these sets, or more simply put, the number of equality and inequality constraints, respectively; while n_{dims} is the number of problem dimensions and so, $x \in \mathbb{R}^{n_{dims}}$ forms a convex set where both $f_i(x)$ and $g_j(x)$ are defined [331].

When approaching multivariate constrained optimization (typically via an unconstrained Lagrangian expression, similarly to the penalty approach presented in equation 4.6), the first-order and second-order conditions presented in sub-section 4.2.1 still hold. However, we now specifically look for points where the gradients of the objective function and the constraint functions are parallel. This means that at these critical points, the gradient of the objective function is a linear combination of the gradients of the constraint functions. Then, by examining the Hessian of this Lagrangian equation, the definiteness sign determines the nature of these critical points.

Penalty methods provide an effective manner to solve P_{co}, by solving a sequence of modified unconstrained optimization problems, i.e., the feasible (problem) region is expanded to the entire $\mathbb{R}^{n_{dims}}$, where a sufficiently large non-smooth penalty term is added to $OF(x)$ for solutions that end-up outside the feasible region, thus, expectantly, driving the search

space of the optimization, back to viable solutions, i.e., with much smaller objective values (in the case of minimization) [332]. The first step is to convert the n_{eqc} equality constraints into two additional inequality constraints, $g_j(x) \leq 0$ and $-g_j(x) \leq 0$ [329], so, that $n'_{ineqc} = n_{ineqc} + 2n_{eqc}$, $j \in \left[1, n'_{ineqc}\right]$. With that, for P_{co} a penalty function can be defined, $pen(x) : \mathbb{R}^{n_{dims}} \to \mathbb{R}$, that satisfies the following:

$$pen(x) = \begin{cases} 0, & g(x) \leq 0 \\ ]0, +\infty[, & g(x) > 0 \end{cases} \tag{4.5}$$

Subsequently, the second step is to rewrite the now unconstrained problem, $P_{co}(c)$ as:

$$\begin{aligned} P_{co}(c): \quad & \underset{x}{\text{minimize}} && OF(x) + c\, pen(x) \\ & \text{s.t.} && x \in \mathbb{R}^{n_{dims}} \end{aligned} \tag{4.6}$$

where c is a scalar known as penalty parameter/weight, and a penalty factor should be assigned for every (jth) violated constraint, if any. For instance, considering a one-dimensional feasible region given by $x \geq lb \wedge x \leq ub$, which after transforming results in the constraints, $g_1(x) = -x + lb \leq 0$ and $g_2(x) = x - ub \leq 0$, Figure 4.4 showcases the effect of the penalty parameter choice in the penalty term (with quadratic growth), $c\, pen(x)$ for unfeasible solutions (outside the delimited region). Given a sufficiently large value of c, the minimum of 4.6 will be found in a region where $pen(x)$ is the smallest. In fact, as $c \to +\infty$, the solution of $P_{co}(c)$ will converge to a solution of the original constrained problem [333]. This result is given by the penalty convergence theorem, as derived from a series of penalty lemmas (4.3.1), but first we have to establish (as given below) some assumptions and associated notation.

Let c_k be a monotonic sequence of penalty parameters, where $0 < c_k < c_{k+1}$ with $\lim_{k \to +\infty} c_k = +\infty$; let $q(c, x) := OF(x) + c\, pen(x)$; let x^k be the exact solution of $P_{co}(c)$, i.e., $x^k = \arg\min_x q(c_k, x)$, and x^* be any optimal solution of the original problem, P_{co}; Then, the following lemmas can be extrapolated:

Lemma 4.3.1 (Penalty Lemmas)**.**

$$\begin{gathered} q\left(c_k, x^k\right) \leq q\left(c_{k+1}, x^{k+1}\right) \\ pen\left(x^k\right) \geq pen\left(x^{k+1}\right) \\ OF(x^k) \leq OF\left(x^{k+1}\right) \\ OF(x^k) \leq q\left(c_k, x^k\right) \leq OF(x^*) \end{gathered} \quad , \forall k \in \mathbb{N}$$

From the lemmas above, the Penalty Convergence Theorem emerges as:

Theorem 4.3.2. *Assume that $pen(x)$, $g(x)$ and $OF(x)$ are continuous functions and let $\{x_k\}_{k=1}^{+\infty}$ be a sequence of solutions of $P_{co}(c_k)$, with $\bar{x}$ as an arbitrary limit point of $\{x_k\}$ that solves P_{co}.*

Proof. Given that the limit point is defined as $\bar{x} = \lim_{k \to +\infty} x_k$, provided the continuity of

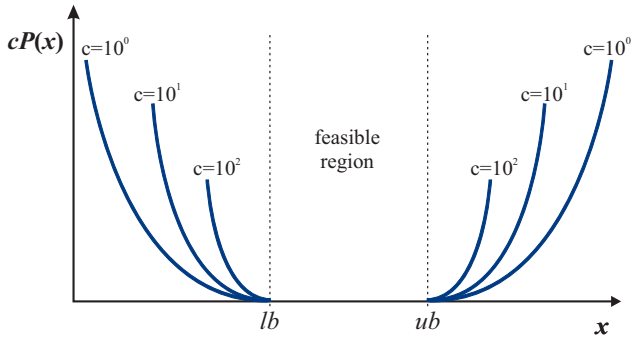


Figure 4.4: Quadratic penalty term value for several penalty parameters.

the involved functions, then $OF(\bar{x}) = \lim_{k \to +\infty} OF(x_k)$. From the penalty lemma, we know, that $q^* := \lim_{k \to +\infty} q(c_k, x^k) \leq OF(x^*)$

Thus, $q^* = \lim_{k \to +\infty} OF(x_k) + \lim_{k \to +\infty} c_k pen(x_k) \leq OF(x^*)\, q^* - OF(\bar{x}) = \lim_{k \to +\infty} c_k pen(x_k) \leq OF(x^*)$ ■

Since, both $q^* - OF(\bar{x})$ and $OF(x^*)$ are finite quantities, it means that $\lim_{k \to +\infty} c_k pen(x_k)$ is also finite. Moreover, given that $c \to +\infty$, it implies that $\lim_{k \to +\infty} pen(x_k) = 0$, which in turn means that $pen(\bar{x}) = 0 \Rightarrow g(\bar{x})$, which signifies the feasibility of the solution $\bar{x}$ in terms of the constrained problem. From Lemma 4.3.1, we know that $OF(x^k) \leq OF(x^*)$, thence, implying that $OF(\bar{x}) \leq OF(x^*)$, this means that $\bar{x}$ is not only feasible but also the optimal solution of P_{co}. That is to say, the global solution convergence of the penalty method is attested (Theorem 4.3.2).

The common class of penalty functions is given in 4.7. If $q_{exp} = 1$, $pen(x)$ is a linear penalty function, which may not be differentiable at points where $g_j(x) = 0$, while more frequently, a quadratic penalty function is used, $q_{exp} = 2$.

$$pen(x) = \frac{1}{2} \sum_{j}^{n'_{ineqc}} [\max\{0, g_j(x)^{q_{exp}}\}],\ q_{exp} \geq 1 \quad (4.7)$$

Alternatively, Barrier methods' working principle is to avoid the constraint violation in the first place, and so for the original constrained problem, P_{co}, a barrier function, $b(x) : \mathbb{R}^{n_{dims}} \to \mathbb{R}$, with $b(x) \geq 0$ for all x that satisfy $g_j(x) < 0$ and $b(x) \to \infty$ when $\lim_{\substack{x \to lb \\ x \to ub}} g(x) = 0$. That is to say, a very large cost is imposed to feasible points that lie close to the feasible region boundary. The logarithmic function is a common example of a barrier function that displays this behavior. So, similarly to equation 4.6, one can write:

$$P_{co}(c): \quad \underset{x}{\text{minimize}} \quad OF(x) + \frac{1}{c} b(x) \qquad \text{s.t.} \quad \begin{matrix} g(x) < 0 \\ x \in \mathbb{R}^{n_{dims}} \end{matrix} \tag{4.8}$$

As the now named, barrier parameter, c, increases, the barrier term dominates the new objective function, and so the resulting unconstrained problem becomes increasingly unfavorable to solve (ill-conditioned around the boundary region) [333], which may lead to a slow convergence, thus, rendering them a less popular choice versus the penalty-based method.

4.4 Multi-objective problems

So far, we have been dealing with single-objective optimization problems. However, many optimization problems are, in practice, multi-objective, as real-world applications often require balancing competing criteria — such as efficiency versus cost or speed versus accuracy, etc. — rather than optimizing a single criterion (quantitative metric). As such, we can begin by defining a multi-objective problem in a manner analogous to the notation used for the single-objective problem (Equation 4.4), but where at least two independent objective functions (mapping the same decision variables/potential solutions) are to be minimized (or maximized) simultaneously [334]:

$$P_{co}: \quad \underset{x}{\text{minimize}} \quad OF_k(x) = 0, \quad \forall\, k \in [2, n_{objecs}] \qquad \text{s.t.} \quad \begin{cases} f_i(x) = 0, & \forall\, i \in [1, n_{eqc}] \\ g_j(x) \leq 0, & \forall\, j \in [1, n_{ineqc}] \\ x \in \mathbb{R}^{n_{dims}} \end{cases} \tag{4.9}$$

As shown in equation 4.9, we can now consider up to $n_{objects}$ criteria for evaluating feasible solutions. The objective is to find the best trade-off among these often competing — and sometimes even contradictory — goals. The particular characteristics of this type of optimization problem allow for a sub-classification of algorithms to be established. While various branching possibilities exist, as seen in Figure 4.2, a fundamental distinction between exact (deterministic) algorithms, such as multi-objective linear programming, and stochastic (approximate) algorithms remains valid. In this work, however, given the global scope of the problem and its more intuitive approach, we focus on the approximate branch, specifically examining how metaheuristics can contribute to solving simultaneous optimization tasks.

But first, we need to grasp some underlying concepts. Aggregating goals — specifically, the weighted sum of multiple objective functions, $\min \sum_{k=1}^{n_{objects}} w_k OF_k(x)$ — effectively trans-

forms the problem back into a single-objective one. This approach simplifies the solving task, as it eliminates the need to adapt the optimization algorithm to handle multiple objectives. However, choosing appropriate weights factors, w_k is not trivial, especially when objectives are numerous (which imply a time-consuming refinement task), with very different scales, varying importance, and/or conflicting natures.

The weighted sum method assumes a linear trade-off between objectives, which tends to favor solutions along convex edges of the Pareto front (as will be explored below), while skipping intermediate concave regions. This assumption can lead to gaps, where certain solutions remain unreachable, regardless of the optimization algorithm's capabilities. Furthermore, this collapse into a single-objective problem produces only one solution, rather than the set of trade-offs provided by the Pareto front approach. As a result, it limits the understanding of correlations between objectives — specifically, how improving one affects (may deteriorate) others.

Other noteworthy (there are other instances) and fairly easy to implement alternative approaches to a truly multi-objective optimization implementation — i.e., non-elitist, non-Pareto-based algorithms — include the Vector Evaluated Genetic Algorithm (VEGA) — a key inspiration for several multi-objective metaheuristics, including the multi-objective guaranteed convergence PSO, as demonstrated in [335] — and the ε-Constraint method.

In general terms, the first approach divides the population of candidate solutions (positions) into sub-swarms, with each sub-swarm focusing on optimizing a single objective function (independent best local value/fitness value). Typically, these methods integrate a knowledge transfer strategy between sub-swarms, facilitating global exploration, often in a randomized manner. Some clear drawbacks, though, lie in its ability to find the proper compromise — especially when objectives are numerous or highly conflicting (similar to the limitations of the linear aggregating sum). In such cases, the decision space required to optimize one objective may be significantly different from that needed to minimize another. As a result, not only does this approach fail to maintain diversity, but it also often pushes solutions toward the middle of the objective space — a phenomenon known as the "middling problem" — instead of properly exploring what we will define as the Pareto front. By not preserving non-dominated solutions, it ultimately leads to sub-optimal convergence.

As for the ε-Constraint method, the multi-objective optimization problem is transformed into a series of single-objective problems. Specifically, one chosen objective function is optimized while treating the remaining objectives as bounded constraints, each with an allowable threshold denoted by ε_k. After solving for one objective, the constraints ε_k are systematically adjusted, and the next objective undergoes single-objective optimization. This iterative process continues for each objective until a set of equivalent to Pareto optimal is obtained, effectively approximating the Pareto-optimal front.

Alternatively, and to fully embrace the shift in the nature of the problem, it is necessary to move away from the traditional pairwise comparisons (using relational operators) employed in single-objective optimization. Instead, optimality is defined through Pareto

Dominance, where in lame terms, a solution is considered optimal if no other solution in the search space outperforms it across all objectives. This approach allows one to address all objectives simultaneously, without the limiting assumptions of earlier algorithms, thereby enabling the discovery of trade-off solutions. It introduces a more nuanced notion of optimality — one that does not imply complete superiority in every objective, but rather reflects a well-balanced compromise.
Accordingly, the next step is to introduce the formal definition of Pareto dominance - used to compare solutions in a multi-objective optimization problem - as defined below in 4.4.1.

Definition 4.4.1 (Pareto Dominance)**.** *A solution x_1 is said to dominate another solution x_2, denoted as $x_1 \succ x_2$, if:* $OF_k(x_1) \geq OF_k(x_2) \quad \forall k \in \{1, \ldots, n_{objects}\} \ \wedge \ \exists i \in \{1, \ldots, n_{objects}\} \mid OF_i(x_1) > OF_i(x_2)$.

In words, this means that a solution Pareto-dominates another if it is no worse in all objectives and strictly better in at least one.
Based on this, we define a feasible solution as Pareto-optimal if it is not dominated by any other solution in the problem's search space. In practical terms, this implies that a solution cannot be improved in any objective without degrading its performance in at least one other objective function, denoted as OF_k. Accordingly, the Pareto optimal set consists of these Pareto-optimal solutions. Their corresponding values in the objective space — also known as the co-domain of the objective functions — form a smooth surface (idealized boundary of the different trade-offs between objectives) that interpolates (or connects) the Pareto-optimal set. This surface is referred to as the Pareto-optimal Front or Non-Dominated Front, as illustrated in Figure 4.5. In practice, however, finding this front is very difficult, so we instead seek a global Pareto front that provides a good approximation.
These definitions — particularly definition 4.4.1 — are used to adapt optimization algorithms, such as the metaheuristics introduced in sub-section 4.2.2. In other words, candidate solutions, which are iteratively updated, are evaluated across all objectives, and the non-dominated ones are identified and stored in a repository (also known as an external population).
This is essential because, unlike in single-objective optimization, the "best" solution is no longer simply the one with the lowest objective value. Instead, there are multiple non-dominated solutions, none of which is intrinsically better than the others. Therefore, it becomes necessary not only to preserve these non-dominated solutions — acting as a form of memory to maintain elitism and assemble the Pareto Front for interpreting the final results — but also to use them to guide the search process.
When the metaheuristic requires a collective best to guide the search—as in swarm-based methods—a common strategy is to use a roulette wheel selection mechanism. This involves selecting a non-dominated solution from the repository in a purely probabilistic manner, often influenced by diversity or density measures, to serve as a leader for updating the positions of candidate solutions.
However, due to the evolving nature of these algorithms and the multiplicity of candidate solutions, the repository can quickly grow unbounded — containing a very large number

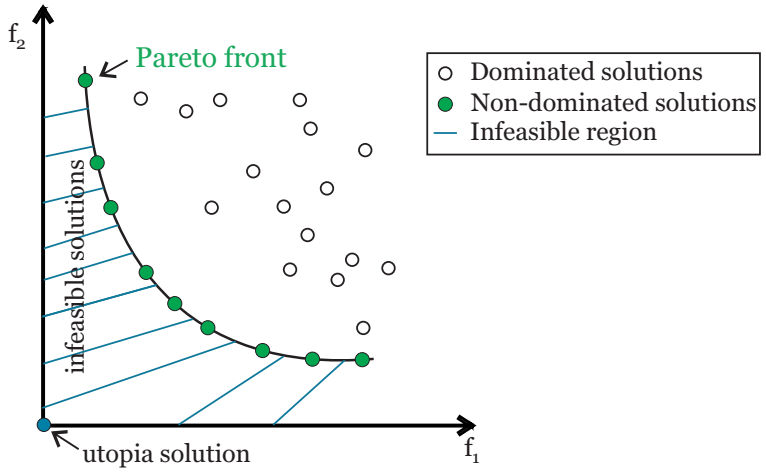


Figure 4.5: Pareto Front (green curve) in a two-objective space, Non-Dominated/Pareto-optimal solutions (green circles), Dominated solutions (white circles), and Infeasible region (delimited in blue).

of non-dominated solutions. This leads to redundancy and potentially unbearable computational costs. Therefore, diversity preservation becomes a key concept when bounding (i.e., limiting) the repository size. It ensures a well-distributed range of decision trade-offs (points) from which the final optimization decision can be made. Put simply, it helps maintain a smoother surface interpolation of the Pareto front without overwhelming it with excessive and redundant solutions.
Key breakthroughs in this field were the introduction of a Pareto archived evolution strategy as well as the non-dominated sorting genetic algorithm (popularized as NSGA-II), this last which introduced an elitist procedure where first a non-dominated sorting takes place, rejecting dominated solutions and then crowding distance sorting is used to address the issue of ”coverage” or in other words to explicitly preserve diversity in the Pareto non-dominated solutions.

4.5 From metaheuristics to surrogate-assisted evolutionary algorithms

Real-world engineering often involves complex optimization problems with a large number of variables — commonly referred to as expensive problems — which can significantly degrade the performance of metaheuristic algorithms during their iterative search for optimal solutions. Computational electromagnetics, mechanical design, finite element analysis, and fluid dynamics are just a few well-known examples of engineering domains that frequently involve nonlinear, non-explicit equations and high-dimensional optimization spaces, often requiring costly high-fidelity simulations [336]. These problem characteristics mean that conventional metaheuristics often require a substantial number of fitness evaluations (FEs) just to begin converging. In some cases, the computational burden becomes prohibitive — a phenomenon commonly referred to as the curse of dimensionality.

To lessen this time-consuming burden, an increasing amount of research works have been gradually shifting the attention towards the use of surrogate models, i.e., relatively computationally cheap approximated models (surrogates), that in turn replace the expensive FEs using the real model/function [337]. These models work by approximating the output from a set of input data based on the behavior of a relatively complex system, yet given the deviation between the actual model and the surrogate model output, is often necessary to run the original model in a few instances [338].

Accordingly, several surrogate-assisted evolutionary algorithms (SAEAs) have been developed [339, 340], and these typically employ polynomial regression models [341], Gaussian processes (GPs) also known as Kriging models [316, 342, 343], support vector machines (SVMs) and support vector regression (SVRs) [344, 345], radial basis functions (RBFs) neural networks (RBFNNs) [346, 347] and other type of neural networks (ANNs) [348, 349] to construct the surrogate model. The different traits of these approximation models is explored in a comprehensive manner. In [350], a SAEA is proposed where a more global model with faster convergence is achieved through a RBFNN, whereas a Kriging model is employed to obtain a more local model.

Other examples of SAEAs include a two-layer surrogate-assisted PSO, where by managing a database of historical (particle) positions, a global surrogate is employed in conjunction with several local surrogates to perform the model approximation, where a RBFNNs is used to build these surrogates [336]. A similar, two-level optimum search approach is followed in [351], where the water cycle algorithm is assisted by an hierarchical surrogate. While a a multi RBF-high fidelity surrogate model is developed in [352], as a suitable option for expensive problems. A decision space partition, where different surrogates are built for different clusters of positions, is suggested in [353], as an effective global search strategy, which is then smoothly complemented with an adaptive selection strategy for the local search. Fuzzy logic has also been applied to this field, and an hierarchical surrogate with a probabilistic PSO search based on global and local assist models is proposed in [354]. An engineering application of SAEA's is shown in [355], where the wireless sensor network coverage problem is solved by employing an add-point strategy based on the information from historical surrogate models (RBFs), coupled with a restart strategy. With this methodology, the authors are able to select better candidate (promising) multidimensional positions to evaluate through the real objective function. Another issue concerns the scarcity of training data used to build the surrogate models associated with hard engineering problems, as such authors in [356] present a global transfer optimization framework, where similar information is inherited.

Due to the fitting capabilities of radial basis functions and its fast training, RBFNNs are a common choice to build the dynamically updated surrogate models [357–360]. A two-phase optimization framework, that fuses the exploitation capabilities of surrogates with the exploration of metaheuristics is proposed in [357]. This constitutes another trend in this field, since the combined use of different metaheuristics or surrogate types is often a strength to achieve higher accuracies [336, 346, 361–363]. An uncertainty-based crite-

rion considering the distance and fitness value information simultaneously is proposed in [364], with two pre-screening criteria to balance exploration and exploitation, and the surrogate is assisted by PSO algorithm. A novel evolutionary strategy based on two co-evolutionary mechanisms is proposed to improve the performance of the Jaya algorithm by replicating an optimal directional guidance and historical learning [365], while in [340] the authors propose a separation of the swarms, for the different optimization stages during the search space, as a better alternative. Additional features like a pre-screening criterion, evolution control strategies and restart strategies are also a trend in the field as illustrated by authors in [366].

4.5.1 Surrogate Models

Function approximation is in general an important task in engineering problems. Following the Weierstrass approximation theorem, one can state that for a continuous function, $f(x)$, over a closed domain interval $[a, b]$, $f(x)$ can be approximated by a polynomial $\varphi(x)$ of a sufficiently large degree n, such that $|f(x) - \varphi_n| \leq \varepsilon$, where $\varepsilon > 0$ is an acceptable threshold. With this theoretical background, approximation or surrogate models $\hat{f}(x)$ are built upon previously evaluated/known training individuals, using the real (objective) function $f(x)$, and depending on the quality and representativeness of the training dataset, as well as the chosen method's ability to properly map the corresponding training space, different magnitudes of the surrogate error $\Delta_f = \left|\hat{f}(x) - f(x)\right|$ will emerge. As seen in the Introduction, these models have attracted researchers from a wide range of fields, due to their ability to substitute expensive simulations/complex (real) models [367].

In terms of the surrogate optimization, meaning the use of the approximated model to find good estimates for candidate solutions, there are two main approaches [368]. The first is the Offline (Direct) approach, where the surrogate $\hat{f}$ is initially constructed by using a sufficiently large set of well-distributed (expensive) individuals (quite a few pairs of candidate solution and its corresponding real fitness). While for the remainder of the optimization search, the model is kept unchanged (static), which means the optimization is made offline (without further information on the real function/simulation). In contrast, with an Online (Dynamic) approach, where a coarse surrogate model is built initially, i.e., with a relatively small amount of training data, and then as the iterations progress the model is augmented with additional (expensive) training samples. This online training procedure is generically illustrated in Figure 4.6, constituting a type of infill strategy, where the primary objective is to balance between the exploration of unknown search space regions and the exploitation of the present promising regions.

With both approaches to surrogate optimization, it is important to have pairs of training data that are well distributed across the entire search space in order to navigate between the different local optimums fairly rapidly, i.e., without the need for extensive exploration, and less prone to being trapped in one of those optimums [336]. This smoothness feature in the approximation curve, which "blinds" the search to the existence of several local

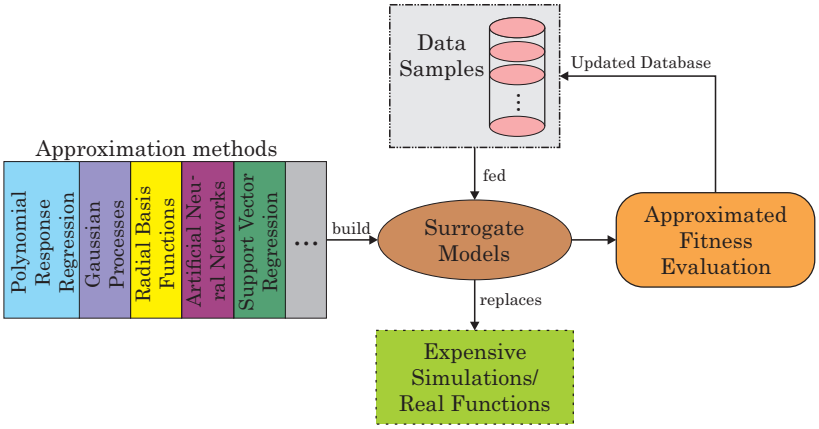


Figure 4.6: Online surrogate model working scheme (generic).

optima in comparison with the real objective function is shown in Figure 4.7, and it illustrates the "bless of uncertainty" principle. On the flip side, the "curse of uncertainty" translates the downsize due to a bad approximation and $\hat{f}$ leads to a wrong (nonexistent) minimum, as it can be seen by the cross-sign point highlighted in the same figure.

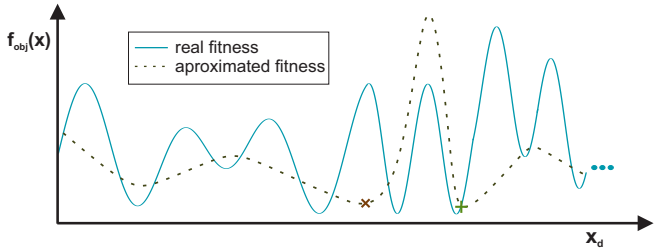


Figure 4.7: Surrogate Model: (rough) smoothing effect.

4.5.1.1 Radial Basis Function Neural Network

When dealing with higher dimensional approximation problems, a common trend in the SAEAs is to use RBFNNs to build the surrogate model, a type of machine learning architecture based on supervised learning, rather than using more conventional polynomial-based techniques. As the name suggests, these networks (which will be further contextualized and classified in a dedicated chapter and sub-section 5.3.3) take advantage of the well-documented ability of radial basis functions (RBFs), φ, to adequately interpolate a scattered set of points [369]. This type of real valued functions, $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$, set the activation threshold in the network neurons to one (maximum) when the input is null, and this is what effectively differentiates RBFNNs from the common feed forward type ANNs [336]. As such, the architecture of RBFNNs, depicted in Figure 4.8, consists of an input layer, a single hidden layer, where all the hidden neurons have a radial basis function, and a linear output layer.

With this regard, the Gaussian function is often the preferred choice to serve as the RBF, in the hidden (layer) neurons, and its processing expression, equation 4.10, is given below:

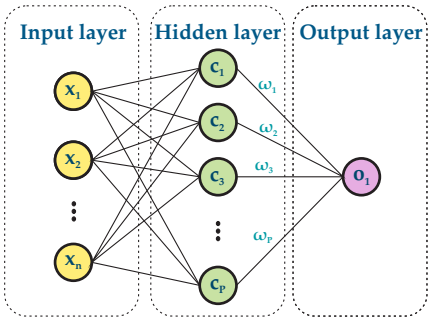


Figure 4.8: RBFNN: a default forward ML architecture.

$$\varphi_j(x_i) = \exp(-\frac{d_j(x_i)^2}{2\sigma_j^2}) \quad (4.10)$$

where $\varphi_j\,(x_i)$ is the jth hidden neuron response to the ith input x_i; $d_j(x_i) = \|x_i - u_j\|$ denotes the Euclidian distance between the ith input and the corresponding centroid of the Gaussian function u_j (Euclidian norm of the RBF), with $\varphi \in \mathbb{R} : 0 < \varphi \leq 1$; $\sigma_j > 0$ is the spread or width of the Gaussian. Besides, radbas is the common designation when referring to the transfer function in a radial basis neuron, and rather than just computing the activation cost solely based on the input x_i, it gauges the activation cost based on a distance $d_j(x_i)$. This feature marks a distinctive trait between regular hidden neurons and radbas neurons. Meaning that for each radbas neuron a connection weight ω_{ij}, representing the centroid of the RBF, is established with each of the P inputs, x_i, where $i \in \mathbb{Z}^+ : 1 \leq i \leq P$, in addition every radbas neuron has its bias term, b_j. Thus, the output response of the jth radbas neuron is computed as in equation 4.11, and the same will produce an output of 1 whenever the input x_i is equal to its centroid weight vector ω_{ij}.

$$a_j(x_i) = \exp\left(\sum_{i=1}^{P}\left(-\frac{\|x_i - \omega_{ij}\|^2}{2\sigma_j^2}b_j\right)\right) \quad (4.11)$$

Nevertheless, in most of the SAEAs including in this work, the considered radial base function is a cubic function, i.e., $\varphi_j\,(x_i) = d_j(x_i)^3$. Also, it is important to notice that when dealing with hard multidimensional problems, with a finite set of available positions (sparsity in the explored regions), we transform an interpolation problem into an approximation problem. Limiting the computation cost by defining a certain (approximation) error goal, ε_{RBF}, as the chosen stopping criteria (allowable tolerance in the approximation), which should be coupled with a tailored spread value to achieve a smoother function approximation.

4.5.2 Slime Mould Algorithm

The slime mould algorithm (SMA) is a fairly recent swarm intelligence metaheuristic proposed in [370]. It mimics the morphological and foraging behaviors of an acellular slime mould protist organism, Physarum polycephalum, that inhabits humid and cold places. Particularly, its food seeking mechanism where the cytoplasmic flux of slime mould surrounds and digests the food, by searching with a front end that resembles a fan-shaped shape, which is interconnected trough a venous network. Then, through a sequence of negative and positive feedbacks, depending on the odor, the slime mould finds its way to the target food by adjusting its propagation wave to alter the cytoplasmic flow in the veins, as is illustrated in Figure 4.9.

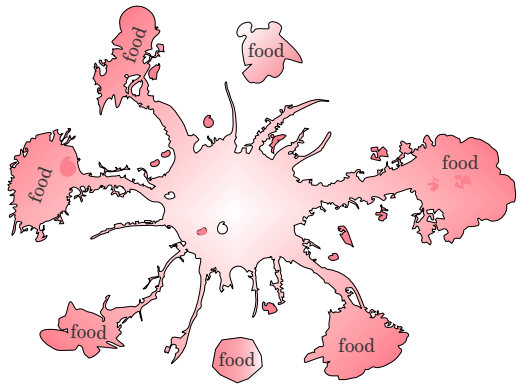


Figure 4.9: Slime mould foraging mechanism.

In terms of mathematical modeling, the SMA replicates the two main stages of this foraging behavior, namely the food approach and the food wrapping. In accordance, the algorithm begins $(iter = 0)$ by randomly initializing the population, $X\,(iter) = X(0) = lb + \mathrm{rand}\,(ub - lb) = X_{nPop \times nDims}$, in a search space defined by the respective problem' lower and upper bounds at each dimension, lb and ub, with a total of $nPop$ (different) search agents (population size) spread all across the problem' dimensions, $nDims$.

The slime mould position update is then modeled by the expressions in equation 4.12, where $X_{i,d}\,(iter + 1)$ indicates the updated position of the ith search agent (swarm individual) in a given dimension d; lb_d and ub_d represent the lower and upper bound of the search space in a given dimension $d \in [1, nDims]$; r_1, r_2 and rand denote uniformly distributed random values in the interval of $[0, 1]$; z is a sensitive small threshold parameter that ensures diversification in the population by assigning a newly generated random value to the aforementioned updated position; $X_{best,d}\,(iter)$ where the index $best$ stands for the best swarm individual (the slime mould position with the highest odor concentration) encountered so far, and the index d refers to a specific dimension of this individual; DF is its corresponding (swarm's best) fitness value; v_b is a randomly uniformly distributed vector that is generated for each ith search agent for all $nDims$dimensions (implying a different v_b when updating different slime mould positions), oscillating between $[-a, a]$, such as $v_b = -a + 2a \cdot \mathrm{rand}$, where $a = \tanh^{-1}\left(\frac{-iter}{iter_{\max}} + 1\right)$, which means

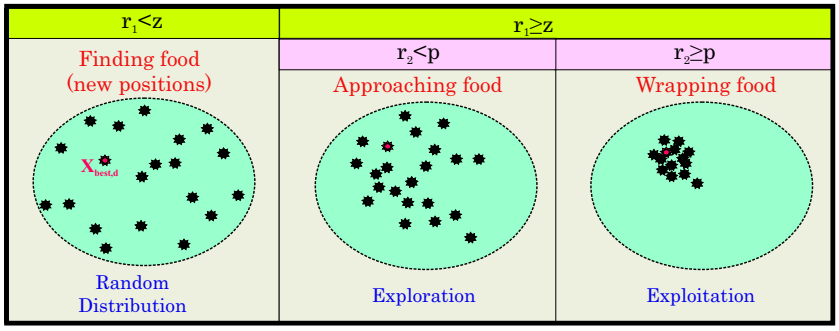


Figure 4.10: SMA: balance between different space search mechanisms.

it gradually decreases towards zero, $\lim_{iter \to iter_{\max}} a\left(iter\right) = 0$; v_c is an analogous vector to v_b since it also converges to zero and is generated for each swarm individual, but instead it oscillates between $[-1, 1]$, i.e., $v_c = -b + 2b \cdot \text{rand}$, where $b = \frac{iter_{\max} - iter}{iter_{\max}}$; W_i is a weight vector assigned to each slime mould position and is computed using equation 4.13 and together with v_b and v_c is one of the parameters that drives the swarm individual's search; $X_A\left(iter\right)$ and $X_B\left(iter\right)$ is a notation used to denote two randomly selected (current) swarm individuals, i.e., with indexes $A \wedge B \in [1, nPop]$, which are selected to update each current position when $r_2 < p_i$; lastly, p_i is an individual threshold value that determines which of the two main position update equations is selected to perform the position update, given by $p_i = \tanh\left(\left|S\left(i\right) - DF\right|\right)$, where $S\left(i\right)$ denotes the fitness value of the ith agent. An overview of the SMA space search mechanisms and the balance between exploration and exploitation is given in Figure 4.10.

$$X_{i,d}\left(iter+1\right) = \begin{cases} lb_d + \text{rand} \cdot \left(ub_d - lb_d\right), r_1 < z. \\ X_{best,d}\left(iter\right) + v_{b,d} \cdot \left(W_i \cdot X_{A,d}\left(iter\right) - \right. \\ \left. X_{B,d}\left(iter\right)\right), \ r_2 < p_i \wedge r_1 \geq z. \\ v_{c,d} \cdot X_{i,d}\left(iter\right), \ r_2 \geq p_i \wedge r_1 \geq z. \end{cases} \tag{4.12}$$

$$W_i = W\left(SmellIndex\left(i\right)\right) = \begin{cases} 1 + r_3 \log_{10}\left(\frac{bF - SmellIndex(i)}{bF - wF}\right), \\ \text{if } i \leq \lfloor pop_{size}/2 \rfloor. \\ 1 - r_3 \log_{10}\left(\frac{bF - SmellIndex(i)}{bF - wF}\right), \\ \text{otherwise.} \end{cases} \tag{4.13}$$

In turn, to calculate the weight vector for each search agent, the variable $SmellIndex\left(i\right)$ that is nothing more than the sorted fitness values of the entire population, $SmellIndex = \text{sort}\left(S\right)$, is introduced, and then the log of the quotient between its difference to the current worst fitness wF and the difference between the current best, bF, to the same wF, is obtained. The random value r_3 is responsible for the uncertainty in the venous contraction phase and it is multiplied by the log of the quotient. For the swarm individuals ranked in the best/higher half of the population, this term is added by 1 (for larger weight)

and subtracted from 1 for the lower half of the population.

4.5.2.1 Bound checking

An important detail concerns the bound checking after each position update (equation 4.12), given the possibility of agents traveling beyond the established (test problem) bounds (lb, ub), i.e., the defined search space. As such, throughout this thesis work, the commonly used deterministic back confinement [330], given by equation 4.14, is considered not only for the proposed SAEA, but also for the benchmark algorithms.

$$\begin{aligned} X_{i,d}\left(iter+1\right) &= \min\left(X_{i,d}\left(iter+1\right), ub_d\right), \quad i = [1, nPop] \wedge d = [1, nDims] \\ X_{i,d}\left(iter+1\right) &= \max\left(X_{i,d}\left(iter+1\right), lb_d\right), \quad i = [1, nPop] \wedge d = [1, nDims] \end{aligned} \tag{4.14}$$

4.5.3 Proposed Methodology

Considering all the contributions and development paths outlined in this chapter, this thesis proposes a novel surrogate-assisted algorithm, designated as "SASMA". An algorithm that leverages the balanced exploration and exploitation capabilities of SMA, applied for the first time within a surrogate-assisted evolutionary algorithm (SAEA) framework. Additionally, SASMA introduces an innovative approach to database management, based on a dual-criteria update strategy. This strategy governs both the storage (insertion and removal) of swarm positions in the global database and their subsequent selection for assembling the training dataset used to build the surrogate model at each iteration.
In terms of methodology, an overall view of the SASMA main mechanisms and the flow of information between them is provided in Figure 4.11. Starting with the familiar Latin hypercube sampling (LHS) initialization of the search agents' positions (shared with the benchmark algorithms), followed by the computation of its respective real objective value, which are then assigned to the global database, forming the initial set of training data for the first surrogate build. Thus, emphasizing that throughout the iterations, only positions deemed as "promising" are assigned with its real objective value, and from these only a handful will enter the DB, grounded on a dual-based criteria explained in further detail in sub-section 4.5.3.1, a key feature to ensure a more accurate function approximation.
The subsequent stage involves the chosen metaheuristic model, SMA, which is used in an adapted manner to find the optimal solution of each test problem, while the surrogate model is tasked with replacing the objective function, meaning that the swarm agents navigate through the search space based on an approximate model information, $\hat{f}\left(x\right)$, rather than the usual expensive real test problem. Sub-section 4.5.3.2 dives into the specifics of the surrogate building and how the training (positions) and target (real objective values) are selected. These stages will run in cyclic way until the chosen stop criterion is met, which, as common in the SAEAs literature, is a maximum number of functions evaluations $(maxFEs)$. Additionally, as we saw in sub-section 4.5.1, the added uncertainty of using an estimate of the real value can also work in our benefit, by "flattening" some of

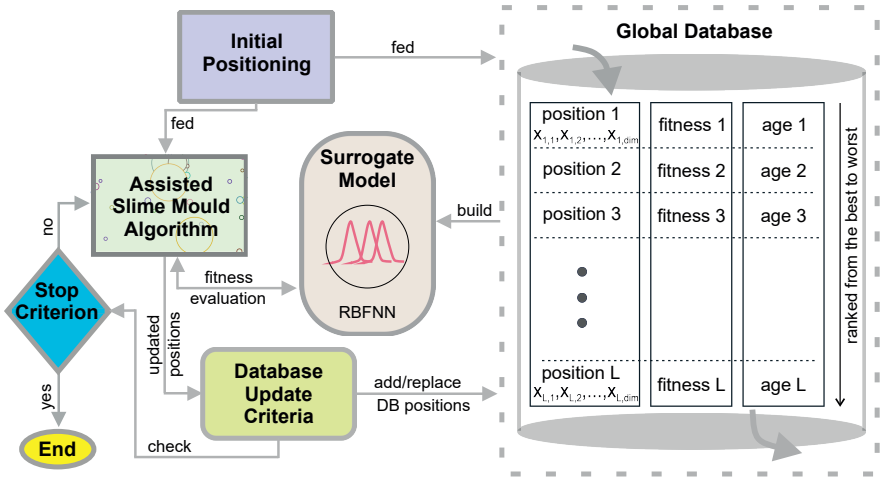


Figure 4.11: Proposed Surrogate Assisted SMA algorithm (SASMA): Overview of the flow between the main mechanisms.

the real model "valleys" during the search space process, i.e., minimizing the local minima phenomena.

4.5.3.1 Global database management strategy

To build the surrogate model that evaluates the different potential solutions (slime mould positions) it is necessary to have a dynamic database, working as a stack of (available) positions at each iteration. As illustrated in Figure 4.11, these stored (search agents) positions, and their respective fitness, constitute the input data and are the target data. In this work, a fixed length, $lengthDB$, is considered for the DB, independent of the test problem dimension (after performing several test runs). As we saw earlier, the database is initialized with the initial positions, its respective real fitness, as well as the stored position age, i.e., the iteration number in which this given position entered the database. This control variable purpose is used to evaluate the possible stagnation of the stored swarm agents and, thus, it can also work as a secondary update criterion, if needed.

Then, with the iterations, new positions are added/replaced on the DB, which means the accuracy of the surrogate model will improve over time, otherwise it would require a very large population size. Unlike more traditional surrogate-assisted approaches, the criteria used to add/replace positions in the DB is based on a best weighted metric [371], that considers two different factors rather than only one. The first criterion takes into consideration the surrogate value of the recently updated search agents (via SMA), $\hat{f}\left(X_i\left(iter+1\right)\right)$ or simply $\hat{f}\left(X_i\right)$, which objective is to minimize, and it scales this value based on the current best, $\hat{f}_{\min}=\min\left(\left\{\hat{f}\left(X_1\right),\hat{f}\left(X_2\right),...,\hat{f}\left(X_{nPop}\right)\right\}\right)$, and worst, $\hat{f}_{\max}=\max\left(\left\{\hat{f}\left(X_1\right),\hat{f}\left(X_2\right),...,\hat{f}\left(X_{nPop}\right)\right\}\right)$, surrogate values, i.e., the current positions that achieved the minimum and maximum fitness approximation values. So, for each swarm agent, i, its scaled surrogate value, $S\left(X_i\left(iter+1\right)\right)$, is given by equation 4.15.

$$S\left(X_i\right) = \frac{\hat{f}\left(X_i\right) - \hat{f}_{\min}}{\hat{f}_{\max} - \hat{f}_{\min}}, \quad S \in [0,1] \ \wedge \ i \in [1, nPop] \quad (4.15)$$

The second criterion, is centered around the distance to the existing DB positions, since when evaluating the potential of an updated search agent to enter the global DB, it is not only important to have a low (estimated) fitness value, but also to be farthest way possible of the previously evaluated positions, consequently promoting a more global search space, which in turn helps avoid local minima problems. Thence, we start off by obtaining the Euclidian distance of every position of the current swarm with respect to all DB stored positions, i.e., for each swarm position, $X_i\left(iter+1\right)$, compute its distance to every j position on the DB, X_j^{DB}, such as $d_{ij} = \left\| X_i\left(iter+1\right) - X_j^{\mathrm{DB}} \right\|$. Then, from this matrix of Euclidean distances, we not only find the overall minimum and maximum distances,
$d_{\min} = \min\left(\{d_{11}, d_{21}, ..., d_{nPop,1}, d_{1,2}, ..., d_{nPop,lengthDB}\}\right)$ and $d_{\max} = \max\left(\{d_{11}, d_{21}, ..., d_{nPop,1}, d_{1,2}, ..., d_{nPop,lengthDB}\}\right)$, respectively, but we also obtain a vector with the minimum distances of every ith position of the current swarm to any given stored position in the DB, i.e., $d_{single}\left(X_i\right) = \min\left(\{d_{i,1}, d_{i,2}, ..., d_{i,lengthDB}\}\right)$. With that, a scaled distance metric, $D\left(X_i\right)$, can be obtained using equation 4.16.

$$D\left(X_i\right) = \frac{d_{\max} - d_{single}\left(X_i\right)}{d_{\max} - d_{\min}}, \quad D \in [0,1] \quad (4.16)$$

With both metrics, the closer to zero indicates a better candidate to enter the DB, however they can be somewhat conflicting, i.e., a certain position can have a small (good) scaled surrogate value but at the same time have a large, scaled distance value (be very close to the existing DB positions), or vice-versa. For this reason, the merit function, f_{merit}, is a combination of the two metrics, equation 4.17, where through a weight, $0 < \phi < 1$, it is possible to modulate which metric should receive higher priority.

$$f_{\mathrm{merit}}\left(X_i\right) = \phi S\left(X_i\right) + \left(1-\phi\right) D\left(X_i\right) \quad (4.17)$$

A small value of ϕ denotes a greater importance to the more distant positions, which is key to the exploration stage, while a larger value of ϕ implies a prioritization of the smaller scaled surrogate values, which in turn favors the later exploitation phase. To achieve this outcome, a linear increasing function, $\phi = \phi_{\min} + \frac{\phi_{\max} - \phi_{\min}}{iter_{\max}} iter$, is considered in this work, rather than the often suggested four cycles of constant ϕ values.
Calculating the f_{merit} is just the first step. The entire DB management, as a key feature of the proposed SAEA, comprises a series of steps that must be done at each iteration, in order to initiate and dynamically update the DB, i.e., the variables X^{DB}, $fitness^{DB}$ and age^{DB}. These recurring steps are fully described in a flowchart form in Figure 4.12,

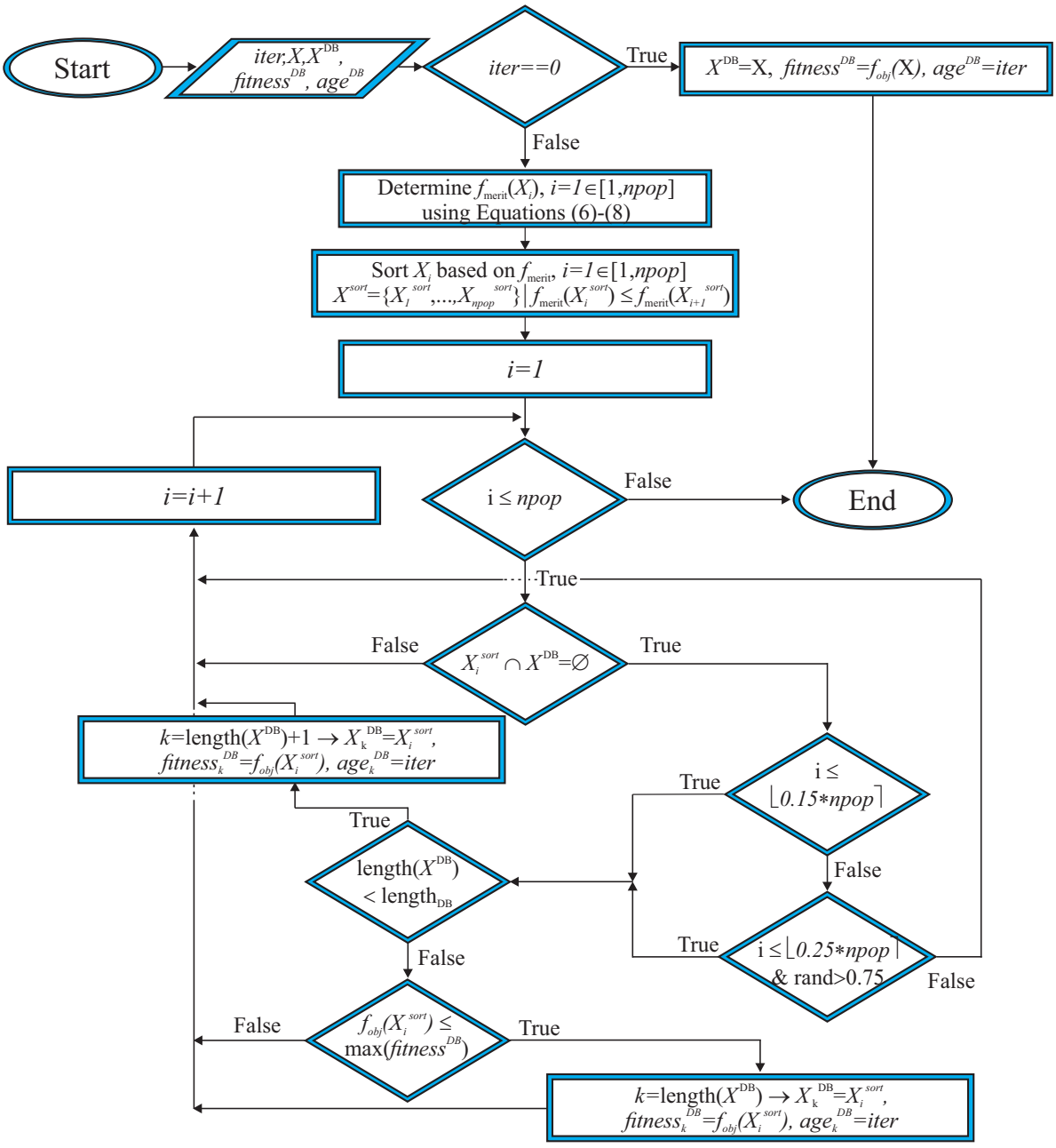


Figure 4.12: Detailed flowchart of the global database management: DB update and integrity check mechanisms.

and they involve: sorting the current SASMA swarm based on the respective $f_{\text{merit}}\,(X_i)$, which leads to a sorted set of positions, X^{sort}; a verification to see if the position is already stored in the DB; two "semi-elitist" conditions, that verify if the potential DB entries are "worthy". So, the updated search agents ranked in the first 15% in terms of f_{merit} or for 25% of the times, even the top quarter ranked positions, are deemed as valid thresholds, that strengthen the diversity in the DB management strategy.

Finally, even if it passes these conditions, the DB is checked to see if it is already full (reached its $lengthDB$limit), and, if so, we need to decide which position to remove, or conversely, maintain, in the DB. Accordingly, the considered criteria is to use the real

fitness value as an indicator of the position potential to the subsequent surrogate building stage, so the worst ranked stored position, $\max\left(fitness^{DB}\right)$, arbitrarily identified by an index k, X_k^{DB}, is compared with the real fitness value of the potential DB entry, $f\left(X_i^{sort}\right)$, and case if it is higher, the position is replaced, otherwise, X_i^{sort} is disregarded.

4.5.3.2 Surrogate building

A crucial factor when assembling the surrogate model is the chosen training data, meaning what positions are handpicked to fit the RBFNNs. In other words, as we saw previously, it is important to have a dispersed set of positions to accurately map the search space regions, particularly in the exploration phase, which means that the training data length also matters. But also, the "worthiness" of those positions, to properly narrow in on the optimum during the exploitation phase, i.e., in regions with bigger slopes in terms of fitness accuracy (smaller variations in the positioning leading to bigger fitness variations), it is more important to have more scattered positions around those, thus ensuring a better fitted surrogate model, which in turn can mitigate the "curse of uncertainty".
Understanding that the global DB tends (through the course of the iterations) to have a much wider distribution of positions, its "global nature", it is important to narrow down the search space defined to effectively fit the surrogate. This implies that after the first iteration, having concluded the first position update, probably only a portion of the global DB data is used to train the surrogate model, which also saves some computational time. This is an adequate approach to find the balance between the current swarm positioning and the overall upper and lower limits of the minimization problem, and therefore, allowing the surrogate model to shift its global nature to a more local nature in the closing stages of the search, i.e., with a concentration of swarm agents around the most promising subspace regions.
To accomplish this outcome, and inspired by the fitness approximation strategy proposed in [336], we start by finding the upper and lower limits of the current swarm agents, $X\left(iter+1\right)$, i.e., the subspace defined by the current maximum and minimum values of all the $nPop$ agents at each dimension d, denoted as $Xcurr_{ub}$ and $Xcurr_{lb}$, and computed as in equations 4.18 and 4.19.

$$\begin{aligned} Xcurr_{ub}\left(iter+1\right) = \{&\max\left(\{X_{1,d}\left(iter+1\right), X_{2,d}\left(iter+1\right), ..., \right. \\ &\left. X_{nPop,d}\left(iter+1\right)\}\right), \quad d=1,2,...,nDims\} \end{aligned} \tag{4.18}$$

$$\begin{aligned} Xcurr_{lb}\left(iter+1\right) = \{&\min\left(\{X_{1,d}\left(iter+1\right), X_{2,d}\left(iter+1\right), ..., \right. \\ &\left. X_{nPop,d}\left(iter+1\right)\}\right), \quad d=1,2,...,nDims\} \end{aligned} \tag{4.19}$$

Then, we can compute the subspace, $Xsspace$, used to select the training data from the global DB, i.e., the positions of the DB that are within a certain upper, $Xsspace_{ub}$, and lower bounds, $Xsspace_{lb}$, as computed in equations 4.20 and 4.21.

$$Xsspace_{ub}\left(iter+1\right) = \min\left(Xcurr_{ub}\left(iter+1\right) + \alpha\left(Xcurr_{ub}\left(iter+1\right) - Xcurr_{lb}\left(iter+1\right)\right), ub\right) \tag{4.20}$$

$$Xsspace_{lb}\left(iter+1\right) = \max\left(Xcurr_{lb}\left(iter+1\right) - \alpha\left(Xcurr_{ub}\left(iter+1\right) - Xcurr_{lb}\left(iter+1\right)\right), lb\right) \tag{4.21}$$

Where α is a spread coefficient, between 0 and 1, that determines how far beyond the current positioning the selected subspace goes. A value of 1 implies that the entire DB (all search space) is considered for building the surrogate, while a value of 0 implies that only the current swarm positioning determines the subspace bounds.

Modifying the original approach to this spread coefficient, authors have opted for an exponential decreasing coefficient, i.e., $\alpha = \alpha_{\max} e^{-\gamma(iter+1)}$, hence, fulfilling the objective of a wider subspace in the exploration phase versus a reduced promising region in the later exploitation stages. In addition, the minimum and maximum functions in equations 4.18 and 4.19., tend to usually return the first term (expression), constituting the bound values, and it explains why we said earlier that the surrogate model neglects the less "interesting" positions of the DB and is driven by the current swarm evolution.

An illustration of how equations 4.20 and 4.21, with their varying spread coefficient, determine the subspace used to build the surrogate, is provided in Figure 4.13. Importantly, this figure showcases the difference between the resultant subspace domain bounds, $Xsspace$, and the effective training domain, $Xtrain^{space}$, which is defined by the range between the upper and lower bounds per dimension of the n selected DB positions inside the computed subspace, i.e., $\left\{\max\left(\left\{X_{1,d}^{train}, X_{2,d}^{train}, ..., X_{n,d}^{train}\right\}\right) - \min\left(\left\{X_{1,d}^{train}, X_{2,d}^{train}, ..., X_{n,d}^{train}\right\}\right), \quad d = 1, 2, ..., nDims\right\}$. To ensure that the resultant training subspace lead us to a balanced training domain for the surrogate model, i.e., there are sufficient data points without a very expensive computation cost in the RBF model fitting, a verification of the X^{train} length, given by the index n, is made (as stated in Algorithm 1). A failure to comply with the established (balanced) length limits leads to either a truncation or an extension of the $Xtrain^{space}$.

When employing the RBFNN as the surrogate model, an important control parameter concerns its spread, σ_j, given in equation 4.11. A larger spread will require many neurons to find a rough fitness function approximation, while, for a smaller spread, the large number of neurons is the prerequisite to get a smooth function approximation, but it risks overfitting. To achieve this fine balance, an adaptative $spread$, equation 4.22, based on the current training set $\left(X^{train}\right)$ subspace (bounds), is used [336].

$$\sigma = \min\left(\left\{\max\left(\left\{X_{1,d}^{train}, X_{2,d}^{train}, ..., X_{n,d}^{train}\right\}\right) - \min\left(\left\{X_{1,d}^{train}, X_{2,d}^{train}, ..., X_{n,d}^{train}\right\}\right)\right\}\right), \quad d = 1, 2, ..., nDims \tag{4.22}$$

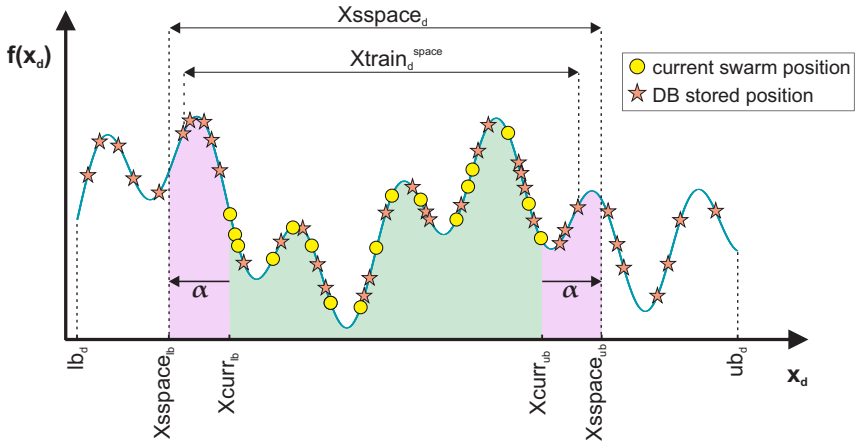


Figure 4.13: Selected subspace for surrogate modeling (spread coefficient effect).

Moreover, as hinted in section 4.5.1, an error goal is defined as a proxy of the approximation accuracy of the surrogate model. With this regard, a typical approach is to use two surrogate levels [336, 362], i.e., a more global surrogate model where a greater error is allowed, and a more refined model for each/or part of the swarm agents, often defined as local surrogate model, where accuracy is targeted, to find a nearby solution. Acknowledging the benefits of this approach, in this work we opted for a single global surrogate that transmutes to a more local type of surrogate model, by defining a linearly decreasing error goal, equation 4.23. Hence, constituting a slightly different approach that shares the same principle. In other words, targeting less accuracy when the DB is filled with a more disperse set of positions (exploration phase), while favoring more accuracy in the latter stages, when the DB is composed of neighboring positions given the convergence properties of the SMA.

$$\varepsilon_{RBF} = \varepsilon_{\max} - \frac{\varepsilon_{\max} - \varepsilon_{\min}}{iter_{\max}} iter \quad (4.23)$$

Overall, the careful fine-tune of the the control parameters, ϕ regarding the database management strategy, as well as the surrogate building related parameters, γ, α, σ and ε_{RBF}, which with exception of the first, all vary with the course of the iterations, we are able to gradually shift from a more global to a more localized surrogate model. A shift that is crucial to capture the error differences, with the increasingly smaller position updates during the exploitation phase.

4.5.4 Case Study and Benchmarking

To conduct a fair evaluation of the proposed SASMA, well-known SAEAs with publicly-available code, namely TLSAPSO [336], from which many of the SASMA features were inspired as mentioned in sub-section 4.5.3, SAEA-RFS [372], SHPSO [360], TL-SSLPSO [373], CALSAPSO [374] and GORS-SSLPSO [359] were selected and the their code was

Table 4.1: Parameter settings for the benchmark algorithms and the proposed SASMA algorithm.

Algorithm	Parameter settings
TLSAPSO	$\omega = 0.7, c_1 = c_2 = 2.05, v_{\min} = -100, v_{\max} = 100$ $fit_{treshold} = 10^{-4}, nb_{treshold} = 10^{-3}$ $globalsurr_{RBFgoal} = 10^{-1}, globalsurr_{RBFgoal} = 10^{-2}$ $RBF_{maxneurons} = 20, RBF_{addneurons} = 5$
SAE-ARFS	$F = 0.8$(scaling factor), $CR = 1$(crossover rate) Exponential Crossover, $\varphi_j\left(x_i\right) = d_j(x_i)^3$ $lenghtXtrain_{lb} = 200, num_{subproblems} = 5$ $\max dimsize_{subproblem} = 20, numiters_{subsubproblem} = 5$
SHPSO	$\omega = 0.7298, c_1 = c_2 = 2.05, v_{\min} = lb, v_{\max} = ub$ $lenghtXtrain_{lb} = 100,\ nDims < 100$ $lenghtXtrain_{lb} = 200,\ nDims \geq 100$ $M = 100, \beta = 10^{-2}$
TL-SSLPSO	$\omega = 1, c_1 = c_2 = \text{rand}, \omega = 1, \varphi_j\left(x_i\right) = d_j(x_i)^3$
CALSAPSO	$\theta = \frac{g}{100}, \omega = 0.9 - \frac{\theta}{2}$ $c_1 = c_2 = 1.49445,\ lenghtXtrain_{lb} = 100$
GORS-SSLPSO	$\omega = 1, c_1 = c_2 = \text{rand}, \omega = 1, \varphi_j\left(x_i\right) = d_j(x_i)^3$
SASMA (SMA)	$z = 0.03$

modified accordingly. For verification and fair replication purposes, the same initial positioning is seen by each one for every test run, and the main control parameters (for the benchmark algorithms) are shown in Table 4.1 and the full test code can be accessed in [375].

With regards to the optimization, a total of 35 runs are performed, which means that the same 35 different initial positions are used to benchmark all the algorithms, thus dismissing the initial positioning bias. Moreover, it allows us to assess not only the methodology' accurateness but also its precision (through the standard deviation), i.e., judge the ability of each method to consistently find the best solution. The swarm size, $npop$, is set to 30, and these individual agents will survey the problem's search space up until a total of 330 or 1000 (real) FEs, for 30D and 100D, respectively, is reached (following literature standard). Unlike in traditional metaheuristic algorithms, where there is an equivalence between the number iterations and the FEs as a stop criterion, this is no longer the case when using surrogate assisted algorithms, since most of the times, on a given iteration, we end up evaluating only a limited amount of the updated positions, and not the entire population.

The difference between the best (real) fitness value achieved in each run by each algorithm, and the problem/test function global minimum, represented as $f_{\min}$ in Table 4.2, is the chosen error metric. To analyze the error performance of the proposed SASMA versus the chosen benchmark algorithms, beyond the common descriptive statistics, i.e., the mean, standard deviation, and the minimum and maximum error values, we followed the common approach of using the non-parametric Wilcoxon Signed-rank and the Friedman statistical tests with 5% significance [376].

The first test gauges if there are substantial differences in the central tendency of two competing data series (SASMA vs all the benchmark algorithms). A one-on-one compar-

ison of the signed errors in the 35 runs for all the problems/test functions allow us to say if the algorithm's error accuracy differs substantially (null hypothesis) and so, attributing the symbol "=", or on the contrary, if there is a definite difference in terms of central tendency (alternative hypothesis), a smaller mean error value for the SASMA versus the benchmarked algorithm, means a better performance of the proposes SASMA, and thus, the symbol "+", whereas the opposite, lead us to the symbol "-" in the results analysis (tables). While the second test allows us to perform a multiple algorithm comparison for the non-normally distributed algorithm's signed error for each of the 35 runs. Revealing if there are any significant differences in the obtained problem/test function errors (alternative hypothesis). The same is done by examining the (Friedman) mean rank, rather than the true value. It is expected that for similar distributions, the mean ranks are approximately identical (null hypothesis).

4.5.4.1 Mathematical Test Functions and Optimization Results

Seventeen classical benchmark test functions, shown in Table 4.2 alongside its global minimum $(f_{\min})$ and properties, are used to evaluate the proposed SASMA, including its scalability, by considering two different number of high dimensions, $nDims = [30, 100]$. Among these test functions, F1–F4 and F14 are continuous unimodal functions; F5 is the Rosenbrock function which can be considered multimodal when the problem dimension is greater than 3 [377]; F6 is a discontinuous step unimodal function; while F7 is a noisy unimodal function; F8-F13 and F15-F17 are multimodal functions, with many local minima, thus, rendering them more difficult to optimize [378]. A detailed account of the optimization bounds and function mathematical expressions is given in Table A.1.

Table 4.2: Case Study: List of unimodal and multimodal test functions.

Function no	Function name	fmin	Properties
F1	Sphere	0	Unimodal
F2	Schwefel 2.22	0	Unimodal
F3	Schwefel 1.2	0	Unimodal
F4	Schwefel 2.21	0	Unimodal
F5	Rosenbrock	0	Multimodal with narrow valley
F6	Step	0	Unimodal
F7	Quartic	0	Unimodal
F8	Schwefel 2.26	$-418.9829 \times nDims$	Multimodal
F9	Rastrigin	0	Very complicated Multimodal
F10	Ackley	0	Multimodal
F11	Griewank	0	Multimodal
F12	Penalized (1)	0	Multimodal
F13	Penalized (2)	0	Multimodal
F14	Ellipsoid	0	Unimodal
F15	Shifted Rotated Rastrigin (F10 in [379])	-330	Very complicated Multimodal
F16	Rotated Hybrid Composition (F16 in [379])	120	Very complicated Multimodal
F17	Rotated Hybrid Composition (F19 in [379])	10	Very complicated Multimodal

As mentioned before, key descriptive statistics and two non-parametric tests are used to assess the algorithm's performance. As such, for all the testing functions, F1-F17, the mean, standard deviation (STD), and minimum error values of all the benchmark SAEAs and the proposed SASMA are presented in Table 4.3 and Table A.3, respectively, for 30D and 100D. CALSAPSO very slow convergence for 100D made the simulation cost very heavy, which why it was not considered in the test case. The best results per metric are shaded in grey, clearly indicating the superior performance of SASMA, which has only failed to achieve the best mean error value (lowest value) on four functions for 30D, and three occasions for the 100D, a scenario that is pretty similar in terms of the minimum error value, where we assess the ability of SASMA to be closest to the global function minimum at least in one of the runs. Despite being an harder problem, SASMA performance for 100D is even better than for 30D, which is explained by the larger number of function evaluations, and that is why for nine out of the seventeen functions we see that SASMA was able to consistently find the global minimum (an STD of zero), while also being very close to this target value for F12.

The same can also be confirmed by the results of the Wilcoxon signed ranked test, at a 5% significance level, where SASMA is compared against every other algorithm. For both 30D and 100D, SASMA records significantly better error accuracies in most of the seventeen test functions (overwhelmingly scoring wins), against the benchmarked algorithms. The few exceptions have mostly occurred in the last three test functions.

For 30D, GORS-SSLPSO was the only benchmark algorithm to show a statistically better error performance in four functions, particularly in F1 and F6 with values in the order of 10^{-5}, while the remainder were only able to surpass or tie with SASMA in F15-F17, which meant that at least for fourteen out of the seventeen functions, SASMA was better. SHPSO and TL-SSLPSO showed consistent standard deviations, and this last SAEA achieved the best performance in the Shifted Rotated Rastrigin function. While for 100D, the advantage of SASMA is even more clear, overall with more wins, one less loss and with only one tie recorded. SHPSO is now the closest competitor, showing an outstanding performance for F6, F15 and F16, while for the remainder, it is clearly outmatched by SASMA, partially confirming the trend already seen for 30D.

The improved (lower) error accuracy is also attested by the Friedman test, which was performed for each function and taking into consideration each test run error, by analyzing the computed mean rank, we then can order individually each algorithm, where the lowest mean rank score indicates a better performance. The results for 30D are shown in Table 4.4, where we can see that SASMA is consistently the best ranked algorithm, with the exception of four functions (F1, F6, F15 and F16). Nevertheless, for the first of the two, it closely follows GORS-SSLPSO as the second best SAEA in terms of mean rank. For most of the test functions, the top 3 ranked algorithms are SASMA, GORS-SSLPSO and TL-SSLPSO. While on the opposite end, TLSAPSO and CALSAPSO were the worst ranked SAEAs.

As for the 100D results, shown in Table A.2, the accuracy of SASMA is even better, achiev-

Table 4.3: Statistical results of the proposed SASMA and the benchmark SAEAs on 30D mathematical test functions.

Test Function/Metric		CALSAPSO		GORS-SSLPSO		SAE-ARFS		SHPSO		TL-SSLPSO		TLSAPSO		SASMA
F1	Mean	2.212E+02		3.180E-05		2.288E+04		1.492E+01		3.880E+02		2.272E+04		1.226E-02
	St. Dev.	4.883E+02	+	2.293E-05	-	5.438E+03	+	1.015E+01	+	9.043E+02	+	4.500E+03	+	3.917E-02
	Min	1.940E-01		4.090E-06		1.362E+04		3.171E+00		1.678E+00		1.282E+04		5.095E-10
F2	Mean	4.826E+12		6.090E+04		5.821E+05		1.142E+04		5.332E+01		1.610E+09		3.258E-04
	St. Dev.	9.374E+12	+	1.781E+05	+	2.815E+06	+	4.704E+04	+	1.483E+01	+	6.128E+09	+	4.560E-04
	Min	1.762E+09		4.288E+01		7.892E+01		3.304E+01		3.866E+01		7.760E+01		2.777E-06
F3	Mean	1.417E+05		1.686E+04		7.193E+04		4.247E+04		2.689E+04		5.702E+04		2.470E+01
	St. Dev.	4.632E+04	+	1.230E+04	+	1.124E+04	+	1.190E+04	+	7.087E+03	+	1.303E+04	+	7.522E+01
	Min	7.582E+04		5.484E+03		4.358E+04		1.946E+04		1.508E+04		3.447E+04		5.311E-09
F4	Mean	7.628E+01		4.721E+01		7.398E+01		3.264E+01		4.088E+01		6.386E+01		5.160E-02
	St. Dev.	5.811E+00	+	8.021E+00	+	6.278E+00	+	6.421E+00	+	7.809E+00	+	6.450E+00	+	1.317E-01
	Min	6.579E+01		2.850E+01		5.888E+01		2.159E+01		2.957E+01		5.264E+01		1.892E-04
F5	Mean	6.988E+05	+	1.346E+05		4.185E+07		7.158E+04		1.015E+05		1.182E+08		2.921E+01
	St. Dev.	8.578E+05		2.920E+04	+	1.559E+07	+	7.039E+04	+	2.317E+05	+	4.750E+07	+	8.241E-01
	Min	9.871E+04		4.274E+04		1.475E+07		8.585E+03		2.488E+03		3.704E+07		2.895E+01
F6	Mean	1.646E+02		3.536E-05		2.235E+04		1.842E+01		2.271E+02		2.629E+04		6.372E+00
	St. Dev.	2.796E+02	+	2.435E-05	-	4.153E+03	+	1.129E+01	+	4.275E+02	+	5.763E+03	+	1.214E+00
	Min	4.659E+00		5.083E-06		1.431E+04		5.132E+00		8.742E-02		1.584E+04		5.842E-04
F7	Mean	8.883E-01		5.761E-01		2.025E+01		5.772E-01		5.688E-01		1.266E+02		9.476E-03
	St. Dev.	4.258E-01	+	2.317E-01	+	8.888E+00	+	3.047E-01	+	3.945E-01	+	2.628E+01	+	7.626E-03
	Min	3.190E-01		2.548E-01		9.262E+00		1.161E-01		1.847E-01		7.957E+01		3.169E-04
F8	Mean	8.448E+03		5.380E+03		6.428E+03		9.051E+03		5.335E+03		7.651E+03		3.094E+03
	St. Dev.	1.371E+03	+	7.296E+02	+	6.831E+02	+	5.593E+02	+	7.648E+02	+	1.056E+03	+	2.298E+03
	Min	5.041E+03		4.288E+03		5.011E+03		7.926E+03		4.029E+03		5.732E+03		1.372E+01
F9	Mean	1.282E+02		1.332E+02		3.115E+02		2.749E+02		1.096E+02	+	3.729E+02		1.432E+01
	St. Dev.	4.302E+01	+	3.656E+01	+	3.806E+01	+	3.760E+01	+	2.674E+01		2.893E+01	+	2.745E+01
	Min	4.873E+01		8.159E+01		2.380E+02		1.757E+02		5.623E+01		3.070E+02		1.930E-07
F10	Mean	2.032E+01		1.434E+01		1.852E+01		8.688E+00		7.867E+00		2.009E+01		3.363E-03
	St. Dev.	3.964E-01	+	6.077E+00	+	1.265E+00	+	6.990E-01	+	4.311E+00	+	2.177E-01	+	3.216E-03
	Min	1.913E+01		3.299E+00		1.468E+01		7.183E+00		2.759E+00		1.912E+01		1.120E-04
F11	Mean	3.796E+00		2.660E+00		1.992E+02		1.116E+00		4.080E+00		2.736E+01		6.739E-04
	St. Dev.	2.915E+00	+	1.529E+01	+	3.328E+01	+	7.297E-02	+	4.817E+00	+	7.017E+00	+	1.253E-03
	Min	1.610E+00		1.113E-02		1.197E+02		1.039E+00		9.886E-01		1.636E+01		4.747E-07
F12	Mean	2.482E+05		1.661E+03		6.228E+07		8.456E+03		1.248E+04		1.921E+08		7.546E-01
	St. Dev.	7.293E+05	+	2.571E+03	+	4.416E+07	+	2.719E+04	+	4.020E+04	+	1.090E+08	+	3.324E-01
	Min	2.943E+01		1.662E+01		9.392E+06		5.588E+00		2.555E+01		4.059E+07		2.928E-06
F13	Mean	8.496E+05		1.283E+05		1.335E+08		4.071E+04		4.430E+04		4.071E+08		2.578E+00
	St. Dev.	1.239E+06	+	6.569E+04	+	6.399E+07	+	1.580E+05	+	6.827E+04	+	1.703E+08	+	9.339E-01
	Min	1.676E+03		1.234E+04		3.547E+07		1.971E+01		6.551E+02		1.005E+08		1.252E-01
F14	Mean	9.632E+03		4.486E+02		2.740E+05		5.951E+03		8.209E+03		3.211E+05		8.987E-02
	St. Dev.	1.118E+04	+	3.414E+02	+	7.496E+04	+	2.136E+03	+	9.588E+03	+	6.634E+04	+	2.538E-01
	Min	2.079E+01		3.969E+01		1.296E+05		2.669E+03		4.844E+02		1.932E+05		1.479E-08
F15	Mean	3.994E+02		2.115E+02		4.647E+02		2.837E+02		1.505E+02		5.226E+02		4.893E+02
	St. Dev.	5.489E+01	-	5.557E+01	-	5.948E+01	=	2.471E+01	-	3.792E+01	-	5.321E+01	=	8.253E+01
	Min	3.108E+02		1.294E+02		3.792E+02		2.236E+02		7.785E+01		4.219E+02		3.236E+02
F16	Mean	6.799E+02		4.803E+02		5.799E+02		4.114E+02		3.548E+02		6.584E+02		5.975E+02
	St. Dev.	1.177E+02	+	1.782E+02	-	8.952E+01	=	8.985E+01	-	1.428E+02	-	1.317E+02	+	1.315E+02
	Min	4.811E+02		2.330E+02		4.489E+02		3.057E+02		1.360E+02		4.424E+02		3.607E+02
F17	Mean	1.190E+03		1.004E+03		1.115E+03		9.806E+02		9.673E+02		1.110E+03		9.620E+02
	St. Dev.	1.574E+02	+	6.867E+01	=	5.178E+01	+	2.229E+01	=	3.050E+01	=	5.151E+01	+	8.828E+01
	Min	9.785E+02		9.177E+02		1.022E+03		9.464E+02		9.257E+02		1.017E+03		9.000E+02
		Win	16		12		15		14		14		16	
		Tie	0		1		2		1		1		1	
		Lose	1		4		0		2		2		0	

[1] According to the Wilcoxon signed rank test at 5% significance level, the symbols "+", "=" and "−" symbolize that the performance of SASMA is better, similar, or worse than that of other SAEAs, respectively. In addition, the best results for all metrics (from the 35 runs) for each test function are highlighted in grey.

Table 4.4: Mean rank scores (based on the Friedman Test) of the SASMA and the benchmark SAEAs on 30D mathematical test functions.

Friedman Test		CALSAPSO	GORS-SSLPSO	SAE-ARFS	SHPSO	TL-SSLPSO	TLSAPSO	SASMA
Mean Rank	F1	4.23	1.46	6.49	3.43	4.34	6.51	1.54
	F2	7.00	3.94	4.54	3.46	2.23	5.83	1.00
	F3	6.97	2.31	5.74	4.06	2.83	5.09	1.00
	F4	6.57	3.71	6.23	2.23	3.11	5.14	1.00
	F5	4.83	3.86	6.00	2.83	2.49	7.00	1.00
	F6	4.29	1.00	6.29	3.46	3.97	6.71	2.29
	F7	4.17	3.40	6.00	3.20	3.23	7.00	1.00
	F8	5.80	2.46	3.91	6.57	2.46	5.20	1.60
	F9	3.20	3.06	5.89	5.29	2.69	6.83	1.06
	F10	6.71	4.06	4.63	2.94	2.46	6.20	1.00
	F11	4.66	2.11	7.00	3.17	4.11	5.94	1.00
	F12	4.20	3.71	6.11	2.60	3.49	6.89	1.00
	F13	4.54	4.11	6.06	2.40	2.94	6.94	1.00
	F14	3.77	2.20	6.26	4.03	4.00	6.74	1.00
	F15	4.49	1.97	5.34	2.83	1.20	6.40	5.77
	F16	5.77	3.26	4.54	2.31	1.86	5.46	4.80
	F17	5.94	3.11	5.63	2.97	2.34	5.66	2.34

[1] According to the Friedman test the best mean rank (of the 35 runs) for each test function is highlighted in grey.

ing the best mean rank in fourteen of the test functions, and as was the case for 30D, it was GORS-SSLPSO with the best mean rank for F6, while for F15 and F16, it is now SHPSO that is ranked the best in detriment of TL-SSLPSO. Completing the top 3 in this test metric, is again SASMA and GORS-SSLPSO, which are now joined by SHPSO. Contrarywise, TLSAPSO and SAE-ARFS occupy the lower (ranked) end of the benchmark algorithms. Notwithstanding all the comparisons, the comprehensive error analysis not only confirms SASMA's ability to mitigate the initial positioning bias but also to substantially outperform TLSAPSO. A very significant result given that many of SASMA features were inspired by it.

To visually illustrate the aforementioned superiority of SASMA in comparison with the benchmarked SAEAs, the convergence curve of the best SASMA run, i.e., the run where the proposed methodology achieved the minimum error value, is shown together with the convergence curves for all the selected algorithms, thus, ensuring that all start from the same initial positioning. Figure 4.14 depicts the results for 30D and the unimodal test functions, and it is no surprise to see that SASMA apart from F6 where GORS-SSLPSO excelled, the minimum fitness value (log) was achieved by SASMA and the most of the times in far fewer FEs.

For the multimodal test functions, Figure 4.15, still considering 30D, the picture is very similar, with SASMA consistently reaching the lowest fitness value with a faster convergence (sharp drop) in test functions F8 and F12 and F13. And is again followed by GORS-SSLPSO and SAHO. The very complicated multimodal test functions, Figure 4.16, reveals a more challenging scenario where SASMA greatly outperforms all the SAEAs in F9 and despite the slower convergence, the same in F17. While for F15 and F16, it is the TL-SSLPSO that shows a greater performance, confirming the Wilcoxon and Friedman test results.

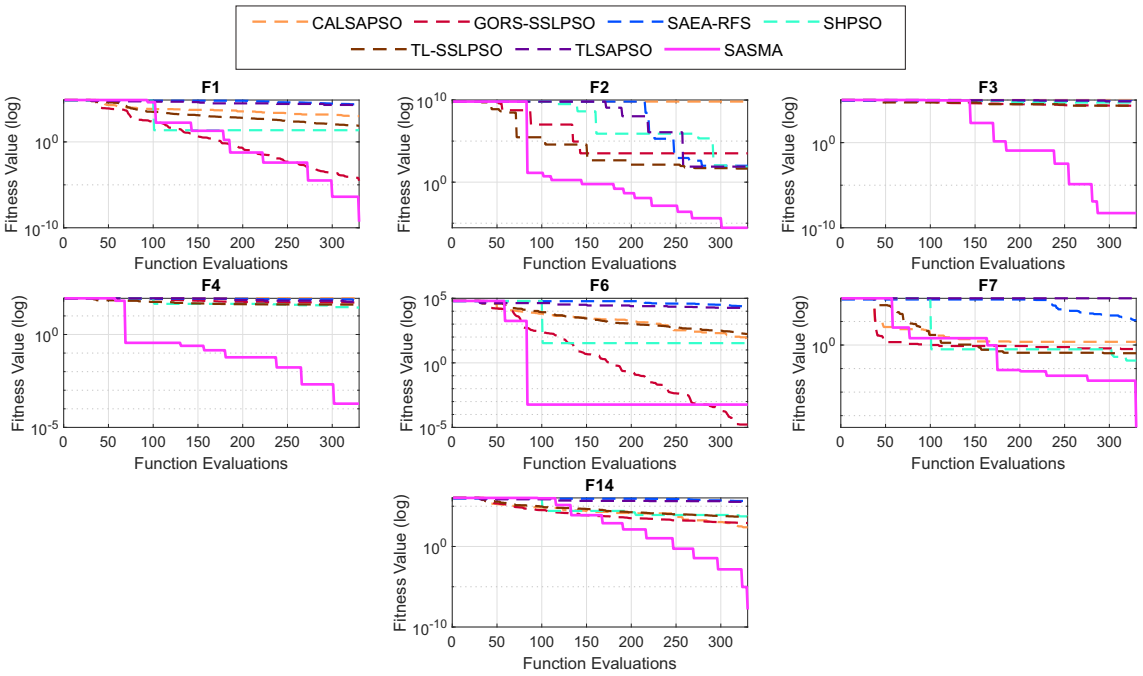


Figure 4.14: Convergence curves (fitness value): CALSAPSO (dashed orange), GORS-SSLPSO (dashed red), SAE-ARFS (dashed blue), SHPSO (dashed cyan), TL-SSLPSO (dashed brown), TLSAPSO (dashed purple) and SASMA (solid pink) on 30D unimodal test functions (F1-F4,F6,F7 and F14).

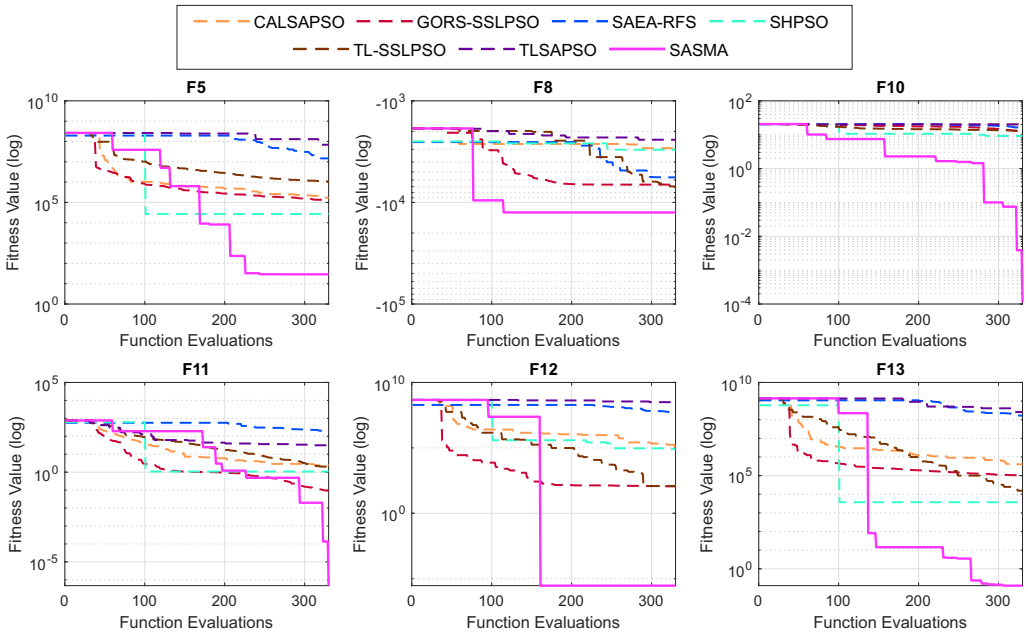


Figure 4.15: Convergence curves (fitness value): CALSAPSO (dashed orange), GORS-SSLPSO (dashed red), SAE-ARFS (dashed blue), SHPSO (dashed cyan), TL-SSLPSO (dashed brown), TLSAPSO (dashed purple) and SASMA (solid pink) on 30D multimodal test functions (F5,F8 and F10-F13).

For 100D, the convergence curves for the unimodal test functions are shown in Figure A.1, the multimodal test functions in Figure A.2 and the very complicated multimodal test

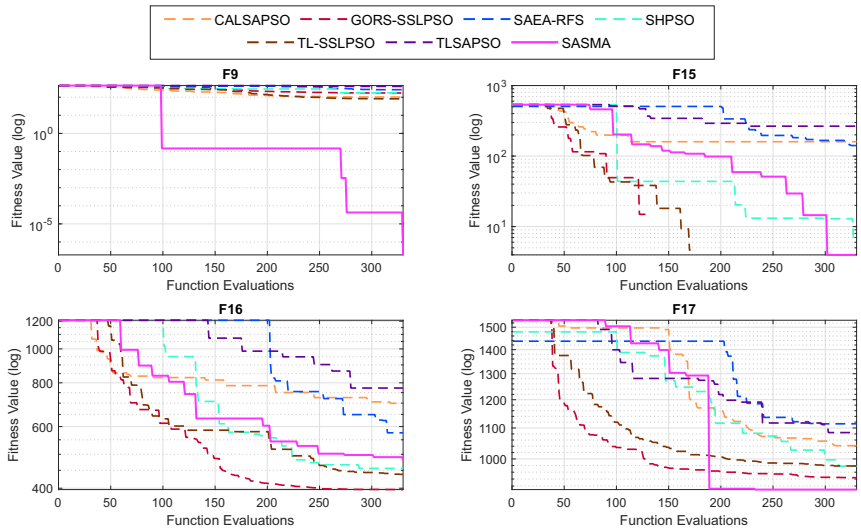


Figure 4.16: Convergence curves (fitness value): CALSAPSO (dashed orange), GORS-SSLPSO (dashed red), SAE-ARFS (dashed blue), SHPSO (dashed cyan), TL-SSLPSO (dashed brown), TLSAPSO (dashed purple) and SASMA (solid pink) on 30D very complicated multimodal test functions (F9 and F15-F17).

functions in Figure A.3. Likewise, we observe that the superiority of SASMA is even more evident, as the convergence curves for F3, F4, F10, F11 and F9 reveal. The exception occurs in F6 and F16, where GORS-SSLPSO and SHPSO, respectively, were able to outperform it. Noticeably, SASMA maintains the fast convergence characteristics and it even reached the global minimum in F11, and so, the log curve vanishes from this point onwards. In all the convergence curve figures, we can also confirm the trend revealed by the several error metrics and non-parametric tests, where GORS-SSLPSO and SHPSO are the closest competitors of SASMA, while the other SAEAs tend to display premature convergence and stagnation features.

In terms of computational cost, a relevant feature in large scale optimization problems, the relative time taken by the benchmark SAEAs in comparison with SASMA is shown in Figure 4.17, and is measured as the ratio between the time taken by each of the SAEAs and the time taken by SASMA. As we can see the x-scale is given in a logarithmic scale given that most of the algorithms take several orders of magnitude more than SASMA, with the exception of TL-SSLPSO and SAE-ARFS, for 30D and 100D, respectively, which are slower but roughly in the same order of magnitude.

To complement the results analysis, SASMA DB properties after each individual run for each test function are analyzed via a bi-variate histogram in Figure 4.18. The bar plots on the left side highlight the relative frequency of the final DB length and average age of the stored positions, whereas the bar plots on the right, highlight the distribution of the rules activated in order to enter the DB for all the performed runs. The distribution for 30D unveils that the most frequent class for the DB size is between 133 and 144 stored positions, while the average age of the positions is around 7 to 8 iterations, meaning that they are fairly new, and that with the considered small number of FEs the DB is being filled

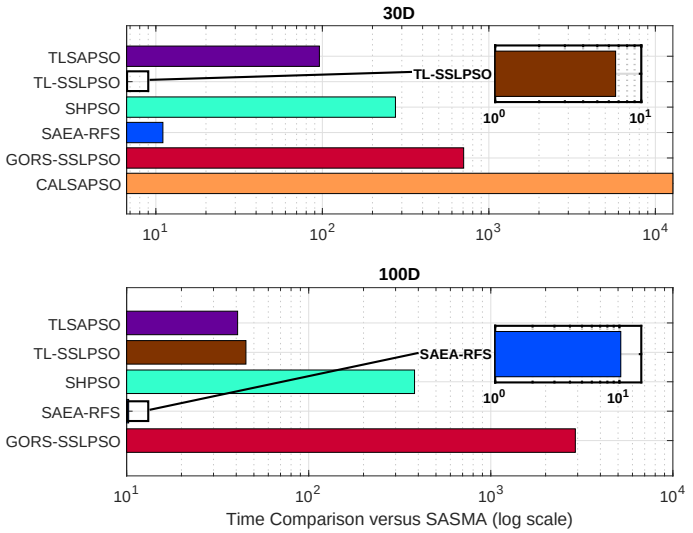


Figure 4.17: Relative time comparison between SASMA and the other SAEAs for 30D and 100D.

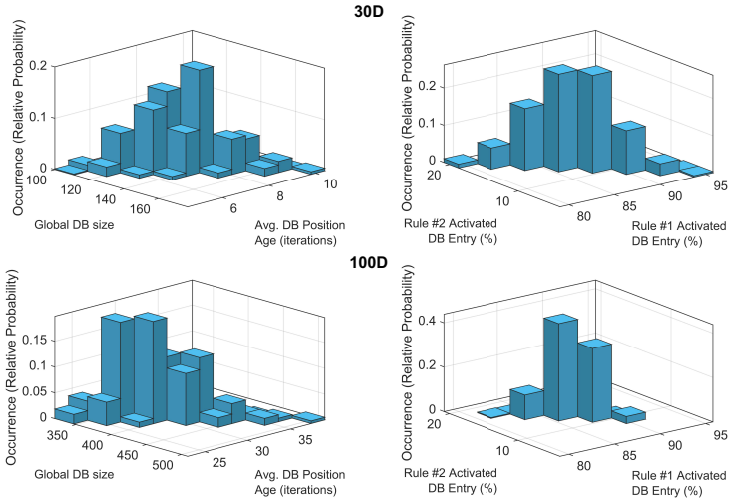


Figure 4.18: SASMA Database: Result Analysis for 30D (first line) and 100D (second line).

with rapidly improving positions. This scenario increases significantly for the 100D case, where the most frequent class for the DB size is between 374 and 401 stored positions, while the average age of the positions is around 26 to 27 iterations. Thus, indicating a more mature level of stored positions, which translates more constancy in the surrogate approximation.

These results are in line with the intended semi-elitist nature of the DB entry rules, designed in Algorithm 1 and shown in Figure 4.12. The same can be verified by the right side bi-variate distributions, where see that independently of the number of problem dimensions, rule 1 and rule 2 are responsible for roughly between 85/89% and 15/11% of the DB position entries, respectively, for more than half of the times (most frequent classes).

Chapter 5

Exploring Machine Learning: a deep-dive into artificial neural networks

Leveraging the vast amounts of data generated by modern power grids, power systems engineering has increasingly benefited from the advent of machine learning (ML) models to support a more efficient, reliable, integrated and sustainable operation (and planning) of EPSs. These algorithms have enabled breakthroughs in advanced data analytics, optimization models, simulation, predictive and modeling tools, real-time data processing, and much more. Breakthroughs, that, directly or indirectly, have enhanced the solution of major EPSs related tasks such as: resource allocation; maintenance scheduling; load, generation and price forecasting; energy management; monitoring, fault detection and diagnosis, among many others.

Given its prominent versatility and scalability features, as well as the treasure trove of data, supervised learning models based on artificial neural networks (ANNs) are a leading trend in the EPSs field development. As such besides some introductory and conceptual themes in ML theory, the main focus of this chapter will be on key practical issues surrounding the implementation of these potent tools.

5.1 Introduction

The renowned Dartmouth Conference of 1956 is considered by most, the outset of artificial intelligence (AI) as a formal field of study [380]. Laying the groundwork for ML to emerge as a subset of this broader field in the mid- to late twentieth century, the basic premise of ML is to enable computer systems to learn from data, without being explicitly programmed, as the name itself suggests. The use of the broad term ”learn”, in and of itself, would lead us to a much bigger discussion, but in this context, it means being able to gradually improve its performance on a specific task (quantifiable metrics that assess the model performance) [381]. It draws inspiration from the conceptual understanding of how the human brain works and performs certain tasks [382], specifically those involving pattern recognition, encompassing a broad-spectrum of statistical and computational algorithms/techniques, involving among several others, combinatorial optimization, convex optimization, mutual information, entropy and graph theory. Common examples of ML algorithms include decision trees, support vector machines, k-nearest neighbors, and also simpler neural networks.

The next revolution in the field of AI came in the early 2010s, with the advent of deep learning (DL), also known as representation learning. By taking advantage of the increasingly available datasets (advent of big data) and, more importantly, by leveraging the advancements of multi-core central processing units (CPUs) and the high parallelism and

memory bandwidth of graphics processing units (GPUs), the training time of DL models has been significantly shortened, making it feasible to scale these models in terms of both architecture and volume of training data [383]. As a general understanding, deep learning models need to have at least three layers [384], but they are typically composed by hundreds or thousands of layers [385]. Shallower models and early neural networks (with fewer layers) fall under traditional ML umbrella.

With these dedicated features, the capabilities of multilayered neural networks to handle raw data more effectively, have become apparent. Namely, the automatic feature extraction features that emerge from the added depth, and so rendering them a powerful tool to approach more complex problems like natural language processing, image and speech recognition, and more [386]. Thence, DL models streamline the workflow by avoiding the need for expensive pre-processing techniques, such as normalization and dimensionality reduction, which otherwise with be common with more traditional ML methods. In other words, the multiple processing layers of DL models, allows them to learn complex representations from raw input data with multiple levels of abstraction, to then produce the desired output, a feature known as end-to-end learning [387]. A notable DL model, ResNet, which stands for residual network, has achieved significant breakthroughs in the field of image recognition. This type of convolutional neural network (CNN) relies on the use of residual blocks with shortcut connections that skip one or more layers, thereby effectively mitigating the vanishing gradient problem [388].

Given the promise of ML-driven technologies, there has been a surge in public interest in their potentialities, particularly their ability to simplify otherwise complex decision-making procedures. Consequently, many AI-related terms are often tossed around casually, that is, without an understanding of the scope and history of the main sub-fields and what they entail, therefore, it was important to begin this section with a brief contextualization, as summarized visually in Figure 5.1, before delving deeper into some more theoretical and specific aspects.

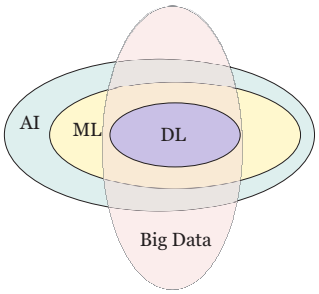


Figure 5.1: Deep learning in the artificial intelligence and big data landscape (diagram adapted from [386].

Meaning, clearly identifying AI as the overarching field, which translates a more illusive/abstract concept, ML as a comprehensive toolbox of techniques, that allow computers to learn from data, involving a series of stages ranging from data curation, pre-processing and expert based feature extraction, model building and finally, model evaluation, as illus-

trated in the Chevron diagram in Figure 5.2. While DL, which is exclusively implemented through various types of neural networks, has the capacity to learn complex representations directly from vast amounts of raw data, a feature known in the field as hierarchical representation learning, i.e., initial layers capture simpler patterns, while deeper layers detect more complex ones. Simply put, it no longer requires the features extraction and selection stage, as evidenced in Figure 5.2, which is now an inherent feature of these models.

Figure 5.2: ML traditional stages with the ”manual” features extraction stage accentuated in a dashed line to signal its absence in DL models (diagram adapted from [389]).

Ultimately, as data-driven algorithms, it is no surprise that the accuracy of ML models hinges heavily on the availability of a steady stream of well-curated/parsed data. Hence, the highlighted intersection of the AI landscape with the field of big data in Figure 5.1. That is, mostly what is defined as structured and semi-structured data (for example HTML or JSON files), where some form of organization/standardization is used to constitute relational databases or tabular formatted files [390].

5.2 Types of learning

A fundamental distinction between ML algorithms/models lies in its learning process, i.e., the way the available data is employed to train the model and how the model learns (supervision). Depending on the data characteristics (availability) and the application requirements, a first divide can be made between offline and online training. In the first approach, the learning samples derive from a static but comprehensive dataset, where all the training data is available and so it is fed as a large (full-) batch to train the model [391,392]. Hence, it is also known as ”batch learning”, an approach that fits relatively data-invariant training sets, i.e., when data is not changing much (low frequency in data collection) and/or the changes are somewhat consistent with the considered training batch. This approach results in a thorough training process, which is computationally expensive, with more pronounced weight adjustments. Leading to a more stable model, which on the downside, is less adaptive to potential data drifts in forthcoming samples (i.e., a ’drift’ in the data distribution properties).”
In contrast, with the increasingly popular online (or incremental) training approach, the training data is fed gradually over time (in a serial fashion as it becomes available), and so, the model is continuously updated, in a more computationally affordable way. Considering either mini-batch learning (small sized batches) or even stochastic learning (model weights are updated after each individual training sample is processed). Again, in opposition to offline (batch) training, the incrementally adapted model (with consecutive, less

pronounced weight adjustments) results in a less stable and noisier model (especially for stochastic gradient descent algorithms). On the upside, it is more responsive to dynamic data environments (common in real-time decision-making) and is suitable when there is a shortage of sufficient (past) training data to consider a batch approach.
Moving on to the main divide in ML, which encompasses the main types of learning approaches, with most of the literature pointing to at least four [381, 390], namely:

- Supervised learning: probably the most intuitive approach to ML, the model learns from labeled training data, provided by the (human) supervisor. In other words, presented with a cherry-picked set of input-output pairs, a ML model is built based on a ”function” that maps the fixed-size input to the desired (target) fixed-size output. Regression (data fitting) and classification (data categorization) are the most common examples of supervised tasks;

- Unsupervised learning: contrary to the prior type, this approach relies only on unlabeled data, without the intervention of any supervisor, to indicate (know beforehand) the target output. That is, the model is ”left on its own” in trying to capture some hidden patterns, clear divisions and grouping results in the processed input data. A learning approach that requires tailored learning rules, and to a great extent relies on feedback and self-organizing mechanisms to uncover underlying patterns and elaborate meaningful data representations. As such, feature learning, clustering and dimensionality reduction are obvious examples of unsupervised tasks;

- Semi-supervised learning: as the name indicates, it is a mixture of the two prior types of learning, where the model is built around a set of labeled and unlabeled data, i.e., an incomplete batch of training data (usually with many missing target outputs). It is a particularly useful approach, when the amount of labeled is scarce, contrary to the amount of unlabeled data, and so it is better to employ a semi-supervised ML model, rather than one exclusively based on labeled data. Clustering and classification are two common tasks associated with this learning type;

- Reinforcement learning: considered an environment-driven approach, it involves models built around a feedback scheme of rewards or penalties (learn by trial and error based on a reinforcement signal), with the ultimate objective of learning to maximize rewards (reinforcement). This approach can be integrated in DL models, with significant improvements in terms of model generalization, which adds to the purpose of reducing the learning computational burden. Again, it is useful when there is not a complete understanding or even lack of information regarding the desired output, and so the error feedback is not given as labeled input-output pairs. Complex control and classification tasks, such as driving a vehicle or playing a game against an opponent, are real-world applications of these models.

Therefore, the structure and availability of the data, as well as the targeted outcome and type of task at hand, will all influence the choice of the type of learning to consider, which

in turn, helps determine which type of ML models/algorithms to consider. Importantly, offline vs. online learning represents a fundamentally distinct approach on how the model is built (and updated), given a different way of handling the available (or soon to be) training data. And so it is applicable to all types of learning.

Overall, and as hinted in the introductory paragraph of this chapter, supervised learning is more commonly applied in power systems engineering than unsupervised learning. Leveraging not only the rich repository of historical data with known outcomes (known as labeled data), which makes it suitable to fit highly competent models (i.e., providing an accurate mapping of the relationship between input features and the output targets) but also the fact that many of the encountered problems are either regression (e.g., load forecasting and power flow analysis) or classification (e.g., fault detection) type tasks. As we saw, these are indeed types of supervised learning models, which serve different purposes depending on the nature of the predicted output [390]. Regression is a predictive modeling problem where the output is a continuous variable or label. Linear regression, polynomial regression, and support vector regression are some of the most common algorithms used for this task. Whereas, classification problems involve an output that is a categorical (discrete) variable. Logistic regression, decision trees, random forests, and support vector machines are some of the most recognized algorithms for solving classification tasks. Moreover, the different nature of the output variable also impacts the performance metrics used to evaluate the model. Mean squared error (MSE), root mean squared error (RMSE), and R-squared are typical options for regression models, while for classification models, the options usually include accuracy, precision, and F1 score.

This distinction can be visualized in Figure 5.3, where in the regression case, the dotted (regression or best-fit) line, models the predictive relationship between the independent variable (input) and the dependent variable (output); while in the classification case, the dotted line represents a decision (in this case linear as well) boundary that separates two classes (i.e., represents the threshold at which the classifier changes its prediction from one class to another).

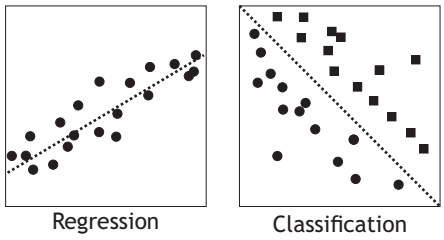


Figure 5.3: Supervised learning: regression vs classification.

Notwithstanding that, unsupervised learning approaches can also be of interest in clustering and feature extraction tasks. For instance, customer segmentation based on usage patterns, identification of trends in aggregate electricity demands or even the use of K-Means clustering to better understand and predict cloud behavior for an improved and more granular PV forecasting.

5.3 Artificial Neural Network: a prominent supervised learning model

The concept of ANNs dates back to 1943, inspired by the functioning of the human brain, particularly the synaptic connections established between the billions of biological neurons (nerve cells), McCulloch and Pitts introduced the groundbreaking concept of a threshold logic unit. The next breakthrough occurred in 1957, when Rosenblatt conceived the perceptron, enabling the training of a single artificial neuron, and so laying the foundations for supervised learning [393]. The subsequent four decades witnessed significant advances, including the work of Werbos in 1974 to formalize backpropagation as a training algorithm; successive works on multilayer perceptron; Kohonen's proposal of the unsupervised learning model, Self-Organizing Map (SOM); incremental works on recurrent neural networks during the 80s, among many others. But as we have discussed, it was not until the 2000s, with developments in computing power and the proliferation of large amounts of data, that the field truly took off, leading to its widespread applications.

Having reviewed some of the key milestones in the history of artificial neural networks, it is crucial to understand their main "building block", the neuron. A biological neuron, as shown in Figure 5.4 a), is composed of several dendrites, which receive impulses from other connected neurons, acting like an boolean input in our "artificial information processing unit" analogy. The cell body (soma) is where the linear summation of the different input signals (arriving from the multiple dendrites) is evaluated, and so it was mathematically modeled via an activation function with a threshold, θ. Then, at the end of the cell body, there is a single axon, with multiple output terminals, channeling the processed information to the other neurons (via their dendrites). Lastly, synapses are modeled as weights, determining the strength of the connections and directly influencing the learning process [394].

This relatively simple building block is then arranged in a more complex multilayered architecture (with different specialization levels), as shown in Figure 5.4 b). Starting with a mandatory input layer that receives data (analogous to information/stimuli), which activates or does not activate a neuron's response. This individual response is then fed to the first hidden layer, where a similar process takes place. Information is sequentially fed to adjacent neurons in subsequent hidden layers (if they exist), until a result or decision is made in the mandatory output layer (analogous to the brain's effector responses). During the feed-forward process, hidden layers apply successive non-linear transformations to the input data, projecting it into higher-dimensional feature spaces. These transformations are the key enablers of learning in ANNs, allowing models to capture complex patterns and relationships. By the time the transformed data reaches the output layer, hopefully it is structured to become linearly separable for classification tasks (allowing for a confident assignment of class labels) or precisely mapped to target values for regression problems (yielding accurate continuous predictions). This outcome, however, depends on the model's architecture, the quality of the training process, the representativeness of

the training data, among other factors.
As we saw when introducing DL, a neural network with an increased (substantial) number of these layers is ”deeper”, hence, leveraging the powerful hierarchical representation learning traits conferred by this extra depth. While a ”denser” network, is normally associated with the number (density) of neurons per layer (also known as width of the network). Translating an added flexibility and representational power, the potential to enhance model's generalization, in a subject that will be further explored in subsection 5.3.2.2.

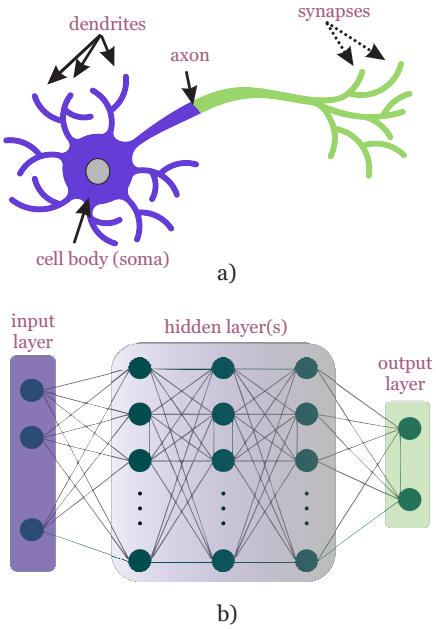


Figure 5.4: Working mechanism of ANNs: a) basic structure of a biological neuron; b) ANN layered structure.

As noted earlier, optimizing EPS operations and planning frequently involves tackling classification problems, and even more so, regression problems. With load forecasting, voltage control, or anomaly detection among the common examples, it is no surprise that supervised learning models, in particular ANNs, are ordinarily employed. This is a justified choice, given their well-documented pattern recognition abilities, versatility and scalability traits. ANNs can be configured in numerous architectures to better suit the task at hand. They are capable of capturing complex, non-linear relationships in large datasets and performing automatic feature extraction, which is otherwise difficult to achieve with ”manual” approaches.Nevertheless, due to their black-box nature, these ”skills” come with the trade-off of reduced interpretability. This stands in clear contrast to simpler statistical models, such as linear regression, where the model coefficients directly reflect the relationship between the input variables and the target output.
As expected, with these powerful abilities, ANNs and other versatile ML models have been applied in many shapes and forms across nearly every research field, resulting in abundant scientific and technical literature. In spite of this, their application is often not accompanied by a comprehensive understanding of several core issues, namely: i) data selection, pre-processing and cross-validation, feature extraction and dimensionality reduc-

tion; ii) model selection, type of layers, depth and hyperparameters tuning; iii) learning rules and methods (gradient descent based), validation and regularization techniques to promote generalization and avoid over-fitting; iv) performance metrics, stopping criteria, testing and model deployment. In other words, issues regarding the preparation of data, the different models, architectures, and its specificities, key training and performance aspects, which by all means, are crucial to assemble/configure a customized approach.

5.3.1 Forward Propagation (Feedfoward processing)

To fully grasp the concepts surrounding ANN model architecture and its tuning, as well as the basis for backpropagation (gradient error calculation and reverse propagation for weight adjustment), it is necessary to first understand the basics of its forward propagation, that is the sequential computation of neuron outputs. Starting with the input data and ending when the network error is computed.
Building on the revolutionary concepts introduced by McCulloch-Pitts neuron model, namely its weighted sum of the (boolean) inputs, and the respective (static) threshold-based activation, as expressed by the processing equation 5.1, Rosenblatt's perceptron neuron/network model broadened its application. By conferring it the ability to process real-valued inputs, this ML model, is able to approximate a wider range of data, and so unlike the McCulloch-Pitts, is no longer confined to approximating only logic functions. An ability that is also consequence of the proposed perceptron learning rule (explored later in subsection 5.3.2.1), i.e, the introduction of trainable weights that can be automatically assigned to each input connection, thus, enhancing the capability to learn from and adapt to labeled data [395].

$$y = u\left(z\left(\boldsymbol{\omega}, \mathbf{x}\right) - \theta\right) = \begin{cases} 0 & \text{if } z\left(\boldsymbol{\omega}, \mathbf{x}\right) < \theta \\ 1 & \text{if } z\left(\boldsymbol{\omega}, \mathbf{x}\right) \geq \theta \end{cases} \tag{5.1}$$

Where y denotes the binary artificial neuron output; $\mathbf{x} = (x_1, \ldots, x_i, \ldots, x_n)$ represents the neuron input vector, with $x_i \in \{0, 1\}$ standing for each individual boolean input (from a total of n network inputs); $\boldsymbol{\omega}$ represents the weights vector, with ω_i expressing the (individual) "strength" of the i^{th} input to the output y; $z\left(\boldsymbol{\omega}, \mathbf{x}\right) = \boldsymbol{\omega}^{\mathrm{T}}\mathbf{x} = \sum_{i=1}^{n} \omega_i \times x_i$ denotes the linear weighted sum of the input vector; and $u\left(z\left(\boldsymbol{\omega}, \mathbf{x}\right) - \theta\right)$ denotes a shifted Heaviside unit step function, with θ indicating the threshold from which the neuron is activated.
Notice that, despite the fact that neuron weights, $\boldsymbol{\omega}$, being iteratively modulated and the inputs not restricted to be Boolean, i.e., $x_i \in \mathbb{R}$, Rosenblatt's perceptron still processes the neuron output in the same manner as given in 5.1. However, a standard practice in modern perceptron representation, as shown in Figure 5.5 (adapted from [396]), is to incorporate (add) a bias term to the weighted sum term, together with an input value of 1 prepended to the input vector. This can be expressed as: $z\left(\boldsymbol{\omega}, \mathbf{x}, b\right) = \boldsymbol{\omega}^{\mathrm{T}}\mathbf{x} + \mathbf{b} = \sum_{i=1}^{n} \omega_i \times x_i + b$, with $b = -\theta$. This modification that simplifies the activation function

to the (normal) unit step function $u\left(z\left(\boldsymbol{\omega}, \mathbf{x}, b\right)\right)$. More importantly, it replaces the manually set threshold, θ, with a trainable bias term, hence increasing the model's flexibility. In other words, the learning process will adjust to the training data and determine if it is better to have a more loosely activated neuron (activating with lower input sums as it occurs with a negative bias) or, on the contrary, have a more stringent activation (requiring higher input sums as it occurs with a positive bias). With these considerations, one can write the modern (scalar/element-wise) notation for the feedforward perceptron processing equation, based on a linear transformation of its inputs using weights and a bias, as given by 5.2.

$$y = u\left(z\left(\boldsymbol{\omega}, \mathbf{x}, b\right)\right) = \begin{cases} 0 & \text{if } \sum_{i=1}^{n} \omega_i \times x_i + b < 0 \\ 1 & \text{if } \sum_{i=1}^{n} \omega_i \times x_i + b \geq 0 \end{cases} \tag{5.2}$$

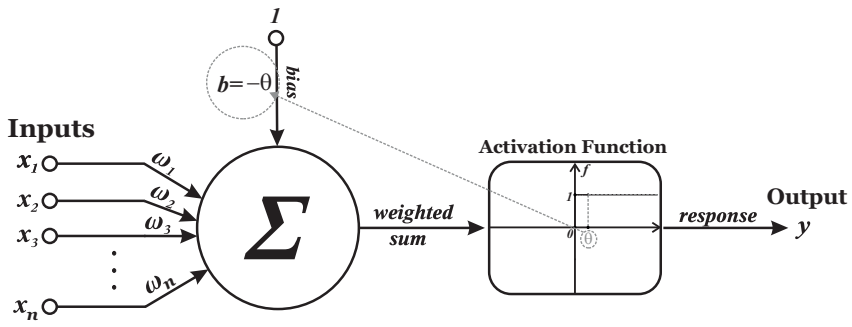


Figure 5.5: Modern perceptron representation.

Despite the progress with Rosenblatt's perceptron, and regardless of the added flexibility brought by the bias term in the modern representation, Minsky and Papert's 1969 work revealed the inability of this single-layered and single-node model to solve linearly inseparable problems (requiring a non-linear boundary), like the XOR logic function [397]. This means that both the McCulloch-Pitts neuron and Rosenblatt's perceptron are limited to linear separability (only able to create linear decision boundaries). This limitation fueled the works of Rumelhart and others on multilayer perceptron (MLP) models and backpropagation. It was found that with a multilayered architecture (including hidden layers, as shown in Figure 5.4) and non-linear activation functions, these models were now able to solve more complex, non-linear problems. With MLPs, forward propagation can be thought of as series of nested equations involving matrix multiplications and the subsequent element-wise activations.

Lastly, it is important to notice, that for historical reasons, and despite the name, MLPs, which are essentially multiple layer networks, can have neurons with any activation function (not restricted to being made up of perceptrons).

5.3.1.1 Activation functions

The decision surrounding which activation function to use is a field of study in itself. With different traits in terms of the function nature and shape, i.e., linearity, smoothness, being symmetric and zero-centered, saturation (when an activation function has regions where its output is constant or changes very slowly over a wide range of input values) and vanishing gradient robustness, computation cost and perhaps the most known trait, which is the range. With most of them scaling the output magnitude of the neuron between $[0, 1]$, $[-1, 1]$, $[0, +\infty]$ or $[-\infty, +\infty]$.

If we recap where we left off with the modern perceptron, the unit step activation function results in a binary "all-or-nothing" action of the neuron (the domain is split into two halves at $u(0)$). This behavior not only results in a "rough" transition, meaning a very small negative number does not trigger a neuron response, while a very small positive one does, but also note that $u(z)$ is non-differentiable and non-continuous at $z(\boldsymbol{\omega}, \mathbf{x}, b) = 0$ and for the rest of the domain has a gradient of 0. That is, no information is available regarding how small changes in weights may affect the output, thus rendering it unfeasible to compute the gradient and apply gradient descent algorithms for the training process. This is key aspect, since ideally, in the training process, a small $\Delta\omega$ should lead to correspondingly relatively small Δy, but when dealing with such an abrupt activation behavior, even a small weight change maybe the difference in triggering a neuron response or not.

Realizing all these shortcomings, many smooth (or smooth in parts) activation functions were proposed through the years, with most of them being non-linear. In fact, this is the key feature of most ANNs, as it enables them to learn complex non-linear relationships in the training data [396] by creating intricate decision boundaries, which are crucial for handling non-linearly separable data. If enough neurons and appropriate training are provided, a single hidden layer can approximate any continuous function to a desired degree of accuracy, known as the universal approximation theorem. Simply put, regardless of the number of layers, it all fundamentally traces back to the non-linearity introduced by activation functions and the network dimensionality, as showcased by Figure 5.6. Nevertheless, not all functions we want to approximate are continuous; most of the time, we are dealing with raw, noisy, and complex datasets. In practice, deeper networks, with their hierarchical representation features, as we will delve into in greater length in subsequent subsections, can learn more effectively.

Sigmoidal (S-shaped) functions (binary and bipolar) constitute the first major family of smooth activation functions in ANNs theory. Provided with relatively small values of $z(\boldsymbol{\omega}, \mathbf{x}, b)$, these functions enable that small changes in the neuron's weights and bias, can translate into a small change in its output, thus facilitating the training process. A change that is quantified below:

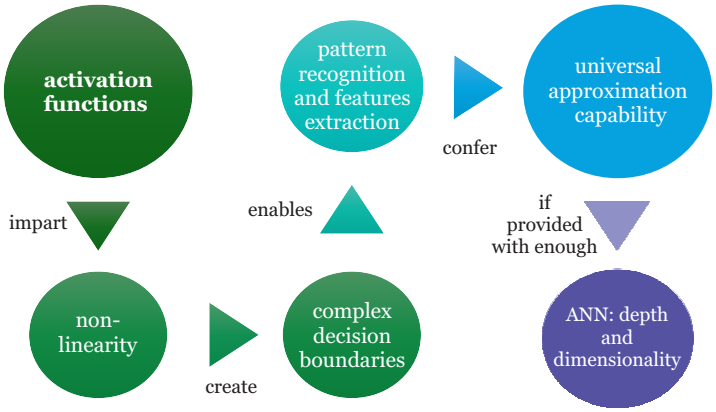


Figure 5.6: Activation functions as the primary (and for most of the architectures the sole) source of non-linearity in ANNs: a key enabler of their ability to capture and model complex data patterns (highlighted in a chain of key ANN characteristics/traits).

$$\Delta y \propto \overbrace{\sum_{i} \frac{\partial y}{\partial \omega_i} \Delta \omega_i + \frac{\partial y}{\partial b} \Delta b}^{\Delta z} \quad (5.3)$$

The partial derivatives of the output with respect to the weights and bias reflect how sensitive is the output to changes in these trainable parameters (the core of ANN training, as will be further explored in sub-section 5.3.2). An outcome that is dependent on the activation function choice. Therefore, the fact that the classic choice, $u(z)$, has a derivative of zero for all points except at $z = 0$ (where is not defined), renders it unusable for gradient-descent backpropagation methods. While for the sigmoidal family of activation functions, one can infer from 5.3, that Δy is proportional to a linear function of the changes $\Delta \omega_i$ and Δb, provided that a non-zero partial derivative can be found around $z = 0$, thus allowing meaningful weight updates. Nevertheless, despite their smoother (linear) in-between behavior, sigmoidal activation functions preserve the 'hard' bounds (asymptotic behavior) for less constrained weighted input sums. This is a key characteristic that ensures the information flow does not grow exponentially as it propagates from layer to layer. Instead, it remains conditioned [398].

The sigmoid function, as expressed in 5.4, which is also known as Log-Sigmoid or Logistic function, is the most well-known binary sigmoidal function. With an output range between 0 and 1, it transitions smoothly between its minimum and maximum value, and so it provides valuable gradient information for weight adjustment. This non-symmetric and non zero-centered function, is basically a smoothed out version (especially around the discontinuity point, $z = 0$) of the unit step function, as exemplified in Figure 5.7, can be computed as follows:

$$\sigma(z) = \frac{1}{1 + e^{-\alpha z}} \quad (5.4)$$

where α is the (slope) steepness parameter, with higher positive values associated with a smoother transition (less sensitive to small changes in z), and the opposite, i.e., a more abrupt transition for lower positive values (more sensitive to small changes in z).

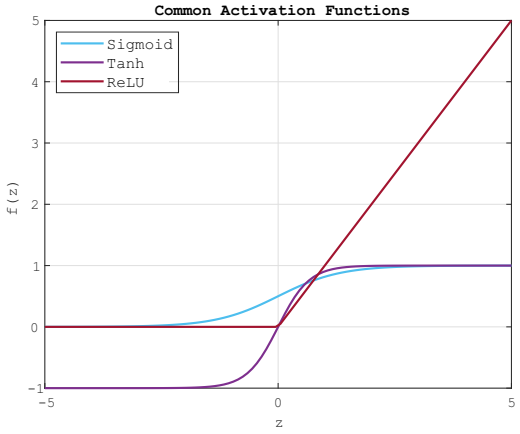


Figure 5.7: Common activation functions: Sigmoid, Tanh and ReLU considering $\alpha = 1.5$.

Similarly to the unit step function bipolar counterpart, the signum function, the bipolar sigmoidal function or hyperbolic tangent function (abbreviated as Tanh or Tan-Sig), as expressed in 5.5, has also an extended range of $[-1, 1]$ and is also symmetric and zero-centered, again as exemplified in Figure 5.7.

$$\tanh(z) = \frac{e^{\alpha z} - e^{-\alpha z}}{e^{\alpha z} + e^{-\alpha z}} = \frac{1 - e^{-2\alpha z}}{1 + e^{-2\alpha z}} \quad (5.5)$$

One trait that both $\sigma(z)$ and $\tanh(z)$ share is that they both saturate for even moderately small positive or negative values of z, i.e., both quickly approach their asymptotes: 1 and 0, respectively, in the case of the sigmoid, and 1 and -1, respectively, in the case of the hyperbolic tangent function. This clearly visible saturation translates into a very small gradient, which in turn results in very small weight adjustments, thus severely slowing down the learning process. Moreover, with deeper networks (MLPs, classic recurrent neural networks, and others), this problem is exacerbated. As the output layer gradient propagates backwards, small gradients become even smaller as they are multiplied, ultimately compromising the training process of these networks.

Unlike previous decades, when sigmoid and hyperbolic tangent activation functions — characterized by smooth nonlinear transformations — were widely adopted, the Rectified Linear Unit (ReLU) has gained significant prominence in more recent deep NN research. It is conceptually similar to a half-wave rectifier graph from electronics theory, as it outputs (preserves) only positive values, as illustrated in Figure 5.7. Mathematically, it is de-

fined by equation 5.6 as a piecewise linear activation function that is continuous, though non-differentiable at $z = 0$, in contrast to the step function $u(z)$. ReLU saturates at zero with a gradient of zero for all negative values and has a constant gradient of 1 for positive inputs. This constant non-zero gradient, applied during weight updates, effectively mitigates the vanishing gradient problem by ensuring stable and substantial gradients for positive inputs. Thereby, fostering a more efficient (faster) and consistent training process in deep NNs.

$$\text{ReLU}(z) = \max(0, z) \quad (5.6)$$

Other examples of non-linear activation functions include: Complementary log-log, Smooth Rectifier, Probit, Logit, Softmax, Softsign, Softplus and Modified ReLu's are other examples of non-linear activation functions.
A last note on the topic of activation functions, is that with the exception of some advanced, more custom-designed network architectures, neurons within the same layer share the same activation function, but different layers can have different activation functions depending on the problem requirements.

5.3.2 Training and backpropagation

As touched upon previously, training an ANN in supervised mode involves adjusting its weights based on a loss function, $L\left(\hat{y}_i\left(x_i, \omega, b\right), y_i\left(x_i\right)\right)$ or simply put, an error function of the (iterative adjusted) weights and biases, $L(\omega, b)$, given that both the input and desired output values are constant for the entire training process). In supervised learning, the choice of loss function varies, with a first clear distinction between those used for regression (continuous outputs) and classification (discrete outputs). So, given that the main focus of this thesis are regression type problems, a preferred choice is the mean squared error function (MSE), as given below:

$$L\left(\hat{y}_i\left(x_i, \omega, b\right)\right) = L(\omega, b) = \text{MSE} = \sum_{i=1}^{N} \frac{\left(y_i\left(x_i\right) - \hat{y}_i\left(x_i, \omega, b\right)\right)^2}{N} \quad (5.7)$$

where N is the number of training samples, $\hat{y}_i\left(x_i, \omega, b\right)$ is the actual (obtained) output response (activation value), and $y_i\left(x_i\right)$ the desired output response. Therefore, training can be thought of as the minimization of loss function(s), such as equation 5.7, which ultimately means finding the weights and biases that, for all the training inputs, approximate the actual output to the desired output.
The popularity of quadratic loss functions is explained by their convex and smooth nature, which allows error gradients to be computed during backpropagation. This, in turn, enables gradient descent algorithms to optimize the weights. Additionally, the quadratic term penalizes larger errors ($L(\omega, b)$>1) more heavily while reducing the impact of smaller

errors ($0<L(\omega, b)<1$), which are more frequent on later stages of training, thus benefiting from more precise weight updates. This provides a mix of fast convergence and model stability for a more accurate model. Notwithstanding, it also can be said that it is an outlier sensitive loss function [399]. Some noteworthy alternatives to MSE for regression tasks, include mean relative estimation error (MRE), mean absolute error (MAE), sum of squared residuals (SSR) and root mean square error (RMSE). Whereas, for classification tasks, cross-entropy loss tends to work well.

As illustrated in Figure 5.8, forward propagation concludes with the network output computation. This enables us to compute the loss score, which triggers the backward propagation of this error to all model parameters. Various gradient descent algorithms (which will be discussed later) then update the weights in a cyclic process. This continues until a specific stopping criterion, which gauges the learning process, is met.

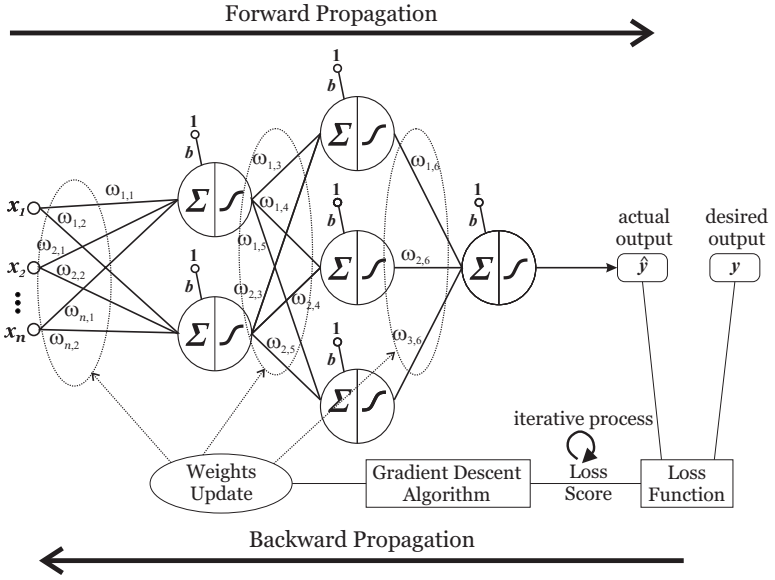


Figure 5.8: Forward and backward propagation in a generic 2-hidden-layer, single-node-output (layer) ANN: an iterative training process.

Backpropagation, which is the most commonly used training algorithm, is responsible for computing the gradient of the chosen loss function with respect to each weight by sequentially propagating the output loss backward through the network. This involves calculating the partial derivatives, which are then used to adjust the individual weights of each neuron, through the use of optimization techniques. This is an important clarification given that a frequent misconception around backpropagation algorithm is that it embraces the whole training algorithm, when in fact is up to gradient descent algorithms (or its variants) to use the gradient information and update the weights. That is to say, training and backpropagation are not interchangeable terms. On one hand, backpropagation is a key component (cornerstone) of the training process, used to compute and distribute the error gradients backward through the network layers ("an automatic" differentiation algorithm). On the other hand, training encompasses a broader process, that is closely related to the type of learning. It usually includes the use of some form of

gradient descent optimization, generically illustrated in Figure 5.9, to effectively reduce the ANN training loss through iterations by adjusting individual weights. The illustrated generic univariate case showcases a positive slope for the assessed weights, meaning that we constantly decrease the weight (move to the left) to find the weight that minimizes the training loss. Conversely, a negative slope signals that steps towards a weight increase (move to the right) lead us to loss minimization. These steps should be proportional to the amplitude of the slope, which helps avoid overshooting the minimum value when approaching a very low loss gradient with respect to the weights and bias [400]. Hence, a common analogy for this iterative process is to imagine a ball rolling on a hill with different velocities depending precisely on the slope of the current position.

Each of these steps, an iterative weights update, translates the predictive model learning. The optimization landscape of neural networks often contains multiple local minima and saddle points, which can create regions where the gradient path appears ”U-shaped” as illustrated in Figure 5.9. Consequently, a continuous step in one direction can eventually lead to moving up the gradient (undesirably). Furthermore, a key trait of practically all resulting loss surfaces (which illustrate how the loss function changes with different values of the network's parameters) is their non-convexity. This is a direct result of the model's complexity and its layered architecture, which establishes nonlinear interactions between neurons. Another source of the ANN's non-convexity is the symmetry between parameters (swapping some of the model parameters has no effect on the final loss computation), particularly in dense models. This means that if there is a combination of parameters leading to a certain local minimum, there will be several other permutations that will also lead to the same outcome. Moreover, merely considering gradient information does not guarantee that we are dealing with a local minima rather than a saddle point. As discussed earlier in subsection 4.2.1, a more robust approach is to use gradient descent methods that account for both first-order and second-order conditions.

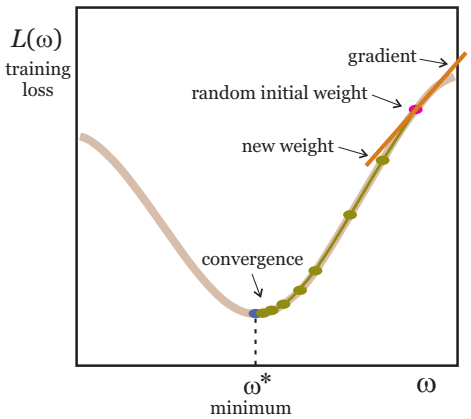


Figure 5.9: Gradient descent for training loss minimization illustration: weight update and convergence.

Withal, for practical cases, we are talking about thousands or even more individual weights and biases with inter-dependencies, and so we extend this univariate derivative to the de-

scending gradient vector (negative first-order information indicating the steepest downward direction) in this high-dimensional space: $-\nabla L = -\Bigg(\frac{\partial L}{\partial \omega_{11}^{(1)}}, \frac{\partial L}{\partial \omega_{12}^{(1)}}, \ldots, \ldots, \frac{\partial L}{\partial \omega_{1\,n_{\text{hn}}}^{(1)}}, \frac{\partial L}{\partial \omega_{21}^{(1)}}, \ldots, \frac{\partial L}{\partial \omega_{2\,n_{\text{hn}}}^{(1)}}, \ldots, \frac{\partial L}{\partial \omega_{n_{\text{inputs}}\,n_{\text{hn}}}^{(1)}}, \frac{\partial L}{\partial \omega_{11}^{(2)}}, \ldots, \frac{\partial L}{\partial \omega_{n_{\text{inputs}}\,n_{\text{hn}}}^{(2)}}, \ldots, \frac{\partial L}{\partial \omega_{i\,j}^{(l)}}, \ldots, \ldots, \frac{\partial L}{\partial \omega_{n_{\text{inputs}}\,n_{\text{hn}}}^{(n_{\text{hl}+1})}}\Bigg)$.
This representation was chosen to easily identify all the partial derivatives with respect to the weights across all the network layers. $-\nabla L$ denotes the negative gradient vector of the loss function, typically a concatenated column vector indicating the direction of steepest descent (determined by/during backpropagation method/phase), and with the amplitude of the slope now being generalized by the length/Euclidian norm of the gradient vector, $\|\nabla L\|_2$; $\omega_{i\,j}^{(l)}$ is the weight connecting the ith input to the jth node in the lth layer; n_{hn} denotes the number of nodes in the considered layer l, and is an abbreviated notation of the function $n_{\text{hn}}\,(l)$; by the same token $n_{\text{inputs}} = n_{\text{inputs}}\,(l)$represents the total number of inputs to a given layer l; which for the weights connecting (the input layer) to the first (hidden) layer, actually represents the total number of ANN inputs, denoted simply by n (following the notation started in equation 5.1), meaning that $\frac{\partial L}{\partial \omega_{n_{\text{inputs}}\,n_{\text{hn}}}^{(1)}} = \frac{\partial L}{\partial \omega_{n\,n_{\text{hn}}}^{(1)}}$; while for subsequent layers it represents the total number of outputs (nodes) from the previous layer, which for the current layer are "seen" as inputs, meaning that $n_{\text{inputs}}\,(l) = n_{\text{hn}}\,(l-1)$; n_{hl} is the total number of hidden layers and so, in accordance, $n_{\text{hl}+1}$ is the superscript used to denote the output layer (weights); hence this notation adapts appropriately for each connected layer.

Perceptron rule, evolutionary algorithms and reinforcement learning approaches are some of the few alternatives to backpropagation, although they might not be as effective or applicable (some are architecture dependent).

Returning to the actual weight update scheme, again as illustrated in Figure 5.9, we can connect the terminology used to describe the different batch size approaches when comparing offline vs online training in chapter 5.2 with the applied gradient descent algorithms. We have previously indicated that backpropagation occurs after each full forward pass (epoch), which is indeed true, but only for (full) batch gradient descent, when an epoch and an iteration are indeed the same, given that the entire training dataset is used to compute the gradient of the loss function with respect to the weights. As such a more precise wording would be: The cyclic model's weights and biases adjustment takes place after each iteration. With the number of iterations per epoch being determined by dividing the number of training samples N (as defined in equation 5.7) by the user-defined mini-batch size (usually a power of 2).

Given that the training data is usually divided into mini-batches, this effectively implies that one epoch consists of multiple iterations, which in turn, means more frequent weight updates compared to the full batch approach. As a consequence, mini-batch learning tends to accelerate model convergence, with a good enough compromise in terms of model accuracy, making it more practical for very large training datasets [401]. Additionally, this

approach introduces randomness (stochasticity) in the sampling (composition) of these smaller batches for every epoch, which results in gradient variability even if the overall direction is the same, thus making it less prone to local minima problems, common with non-convex optimization problems.
Taking this concept to an extreme, stochastic gradient descent (SGD) is a special case of mini-batch learning, where the model is updated after each training sample (a single labeled training pair is used for the forward and backward pass). This means that the number of iterations per epoch is equal to the number of training samples (N), as illustrated in Figure 5.10. SGD is therefore the ”best equipped” approach to escape local minima and thus find the global best loss. Yet it comes with the ”cost” of a greater fluctuation in the gradient decent path (with some erratic jumps), hindering the convergence speed.

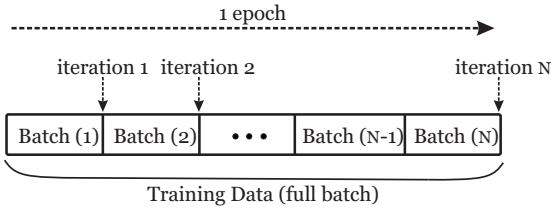


Figure 5.10: ANN stochastic gradient descent training: epoch, iterations and batch size.

In opposition, full batch learning (also known as plain vanilla learning) offers a deterministic gradient descent path provided with the same starting point (training data) and a fixed learning rate. This, makes it particularly suited for approximating convex functions, as the fuller account of the gradient error ensures a steadier convergence path towards the minimum. In more detail, the less frequent weight updates are based on the average gradient over all the training samples, rather than just a portion (a larger value of N in 5.7 implies a more representative average), translating into less variance in the weight adjustments. On the downside, besides the ”prohibitive” computational cost of this approach for very large training datasets, this steadier course, being less susceptible to extremes and outliers, results in a diminished exploration of the function surface and, consequently, may be more easily trapped in local minima. With this said, one can understand why mini-batch gradient descent — which is generally faster than SGD and it provides a greater exploration of the loss surface than plain vanilla (due to the randomness associated with the mini-batches) — is often the preferred approach. An illustration of the deterministic versus the stochastic nature of these different batch size approaches in the path towards the minimum is given in Figure 5.11.

5.3.2.1 Learning rules and gradient descent methods (ANN weights update)

Proposed in 1949, the Hebbian weight update rule is one of the oldest learning rules for ANNs [396]. This very simplistic weight update mechanism is inspired on an emblematic premise that states: ”cells that fire together, wire together”. In practical terms, the weight of a synaptic connection is adjusted according to the synchronization level between the neuron input and its output, through the following expression:

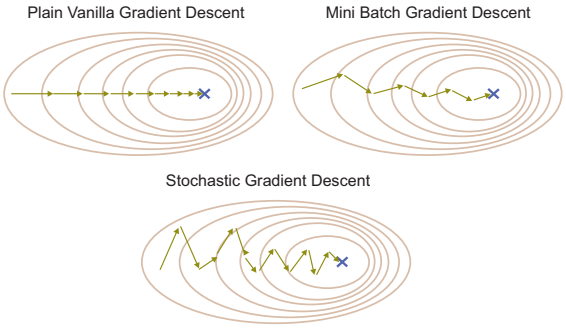


Figure 5.11: Gradient descent behavior for the different batch size variants.

$$\Delta\omega_i = \eta x_i \hat{y} \quad (5.8)$$

where $\Delta\omega_i$ is the computed weight change associated with the ith input feature, x_i (presynaptic connection); $\hat{y}$ denotes the output (predicted label) from the receiving neuron (postsynaptic); η, which up until this point we have not discussed, is a crucial positive constant (usually between 0 and 1) responsible for balancing training speed, convergence, and model accuracy. It is a hyperparameter of many learning (weight update) algorithms, including gradient descent ones. A too high learning rate enables the model to learn rapidly by taking larger strides (larger updates to the weights), at the cost of fast oscillations near the minimum, hindering convergence and risking overshooting (underfitting scenario). On the contrary, a too low learning rate ensures a more refined search (very small $\Delta\omega_i$), at the cost of very slow convergence, increasing the likelihood of the model getting stuck in a local minimum (when the number of epochs is not compensated) [402]. Both scenarios lead to sub-optimal training outcomes, as illustrated in Figure 5.12 for both the loss function as function of the weight plots and the (training) convergence curves (loss function as a function of the epochs plots).

The discussion surrounding the learning rate goes beyond its initial and fixed value. Figure 5.12 clearly demonstrates that a too small value of η ensures that the weight adjustments are always in the right direction. However, as a downside, it implies a very large number of epochs to start converging, which is time-consuming. Conversely, a high value of η is more computationally efficient due to the accelerated initial convergence phase. But this comes at the expense of more erratic steps (notice the change in the gradient direction or sign in the several weight adjustments for this scenario, in clear opposition to the steady convergence of the other two), which as we saw may lead to overshooting the minimum or even divergence (as illustrated in the very high η scenario). For these reasons, the use of a constant learning rate is not recommended. Instead, a varying learning rate throughout the different phases of training can dramatically improve the effectiveness of gradient descent algorithms [404], and so it is deemed a key tunable ANN (training-related) hyperparameter. Based on these assumptions, a straightforward solution is to have the learning rate decay over the epochs, with exponential decay, step decay, and inverse decay being

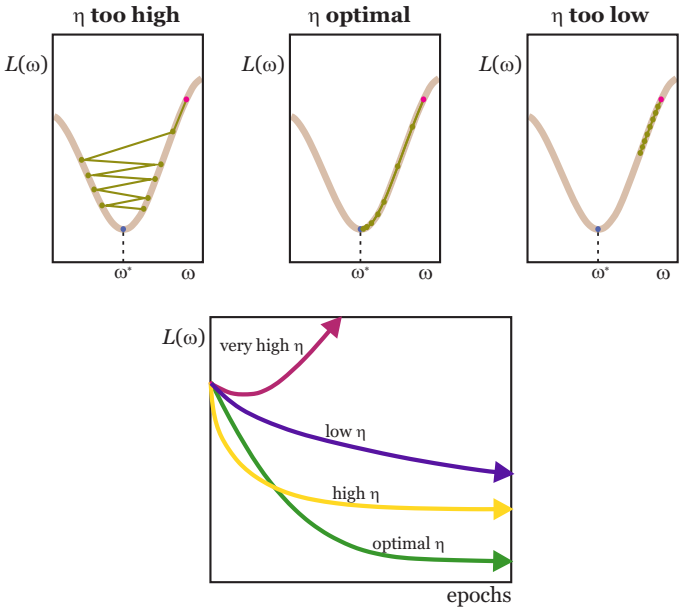


Figure 5.12: Effect of the learning rate parameter η (adapted from [403]): Loss function as a function of a given weight and Loss function as function of the epochs.

popular choices to achieve this goal.

Having concluded a brief detour on the importance of a calibrated learning rate—a tunable hyperparameter shared by virtually all learning rules—for the overall ANN training performance, I will now return to the subject of learning rules. Having described Hebb's learning rule, which was effectively the first, Rosenblatt's perceptron rule, already hinted at in section 5.3.1, was the next development. This rule iteratively adjusts the weights based on labeled training data examples, the difference between the predicted label $\hat{y}$ and the true label (desired output) y, as given below:

$$\Delta\omega_i = \eta\,(y - \hat{y})\,x_i \quad (5.9)$$

Nevertheless, due to limitations of the perceptron neuron model (nature of the output data), the model and its companion learning rule do not explicitly minimize a loss (error) function; instead, it only works as a binary classifier for linearly separable problems. The next breakthrough came with the work of Widrow and Hoff and their adaptive linear neuron (ADALINE). This neuron model resembles the modern perceptron model (as described in section 5.3.1), but instead of using the classic step function, it relies on a linear activation function. This means that the predicted label is continuous rather than binary, making it suitable for regression tasks.

This breakthrough revolutionized the learning framework by incorporating the concept of a loss function for weight adjustment. To this end, Widrow and Hoff proposed an iterative algorithm known as Least Mean Squares (LMS), which encompasses weight initialization, the forward pass, error calculation, and weight updates via the delta rule (an equation

similar to the perceptron rule 5.9, but with the error term now producing a continuous value instead of just 1, -1 or 0). These steps are then repeated multiple times to minimize a loss function, such as the MSE (equation 5.7).

Werbos' groundbreaking work on backpropagation enabled the application of LMS to multilayered networks, a weight adjustment procedure that was subsequently refined by Widrow, Winter, and Bunter in the late 1980s [396]. To properly understand it, one can explore the (landmark) generalized delta rule, as given below:

$$\Delta\omega_{ij} = -\eta\nabla L\left(\omega_{ij}\right) = -\eta\frac{\partial L}{\partial\omega_{ij}} \quad (5.10)$$

where η is again the learning rate; $\Delta\omega_{ij}$ is the computed weight adjustment for the ith input of the jth receiving neuron; and $\nabla L\left(\omega_{ij}\right)$ is the gradient of the loss function L with respect to the weight ω_{ij}, which is a more compact way of expressing the partial derivative. As we can see right way, the first difference is the negative sign, which is no strange fact, given that $\frac{\partial L}{\partial\omega_{ij}}$ indicates the direction of the steepest increase in the error function, so multiplying it by -1, ensures that the weight update moves in the opposite direction (hence the term "descent"). That is the direction that minimizes the loss function. The other difference is that there is no explicit mention of the input layer entries (x_i) in the equation. However, the generalized delta rule still accounts for the input, but through the gradient calculation, as backpropagation eventually reaches the first hidden layer, which connects directly to the input layer. And so, all the ANN inputs, x_i, are still indirectly accounted in the computed weight update.

Considering the composite activation function, $(\sigma\circ z)(\boldsymbol{\omega},\mathbf{x},b) = \sigma\left(z\left(\boldsymbol{\omega},\mathbf{x},b\right)\right)$, knowing that its derivative is given by $(\sigma\circ z)'(\boldsymbol{\omega},\mathbf{x},b) = \sigma'(z(\boldsymbol{\omega},\mathbf{x},b))\cdot z'(\boldsymbol{\omega},\mathbf{x},b)$; with $\sigma\left(z\right)$ given by equation 5.4 with $\alpha = 1$; $z\left(\boldsymbol{\omega},\mathbf{x},b\right) = \sum_{i=1}^{n}\omega_{i,j}\times x_i + b_j$ denoting the jth node response to every ith input (from a total of n), where $\omega_{i,j}$ is the weight associated with the connection from the ith input to the jth neuron, x_i being the ith input (which in this notation can either be an input signal from the input layer or an output signal from a preceding hidden layer neuron response) and b_j being the individual bias term for this jth node; and to simplify gradient calculations, the selected loss metric being the sum of squared residuals (SSR) multiplied by a factor of $\frac{1}{2}$, $L\left(\omega,b\right) = \mathrm{SSR} = \sum_{i=1}^{N}\frac{\left(y_i(x)-\hat{y}_i(x,\omega,b)\right)^2}{2}$, which, much like MSE, favors error magnitude over signal (direction). With this preamble set, we are now able to expand equation 5.10 using the chain rule:

$$\Delta\omega_{ij} = -\eta\left(\frac{\partial L}{\partial\sigma_j}\cdot\frac{\partial\sigma_j}{\partial\omega_{ij}}\right) = -\eta\left(\frac{\partial L}{\partial\sigma_j}\cdot\frac{\partial\sigma_j}{\partial z_j}\cdot\frac{\partial z_j}{\partial\omega_{ij}}\right) \quad (5.11)$$

Now if we compute the individual partial derivatives:

$$\frac{\partial L}{\partial \sigma_j} = \frac{1}{2} \cdot 2 \cdot \left(y_i\left(x\right) - \hat{y}_i\left(x, \omega, b\right)\right) \cdot -1 = -\left(y_i\left(x\right) - \hat{y}_i\left(x, \omega, b\right)\right) \quad (5.12a)$$

$$\begin{aligned} &\frac{\partial \sigma_j}{\partial z_j} = -1 \cdot \left(1 + e^{-z_j}\right)^{-2} \cdot \left(-1 \cdot e^{-z_j}\right) = \frac{e^{-z_j}}{\left(1 + e^{-z_j}\right)^2} \Leftrightarrow \\ &\frac{\partial \sigma_j}{\partial z_j} = \frac{1}{\left(1 + e^{-z_j}\right)} \cdot \frac{e^{-z_j}}{\left(1 + e^{-z_j}\right)} = \frac{1}{\left(1 + e^{-z_j}\right)} \cdot \frac{\left(1 + e^{-z_j}\right) - 1}{\left(1 + e^{-z_j}\right)} \Leftrightarrow \\ &\frac{\partial \sigma_j}{\partial z_j} = \frac{1}{\left(1 + e^{-z_j}\right)} \cdot \left(1 - \frac{1}{\left(1 + e^{-z_j}\right)}\right) = \sigma_j \cdot \left(z\right)\left(1 - \sigma_j\left(z\right)\right) \end{aligned} \quad (5.12b)$$

$$\frac{\partial z_j}{\partial \omega_{ij}} = \sum_{i=1}^{n} x_i \quad (5.12c)$$

From these auxiliary calculations we can now resume equation 5.11 and establish the weight update expression for the generalized delta rule:

$$\begin{aligned} \Delta\omega_{ij} = -\eta \left(-\left(y_i\left(x\right) - \hat{y}_i\left(x, \omega, b\right)\right) \cdot \sigma_j\left(z\right) \cdot \left(1 - \sigma_j\left(z\right)\right) \cdot \sum_{i=1}^{n} x_i \right) \Leftrightarrow \\ \Delta\omega_{ij} = -\eta \left(\delta_j \sum_{i=1}^{n} x_i \right) \end{aligned} \quad (5.13)$$

where $\delta_j = -\left(y_i\left(x\right) - \hat{y}_i\left(x, \omega, b\right)\right) \cdot \sigma_j\left(z\right) \cdot \left(1 - \sigma_j\left(z\right)\right)$ and it represents the error signal for the jth neuron adjusted by the activation derivative.

Ideally, when training ANNs using gradient descent-based approaches, the gradients computed for each layer should be parallel (or nearly parallel). This alignment (similar optimization search direction) typically results in smoother convergence and more stable training dynamics. Nevertheless, the non-convex nature of the loss function, the inherent model complexity, the initial positioning, among a variety of other factors, often hinder this ideal scenario. Therefore, but not exclusively, a common approach to improve learning convergence and avoid unnecessary oscillations, a common technique in training neural networks is the addition of a momentum term [405], as given in equation 5.14. The fact that the current (a certain iteration k) weight update $\left(\Delta\omega_{ij}\left(k+1\right)\right)$ also depends on a fraction of the previous one provides a kind of "inertia" to the gradient descent optimizer. This inertia helps the optimizer continue moving in a direction that consistently reduces the training loss, hence potentially achieving faster convergences. Additionally, by incorporating the cumulative information from past gradients, momentum provides a smoothing effect when navigating through high-curvature regions of the loss function. This helps mitigate the impact of noisy or steep gradients that may lead to erratic direction changes. Therefore, its application goes well beyond the generalized delta rule presented

here. In fact, it is considered in many other gradient descent methods such as the well-known adaptive moment estimation (Adam) method or the Nesterov accelerated gradient method, among others.

$$\Delta\omega_{ij}\left(k+1\right) = -\eta\left(\delta_j \sum_{i=1}^{n} x_i\right) + \phi\Delta\omega_{ij}\left(k\right) \quad (5.14)$$

where ϕ is the momentum coefficient.

The development of backpropagation marked the transition from local learning rules to a more global oriented optimization approach based on gradient descent methods for training ANNs. In a sense, all modern learning rules can be seen as specific implementations or variations of gradient descent, especially in the context of neural network training.

While gradient-based training may not always guarantee finding the global minimum due to the non-convex nature of the loss surface, one often ends up converging to one of the numerous local minima. At face value, this may seem contradictory to the conventional wisdom regarding any optimization problem, as we saw in Chapter 4. However, this is actually not a bad thing in the context of weight optimization, as we will see in Subsection 5.3.2.2. Settling at the global minimum point can often lead to a lack of generalization and poor testing performance of the network.

Despite the progress, it was soon realized that for more complex and high-dimensional problems, weight update methods that rely on the steepest gradient descent direction, such as the generalized delta rule (equation 5.13, can behave poorly in large-scale problems, given the excessive oscillations and slower convergence due to the ”zig-zag” path [406]. In addition, they can also be too sensitive to the chosen fixed or variable η. An answer to these shortcomings came from conjugate gradient (CG) methods, which are considered quasi-second-order methods. As the name suggests, they generate search directions that are conjugate with respect to the Hessian matrix H, without explicitly calculating this matrix. In contrast, standard backpropagation methods, such as the generalized delta rule, usually involve only first-order gradients $\nabla L\left(W_k\right)$, which is the loss gradient vector with respect to all the network weights W at a given training iteration k. These conjugate directions, denoted in a vector form as D_k, are orthogonal to the past search directions. Specifically, $D_{k+1}^T H D_k = 0$, which in practical terms means they avoid retracing already taken paths (oscillatory behavior), which in turn leads to a potentially faster and more stable convergence [407]. With the new direction and the weight update being given as [408]:

$$D_{k+1} = -\nabla L\left(W_k\right) + \beta_k D_k \quad (5.15)$$

$$\Delta W_{k+1} = \eta_k D_{k+1} \quad (5.16)$$

where $\beta_k = \frac{\nabla L(W_k)^T \nabla L(W_k)}{\nabla L(W_{k-1})^T \nabla L(W_{k-1})}$ is the scalar that ensures the orthogonality of the search directions; notice how the new direction (5.15) is computed based not only on the current loss gradient (steepest descent) but also on the information from the previous search direction; finally, η_k in equation 5.16 is now the CG step size (analogous to the learning rate η in traditional gradient descent methods) found via a line search process. In and of itself an expensive iterative process, given that the optimum step size is "unreachable", and so, in practice, the step size is actually determined in what is designated as "inexact line search". Nevertheless, "poor" initialization scenarios, i.e., when the initial weights vectors, W_0, is too far from the desired minimum, line search inadequacies, and the inherent inability to handle non-positive definite Hessian matrices, can often end-up compromising the performance of CG methods by settling the optimization at sub-optimal minima. Addressing one of these issues, Scaled Conjugate Gradient method emerged as a powerful ANN training method, particularly suited to train large networks (with a large W_k), mainly due to its step size scaling mechanism. It avoids an expensive line-search per learning epoch process and it eliminates the dependence on some critical user-defined parameters, requiring less memory usage, thus rendering it faster than more accurate but slower second order methods, like the famous Levenberg-Marquardt method.

Shortcomings, different training approaches, architecture-related variance, and the proliferation of different forward and recurrent processing mechanisms as led to the development of many ANN weight optimizers, that are mostly based on gradient descent methods. Prominent and modern options include the already mentioned Adam, but also root mean squared propagation (RMSProp) and of course, invariably, the stochastic gradient descent (SGD) method. Nonetheless, other methods, including second-order methods like Newton's Method, and metaheuristic algorithms such as Simulated Annealing or Genetic Algorithms can also be employed to this effect (ANN training).

To conclude, while the training process of ANNs is a well-established unconstrained optimization problem—loss function minimization over the training data with the possibility of considering constraints via the penalty method, as will be explored when addressing regularization in sub-section 5.3.2.2—and with a vast repertoire of training methods that have yielded successful results for many regression or classification problems, there are still some intricate details to fully grasping the inner workings of these models. This means there is ground to cover regarding the interpretability and explainability of neural networks, particularly in understanding their powerful universal approximation and interpolation capabilities (black box nature).

5.3.2.2 Initialization, convergence, generalization and regularization

The success of the ANN training phase is a direct proxy for its performance during testing. As described in prior sections, many factors and considerations must be taken into account to properly implement such encompassing ML models. In fact, in-depth knowledge of these related issues allows for a proper understanding of the training behavior via convergence curve analysis (loss function as a function of epochs/iterations). In turn, one can better gauge the adequacy of the training data, fine-tune the gradient descent method, and reassess ANN architecture choices. This knowledge enables further performance gains.

We begin with the issue of **weight initialization**. As one might imagine, as the starting point for gradient descent methods (weight optimizers), a poor initialization can result in slow convergence, getting trapped in local minima, or even divergence [396, 405]. A first broad division regarding different weight initialization techniques concerns initialization without pre-training versus initialization with pre-training. The first is the most common category and it includes random initialization techniques that can be either interval or variance-scaling based. Moreover it also includes data-driven initialization where the network weights are derived from the training data. The second category is less common and it involves the use of an unsupervised model to help place the network weights in a near-optimum place [409].

Regarding random initialization techniques, it is important to clarify that a purely random distribution of weights without any constraints is not a great option for the following reasons: i) Large initial weights can lead to exploding gradients during backpropagation, causing unstable updates and potentially even divergence in the training phase; ii) Conversely, small initial weights can lead to vanishing gradients, where weights are barely updated, leading to slow convergence or even stagnation; iii) The resulting weight initialization can be unevenly distributed which, besides the scaling issues (the aforementioned gradient problems), can hinder the convergence rate and increase the likelihood of the model getting stuck in sub-optimal solutions, particularly for high-dimensional spaces (large networks); iv) It does not ensure symmetry breaking between the network nodes. In other words, despite being highly unlikely, weights that are initialized with the same value in the same layer will end up having the same weight update (as given by equation 5.13), meaning that these neurons end up learning the same features, thus reducing the network's capacity to learn diverse and useful representations of the training data [410].

Focusing on the most common category of initialization techniques without pre-training, interval-based initialization was among the first. The guiding principle is to position the initial weight vectors within an "optimal" interval or maximum bound. In this subcategory, the Nguyen-Widrow initialization technique is considered a reference, providing a scaling framework to initialize the hidden layer(s) weights of feedforward neural networks. Considering the number of hidden neurons, n_{hn} in a given hidden layer m, and the number of inputs n_{inputs} (either outputs of prior hidden layers or the number of inputs in the input layer) arriving from the $m-1$ layer, this technique involves first determining

a scale factor $\beta = 0.7 n_{\text{hn}}^{\frac{1}{n_{\text{inputs}}}}$. Then, for all connections between layers $m-1$ and m, the weights are pre-initialized using a uniform distribution between -1 and 1, such that:

$$\omega_{ij}(0) = \varepsilon, \text{ where } i \in [1, n_{\text{inputs}}] \text{ and } j \in [1, n_{\text{hn}}]. \quad (5.17)$$

Immediately after this random pre-initialization, weights are rewritten based on their norm and the determined scale factor, as given below:

$$\omega_{ij} = \frac{\beta \omega_{ij}(0)}{\|\omega_{ij}(0)\|} \quad (5.18)$$

This scaling ensures that the initial weights are neither too large nor too small, effectively lessening the chances of encountering slow convergence issues during the training phase due to poor weight initialization. As for the bias parameters, the technique suggests a random uniform initialization between $-\beta$ and β.

A more modern approach to weight initialization is given by variance scaling-based initialization techniques. These techniques, after a random weight pre-positioning, scale the resulting vector to maintain a balance in the variance of activations (in the input and/or output layer) and gradients. This balance is crucial to ensure stable gradient backpropagation during training. In this sub-category, Xavier initialization, also refereed to as Glorot initialization, stands out as a popular technique to keep the scale of gradients roughly the same across all layers, which is again beneficial for the backpropagation phase. In simplified terms, and for neurons "fired" using Log-Sigmoid or Tan-Sig activation functions, the weights are initialized using a uniform distribution within the range [411]:

$$\omega_{ij} \sim \mathcal{U}\left[-\frac{\sqrt{6}}{\sqrt{n_{\text{inputs}} + n_{\text{hn}}}}, \frac{\sqrt{6}}{\sqrt{n_{\text{inputs}} + n_{\text{hn}}}}\right] \quad (5.19)$$

following the same notation as expressed in 5.18 for the input and output connections to a given neuron in a given network layer m, this is specifically tailored for layers with ReLU activation functions. This notable and simple non-linear activation function family suffers from the "dying ReLU" issue, where neurons can become inactive and stop learning if they output zero for any input [412], which makes it even harder to train extremely deep models. Therefore, He Initialization offers slightly larger initial weights to prevent the gradients from vanishing. Consequently, the weights are initialized using a uniform distribution within the range:

$$\omega_{ij} \sim \mathcal{U}\left[-\frac{\sqrt{6}}{\sqrt{n_{\text{inputs}}}}, \frac{\sqrt{6}}{\sqrt{n_{\text{inputs}}}}\right]. \quad (5.20)$$

Other technical documentation presents a different range, where only the absolute value of the lower bound is augmented by a factor of $\sqrt{2}$ (resulting in a more negative lower bound), i.e., $\omega_{ij} \sim \mathcal{U}\left[-\frac{\sqrt{12}}{\sqrt{n_{\text{inputs}}+n_{\text{hn}}}}, \frac{\sqrt{6}}{\sqrt{n_{\text{inputs}}+n_{\text{hn}}}}\right]$. Other expressions can be found for both initialization techniques when considering a normal distribution rather than a uniform one.

To address local minima problems, a standard procedure is to restart the training process, ensuring that different weight initializations are tested. Alternatively, random disturbances can be introduced to the weights during training. This procedure shares some conceptual similarities with pre-training techniques.

To sum up, by distributing and conditioning the initial weights appropriately, these techniques set the tone for the early dynamics of the training process (the basis for a more effective error surface exploration). This is why all weight optimizers, including gradient descent methods, rely heavily on proper weight initialization.

When assessing the **convergence** of the training loss score as the training epochs/iterations progress, it is important to keep in mind the critical concepts of underfitting and overfitting. While training loss is a valuable metric that provides initial insights, it is often insufficient as a standalone stopping criterion for the weights optimizer. One key facet of every training process is the plateauing of the training loss, a visual phenomenon translating the significant slowdown in the improvement rate of the model's accuracy, even though training continues. This is also referred to as learning speed slow-down [413]. It is, however, a fully expected behavior and is deeply connected with the issue of overfitting, which we will delve into next. This phenomenon can also derive from other factors such as hyperparameter selection, namely the considered learning rate, as well as the problem complexity itself. Specifically, in the presence of a rough optimization landscape (many local minima and saddle points), the training process can be slowed down or even halted (getting stuck in a local trap).

For instance, a model is said to be overfitting when it "memorizes" the training data, including noise and outliers (viewed as exceptions in the training set), instead of "learning" the "hidden" relationships between the labeled vectors of training data. This results in near "perfect" interpolation for the training data but poor performance on unseen data, as illustrated by the sharp turns and numerous peaks and valleys (zigzag pattern in 1D) in Figure 5.13. Overfitting is such an intrinsic issue in supervised machine learning that it is not only hard to avoid completely but also hard to ascertain its root causes. Despite the fact that overfitting is usually driven by a dataset that is too small, less representative, or too noisy, its root causes can vary extensively and include evaluation procedures, hypothesis complexity (a concept that associates the use of too many inputs with a more accurate but less consistent model), and of course the notable bias-variance trade-off principle (also illustrated in Figure 5.13), among others [413].

The risk of overfitting increases when minimizing the training loss is the sole targeted training metric or stopping criterion. Conversely, underfitting occurs when the model is too simplistic and fails to capture the underlying, less obvious relationships in the pro-

vided labeled training data. This leads to poor performance on both training and unseen data (notice the dissimilitude between the data points and the approximated model in Figure 5.13). Despite this clear theoretical distinction, identifying the ”blurry lines” that separate the underfitting and overfitting zones is far from easy. This highlights the nuanced nature of ANN training, where the goal is to achieve sufficient model complexity to capture meaningful data patterns without falling into the traps of overtraining. The aim is to perform well on new data in a smooth way that captures the overall trend without ”fitting” to every individual data point, as illustrated by the good balance plot in Figure 5.13.

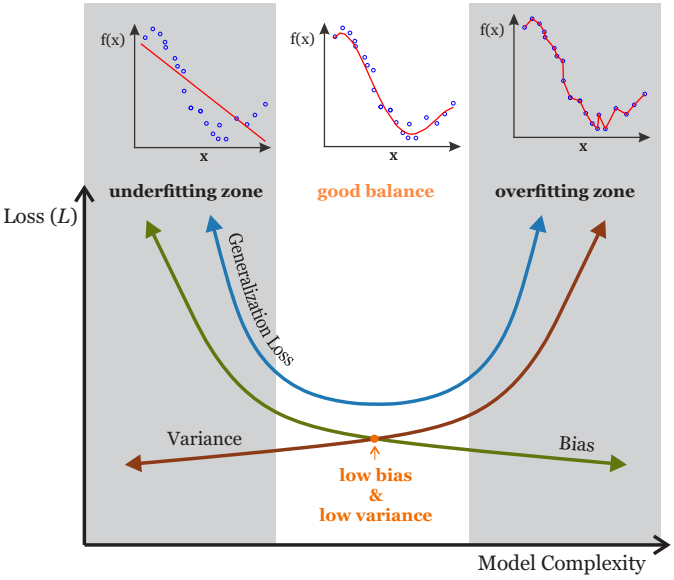


Figure 5.13: Bias-Variance trade-off illustration: underfitting, balanced fitting and overfitting.

The predicates of the bias-variance principle for ML and statistics, which basically states that ”the price to pay for achieving a low bias is high variance” [414], are key to connecting the concepts of underfitting and overfitting with model complexity and training data selection. In this specific context, bias is not another network (neuron) weight or parameter but is the ”error” stemming from the mismatch between the established model assumptions (after training) and the data distribution (including the more complex patterns). A model is said to have high bias when it is too simplistic (few parameters), thus failing to capture the underlying relationships in the training pairs of samples, leading to an underfitting situation. Naturally, the aim is to achieve the opposite, a low bias model, meaning a model that is flexible enough to capture those relationships accurately. That flexibility usually arises from a more complex/deeper model (essentially meaning more parameters). Nevertheless, one should be careful since, as the principle states, the pursuit of low bias can result in a too complex model, which generally leads to an overfitting situation and hence high variance. In turn, variance stands for the sensitivity to different training sets (measure of the fluctuation degree in the model’s prediction accuracy). High variance means that the model is very sensitive even to small changes in the training data, which, as we saw, is closely related to a too complex model that performs poorly on unseen data.

So, the aim is to have more stable and consistent accuracy across different sets of representative training data, put simply, low variance. But again, there is a risk since a too low variance may actually mean an underfitting situation. To sum it up, ideally, this delicate bias-variance trade-off targets a sufficiently complex model to ensure that it can properly capture the intricacies of the training data distribution without being too complex and risking overfitting [415].

This desired balance leads us to the key concept of **generalization**. In other words, an ANN should retain the ability to handle unseen data (generalize), by applying the "knowledge" gathered from the training examples, thus rendering it more than just an interpolation tool. For effective learning, the number of training samples and the input variables should correspond to the complexity of the regression problem at hand [416]. Although no strict rule exists for defining the exact number of training samples required for a given problem, a layman's and minimalist approach is to consider a number of samples that includes most of the temporal, spatial, and other types of relationships/patterns that would otherwise be statistically or visually meaningful from an expert analysis.

This means that theoretically, the main objective of training is to maximize the generalization traits of the network, and so its equivalent to minimize a theoretical generalization loss (L_{gen}). This illusive error can be computed by integrating the training loss, $L\left(\hat{y}\left(x,\omega,b\right),y\left(x\right)\right)$ as given by equation 5.7 multiplied by the joint probability distribution, $p(x,y\left(x\right))$, which represents the likelihood of different input-output pairs occurring. Mathematically, L_{gen} is expressed as:

$$L_{\text{gen}}=\int_{\mathcal{X}}\int_{\mathcal{Y}}L\left(\hat{y}\left(x,\omega,b\right),y\left(x\right)\right)\,p(x,y\left(x\right))\,dy\,dx \quad (5.21)$$

this integral sums up the expected generalization error over the entire range of (possible) data points, not just the ones seen during training—hence, the use of the terms "theoretical" and "illusive", given that the entire data distribution, i.e., set (domain) of inputs $\mathcal{X}$, and outputs $\mathcal{Y}$, is often not known—weighted by their probability of occurrence. More simply, for a given set of weights and biases, L_{gen} translates the expected loss of a trained network on a randomly chosen input-label pattern (with probability $p(x,y\left(x\right))$). The objective of training, is therefore to assemble a representative training set and find the weights and biases (ω^{*},b^{*}), that minimize equation 5.21. Using an optimization notation, this objective can be written as: $\omega^{*},b^{*}=\arg\min_{\omega,b}\,L_{\text{gen}}$.

Given the impracticality/impossibility of the aforementioned approach, and the fact that the testing can greatly vary depending on which of the available data were used to assemble the training set, a common practice to ensure generalization is to consider, during training, a separate set of (available) training data for validation purposes (control) only. This involves using an additional "unseen but known" dataset to gauge model performance via validation loss, after each iteration/epoch, tracking its evolution, and observe when this score decouples from the descending training loss score. In other words, the validation loss typically decreases during the initial phase of training (accompanying

the descent pattern of the training loss). However, when the ANN begins to overfit the training data, the validation loss, usually starts to increase, even though the training loss continues to decline. This is a clear indicator of over-training, which can severely hamper the model's generalization capacity [417]. By considering this complementary loss metric, it is possible to trigger "early stopping" and retrieve the model weights that avoided further degradation of the validation loss score."
The same is illustrated in Figure 5.14, with the dashed line representing this early stopping concept as the recommended training stopping criterion. Given that, before this point, one can assume that the model is still underfitting and after, it has an increased chance of overfitting.

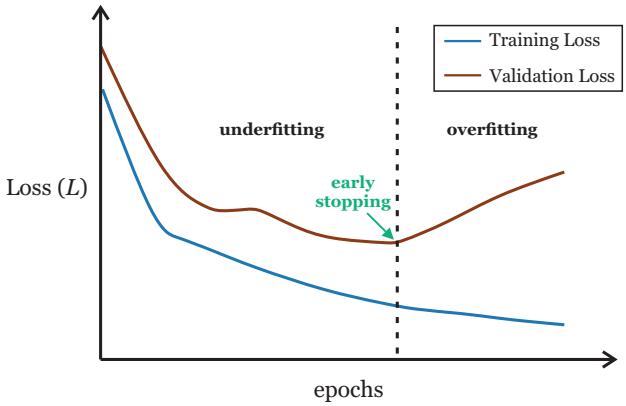


Figure 5.14: Training and validation convergence curves. The concept of early stopping for ANN training.

An important point to notice is that the increase in validation loss is only considered a good candidate for early stopping if that increase is sustained. That is, as shown in Figure 5.14, a sudden rise in validation loss should not be short-lived (i.e., after a few epochs, the score should continue the upward trend and not revert to following the downward trend of the training loss). To ensure this, a counter of the number of consecutive epochs/iterations in which the validation loss is increasing, often labeled as "patience parameter", "max fail" or simply "validation checks", is set. Only if this counter surpasses a user-defined threshold is the inflection point considered a worthy early stopping point [418]. The core idea is that the behavior observed in the validation set should to some extent, reflect the behavior expected in the testing dataset, thus serving as a proxy metric for testing performance. The same does not occur when the model learning relies on spurious correlations, i.e., patterns or relationships that are identified in the training data but are not actually meaningful (do not hold in testing), indicating poor generalization [419]. This reinforces why the size and representativeness of the validation set are crucial factors to consider.
A validation set that is too small may lead to premature early stopping (the validation check counter saturates fairly rapidly), and so leading to a situation of underfitting. An outcome that can be caused by the exact opposite scenario, that is, a validation set that is too large, leaving too few data for training, and so again, the resulting model may fail to capture the underlying patterns in the (totality) of the available training data (graphi-

cally the training loss would start to converge too rapidly, pointing to a large risk of overfitting/memorization instead of a balanced learning, and so ultimately compromise its validation/testing performance). Less frequently (a more hypothetical scenario), is when there are plenty of samples to divide between the training and validation set (with a close to 50%/50% split), a delayed early stopping can occur, and so the model may continue training longer than necessary. The point being, customarily the training set is much larger than the validation and test sets, but there is no universally optimal split between training, validation, and testing data, as it depends on the total amount of samples, the complexity of the relationships within the data, the expected problem/application loss tolerances, etc [420]. Generally, more complex patterns require more training data to capture them effectively. Commonly, most suggested splits range between 70-90% for training, while validation and testing comprises the remaining part, between 30-10% (usually equally split). Nevertheless, despite this broad consensus, some authors have even developed tailored heuristics to determine the best split to improve generalization. For instance, as noted in [421], the ratio of validation samples to the total available training samples should be a function of the total number of ANN weights (model parameters). This approach rightfully connects ANN sizing with the training and validation sets as very impactful hyperparameter choices for the overall ANN performance.

Regarding the question of how the available data is divided, there are several options beyond the contiguous split, where the first percentage of samples is assigned to training, the second to validation, and the third and last 'block' of data is used for testing. However, with this approach, especially when dealing with time-dependent datasets, certain or even a large part of the underlying patterns or seasonalities may only be contained in specific sets (most likely the training set), thus increasing the likelihood of having a high bias model or, on the contrary, a model with poor generalization. An intermediate option is to use an interleaved data division, where samples are retrieved following a specific pattern, distributing data into the three sets in a round-robin fashion. For instance, considering the following percentages for the split, 80%/10%/10%, this would mean that in a randomly assigned repeating pattern, 8 of the first 10 samples are attributed to the training set, while the remaining two are attributed to form the validation and testing sets. This procedure is then repeated up until the end of the dataset, ensuring that samples with the same intervals are always chosen for the same sets. A more common approach, which fully mitigates the aforementioned problem, is to fulfill the split percentages by randomly choosing the indexes of the labeled pairs of samples, i.e., randomly assigning samples to the training, validation, and test sets. This approach is analogous to shuffling the data before splitting and then employing a contiguous split. By doing so, we reduce the risk of overfitting to specific patterns present only in certain parts of the data, thereby leading to better generalization. Nevertheless, with a shuffled data division, the length of the training set, although always a factor (as is the case for most concepts in this chapter, the issues are deeply intertwined), gains new importance. For instance, if one of the network inputs is a time-series composed of hourly information, a random selection of indexes from a

limited pool of available samples can hinder training by distorting the representation of certain attributes or patterns (e.g., substantially fewer occurrences of a certain class). This could lead to sub-optimal learning and generalization, as the model might not adequately capture temporal dependencies or other critical features present in the entire available dataset.

As for the ”representativeness side” of the validation set, even with proper precautions regarding the formation of this set, it is important to recognize that the data in the validation set may differ significantly from the testing data due to various factors. For instance, a specific carbon-related tax or a new regulatory framework regarding the market bidding bounds represents a new reality that could permanently change the mean and standard deviation of the electricity price time-series data. This cannot be accounted for by historical data alone or by a proper data selection/division. This discrepancy can result in a model that performs well on the validation and training sets but poorly on the testing set. In broad terms, the limited nature of the validation set means that it tends to overestimate the test loss of a model trained on the entire dataset. Besides, if the testing set is available during training, its respective loss can also be monitored. If we notice a significant discrepancy between the iteration at which the testing loss reaches its minimum and the iteration where the same occurs for validation (the early-stopping point), it may strongly indicate a poor data division [422].

Despite all the advantages of using the early stopping mechanism as a solo stopping criterion, this discrepancy highlights one of its pitfalls. Overall, these pitfalls are deeply connected to the bias-variance trade-off principle, namely: i) a non-representative validation set; ii) a validation set that might contain several saddle points or local minima, thus prompting the training to stop prematurely; iii) overfitting to the validation set, particularly if the training set is not complementary and robust both in terms of length and representativeness; iv) an effectiveness that can be highly sensitive to the choice of hyperparameters, for example, the patience parameter itself.

The validation set is also essential for hyperparameter tuning, helping to define the settings that control the behavior of the model, such as the learning rate or network size. By experimenting with different hyperparameter values in an iterative process, training, and evaluating performance with the validation set, we can identify the combination of hyperparameters that yields the best loss values. Again, we hope that the same behavior remains consistent for the testing data [413]. Additionally, notice that in this tuning framework the testing set is only used to evaluate the performance of the end (tuned) model, hence why is also refereed to as hold-out data [420].

Beyond the early stopping mechanism supported by a static validation set, it is important to consider more advanced cross-validation techniques. The simplest form, known as holdout cross validation involves subdividing the training data once into training and validation sets. Generally, cross-validation can be described as a model evaluation technique that estimates generalization error based on re-sampling [417].

An improvement over the holdout method is the K-fold cross-validation technique, illus-

trated in Figure 5.15. Instead of a single partitioning, the labeled training data is partitioned into equal-sized K subsets (number of folds). In rotation, $K-1$ subsets form the training set while a single subset is used as the validation set. The training phase is repeated K times, with the training and validation losses for each fold being recorded and then averaged to produce a single estimate for a given combination of hyperparameters, $L_{\text{CV}} = \frac{1}{K}\sum_{i=1}^{K} L_i$. This process is then iteratively repeated to find the combination that yields the lowest average losses . By using the entire training set for both training and validation, we achieve a more comprehensive and unbiased model evaluation that is less sensitive to the specific partitioning of the data compared to the holdout method. Indeed, K-fold cross-validation reduces variance by averaging the model loss across multiple splits, providing a better estimate of the model's generalization error. As expected, the number of chosen folds, K, strikes different bias-variance trade-offs.

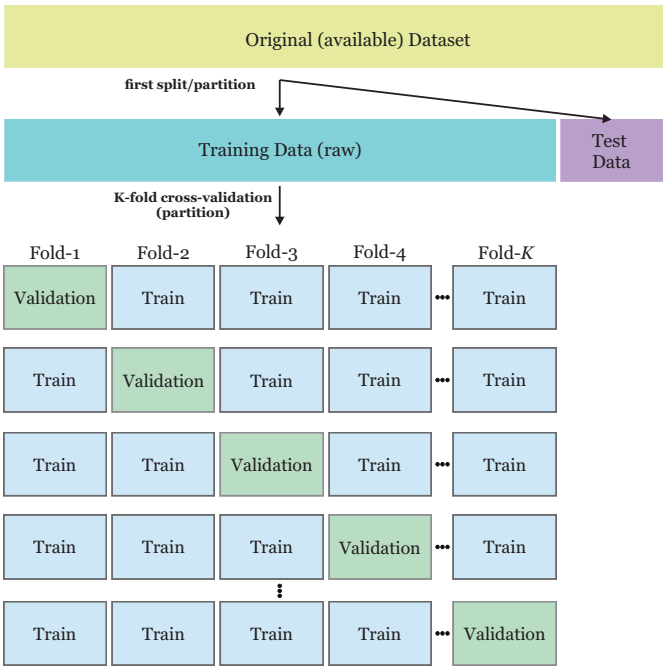


Figure 5.15: K-fold cross validation: data partition procedure.

With smaller values of K resulting in larger validation sets and consequently, smaller training sets, the model bias tends to be higher (due to the fewer samples for learning). However, the variance tends to be lower because the validation sets are statistically more stable across different folds and there are fewer rotations (loss values) to average. Conversely, for larger values of K, the trained models tend to present lower bias but higher variance. This is due to the larger training sets in contrast with the smaller validation sets, resulting in more loss values to average across the different folds. To achieve this delicate balance, the prevailing literature suggests using between 5 and 10 folds (values of K) [423]. Over the years, numerous extensions of the K-fold cross-validation technique have emerged, with particular attention given to the stratified version. This popular extension enhances the bias-variance trade-off through a data rearrangement process that ensures each fold is more representative of the overall dataset.

As we saw, overfitting often occurs in the presence of a limited training set but not exclusively. In fact, as we also covered in depth, a too stringent/one-sided stopping criterion often leads to too many training epochs, thus prompting the same issue. Similarly, too many (free) model parameters (weights) can also induce poor generalization [398].

This leads us right into the topic of ANN sizing and its direct connection with another sensible trade-off in ANN hyperparameter tuning. To put it clearly, while a deeper model with more neurons can strengthen its representational capacity, promote hierarchical feature learning, and substantially increase the network's ability to model complex non-linear relationships, simply adding more neurons without proper precautions, such as regularization, can be detrimental. It not only causes unnecessary computational constraints (increased computational cost, memory requirements, and ultimately, a sizable training time) but, more importantly, can be a concrete inducer of overfitting.

In greater detail, while it is true that an enlarged ANN feature space (manifold created by the input features and the transformations applied by the successive network's layers) expands the number of possible activation combinations, which in turn promotes more flexible decision boundaries, it is also true that the excessive number of model parameters can lead the ANN to poor generalization. This occurs by establishing overly intricate decision boundaries that do not reflect the true training data distribution, memorizing noise, and taking hierarchical feature learning too far, thus capturing irrelevant patterns rather than the underlying trends.

To avoid this problem, the number of samples in the training set (dimensionality of the input space) should greatly exceed the number of model parameters (length of the ANN weights vector) [416, 424]. To put it another way, a larger training set (more patterns) allows for a greater number of model parameters to be used without compromising generalization [425].

Recognizing that this is still a very vague rule of thumb, several studies have been conducted over the years to establish a clearer cut-off threshold to avoid overfitting based on either the batch size or the number of inputs being fed to the network. Despite the case-by-case nature of the problem, some well-known guidelines include: i) the ratio of the number of training samples to the number of network weights should be larger than 30 (it can be seen more as a lower bound) [421]; ii) the number of hidden nodes specifically should be $2/3$ the size of the input layer (number of network inputs, n), plus the size of the output layer (the number of outputs the network produces) [426]; iii) the number of hidden nodes should be between the size of the input layer and the size of the output layer; iv) the number of hidden nodes should be less than two times the size of the input layer; v) or the number of hidden nodes can be determined from the many proposed heuristics through the years, for instance $\left(4n^2+3\right) / \left(n^2-8\right)$ as given in [427]; vi) as mentioned earlier when discussing the cross-validation mechanism and its use for hyperparameter tuning, a more tailored/advanced approach to the network sizing is to use it together with a network constructing or pruning technique (as explained later in this subsection). The point of all of these guidelines is that a small enough network does not have enough rep-

resentational capacity (degrees of freedom), provided by the weights in the feature space, to overfit the data [422]. Additionally, research indicates that for deep architectures, poor local minima are rarely an issue. Regardless of the initial network weights, the solutions tend to converge [387].

To wrap up this subject, while a larger ANN offers undeniable leverage in terms of the model's representation capacity, and consequently, more approximation accuracy for regression tasks, it is crucial to avoid the real risk of overfitting from an over-parameterized model. To help with this task and avoid excessively tight constraints in terms of model sizing (i.e., a fine and time-consuming grid-search for ANN sizing-related hyperparameter tuning), there are a multitude of overfitting mitigation options that experts can employ, thus reducing the strain put on the sizing issue, as shown in Figure 5.16.

Figure 5.16: Overfitting mitigation/prevention options in ANN context.

From the listed techniques/procedures in this radial figure, we can see that besides the already explained cross-validation and early stopping (which go hand-in-hand), there is also a mention of data augmentation as a generalization improvement procedure, as hinted at when discussing an adequate ANN sizing approach. Data augmentation works on the notion that expanding the training data increases the chances of enhancing its diversity and meaningfulness (more patterns for the model to capture).

Moving on to pruning and dropout, which are essential network reduction techniques. Pruning involves the removal of network elements, such as neurons or layers, that are considered redundant based on their contribution to overall network performance. For example, weights with low magnitudes and/or minimal contribution to training loss (or other importance scores) may be pruned. Generally, biases are left "untouched" given that it is more difficult to assess their contribution, and the downside of eliminating them is much higher (as mentioned in subsection 5.3.1, these parameters can effectively shift the activa-

tion value). This technique can operate either in an online fashion, known as train-time pruning, or offline, known as post-training pruning [416, 428, 429]. Conversely, dropout is an online technique occurring during the training process. It randomly selects neurons to drop (usually by applying a binary mask to the neuron outputs, turning off certain neurons randomly) during each training iteration (training using a thinned ANN), based on a hyperparameter known as the dropout rate. This process helps to prevent co-adaptation, a phenomenon where neurons become overly reliant on specific other neurons, leading to poor generalization. Dropout forces neurons to learn independently (breaking these co-dependencies) and share the workload, thereby improving robustness and enabling the network to generalize by learning more general patterns rather than dataset-specific noise [413, 430].

The next listed technique to mitigate overfitting, **regularization**, encompasses a variety of methods. As the name suggests, the goal is to "regularize" the model's behavior by promoting consistency and discouraging overly complex models, which can lead to overfitting. This is typically achieved by constraining or penalizing the model's complexity. To implement regularization, the network's training loss function, usually the MSE value (equation 5.7), is modified to include a penalty term related to the model parameters. One common approach is L1 regularization, derived from the Least Absolute Shrinkage and Selection Operator, also known as Lasso regression. It uses the L1 norm (also referred to as taxi-cab or Manhattan distance), which is the sum of the absolute values of the model parameters [413], $\|W\|_1$. The modified and now compounded loss function then takes the following form:

$$L(\omega, b) = \mathrm{MSE}_{\mathrm{reg,L1}} = \underbrace{\sum_{i=1}^{N} \frac{\left(y_i(x_i) - \hat{y}_i(x_i, \omega, b)\right)^2}{N}}_{\mathrm{MSE}} + \gamma \underbrace{\sum\nolimits_j |\omega_j| + \sum_h |b_h|}_{\text{L1 Penalty}} \quad (5.22)$$

note that this modified MSE includes an L1 penalty term multiplied by a regularization parameter, γ. Its addition to the loss function promotes the shrinkage of the coefficients during training, with the goal of nullifying less important model features. In other words, it effectively targets a more sparse, smoother, and ultimately simpler network. Other instances in the literature show the classic MSE term being multiplied by a factor of $(1-\gamma)$, forming a convex combination. Here, γ is designated as the performance ratio, indicating the strength of the regularization. This parameter controls the balance between fitting the training data and applying the penalty term [422]. Consequently, γ is another hyperparameter that is difficult to determine and affects the bias-variance trade-off.

Another penalty option is L2 regularization, also known as "weight decay," which emerges from Ridge regression. As the name suggests, it uses a second-order (Euclidean) norm of all the network parameters. In the context of regularization, this norm is squared for simplification purposes, $\|W\|_2^2$ (squared magnitude of the weights vector). With it, $L(\omega, b)$ (equation 5.22) takes the following form instead:

$$L(\omega, b) = \text{MSE}_{\text{reg,L2}} = \underbrace{\sum_{i=1}^{N} \frac{\left(y_i(x_i) - \hat{y}_i(x_i, \omega, b)\right)^2}{N}}_{\text{MSE}} + \gamma \underbrace{\sum\nolimits_j \omega_j^2 + \sum_h b_h^2}_{\text{L2 Penalty}} \quad (5.23)$$

with the penalty term now being proportional to the sum of the squares of the model parameters. The main difference is that instead of favoring the rejection of some parameters, this approach leads the training process to favor small weights overall [430]. Larger weights are heavily penalized due to the square effect (and so, they are only kept if they are deemed valuable parameters that truly reflect ”learning” and thus reduce the MSE term), while weights with absolute values below 1 are minimized by the expression (by the same token). Immediately, one can notice the possible incompatibility of a too lenient approach on this (meaning, considering non-small γ values) and the documented problem of the vanishing gradient, which exactly documents the learning adjustment difficulty in denser networks when the gradient is propagated backwards. One can say that the issue would be exacerbated if all the network weights are indeed too close to zero, as targeted by this approach. Thus, this highlights yet again the need for an ”experienced” and tailored approach to each problem to thoroughly consider all these interconnected aspects.

Notwithstanding the focus being put on these classic options, there are other regularization penalties. For instance, elastic net regularization considers a convex combination of both L1 and L2 penalty terms. It is also worth pointing out that, despite not being penalties or constraints added to the loss function, dropout and even pruning techniques, act on the model parameters in a somewhat similar fashion. Temporarily or not, they try to remove redundant parameters/features, and so they are often considered regularization techniques as well, not just in the traditional sense.

Continuing to move on the list of techniques, we see a reference to ensemble methods. This basically implies the combining of multiple models with different characteristics (different bias-variance trade-offs) to achieve a more ”representative” model that is capable of generalization. For instance, an ensemble method can be simply an aggregator of the output predictions of different-sized or even different types of network architectures.

Batch normalization is one of the many less popular regularization techniques, which is why it was categorized under ”others” in our radial diagram in Figure 5.16. As the name suggests, it ”normalizes” the outputs of a given input layer for each mini-batch, ensuring that the nodal activations maintain a stable distribution throughout network training, thereby, according to initial findings, reducing internal covariate shift. It works by adjusting and scaling the activations to have a mean of zero and a variance of one, and then applying learnable scaling and shifting parameters within each mini-batch (i.e., the training set considered in that specific iteration). This helps stabilize and accelerate the training process. Recent findings, however, point instead to a smoother optimization landscape as the main contribution of this technique, rather than the distributional stability of layer inputs as previously thought. In greater detail, its implementation often leads to more

predictive and well-behaved gradients, which in turn make the training process more stable. This smoothing effect is not unique to batch normalization, but in conjunction with other techniques, it can help prevent gradient explosion problems and improve the network robustness to hyperparameter sensitivity [431].

To some degree, batch normalization is related to the pre-processing stage (before it is presented to the neural network), given that the network inputs are normalized within a certain range or standardized to have a fixed mean of zero and a unitary standard deviation. This stage, which precedes training, is followed by a complementary post-processing stage after training is concluded. The post-processing stage reverse transforms the outputs to return them to the original target range. The key difference is that they are used at different stages of the model training process and serve slightly different purposes, with batch normalization being an in-training counterpart. Without pre-processing normalization, the input features may have vastly different scales (magnitudes). For instance, imagine an ANN receiving predicted wind speeds, air densities, and past wind generation outputs as inputs to estimate the future wind generator output (W). Without any kind of pre-processing of the input data, there is a real chance of the model becoming biased towards features with larger scales (in this case, the past power outputs given in W), neglecting those with smaller scales (e.g., air density fluctuating slightly around the value of $1.225\,\mathrm{kg/m^3}$). Another issue is the exploding gradient problem, which may arise in the presence of very high input values, given that the corresponding activations and gradients can also become very large. As these large values are propagated through the network, they can grow exponentially, causing the gradients to ”explode”.

Typically, `mapminmax` type pre-processing and post-processing functions are used to normalize the data by linearly transforming it to ensure all features have the same range, typically $[0, 1]$ or $[-1, 1]$. Besides the already presented reasons, this is also beneficial for matching the nodal activation function ranges (function image). Other pre-processing instances include dimensionality reduction techniques, such as the principal component analysis (PCA), processing unkwown inputs and ”don’t care” targets, feature engineering and extraction, sampling data, data augmentation and outlier removal, among many others.

5.3.3 Architectures

ANNs encompass a very broad class of different architectures. Informed by the information-processing mechanisms that take place in the biological nervous system, ANNs are black-box type, nonlinear processing architectures, where the artificial neurons (nodes) are the key element. Its scalable nature, coupled with the multiplicity of connections, i.e., paths (graphs) between neurons, the chosen activation function, and the way the connection strength (weights) between nodes is updated (training process), results in a very powerful and adaptable ML tool to perform ordinary regression and classification tasks [432]. In this fashion, the presented subsections compile, and list, some of the trending ANN architectures, providing a concise theoretical background (inspiration/main processing

mechanism), which is extended in greater detail for architectures that will be considered in later chapters for real problem implementations/applications.

5.3.3.1 Feedforward type architectures

As the name suggests, this type of architectures contemplate an exclusively forward flow of information between the constituting layers/neurons/nodes, i.e., without any feedback connections. Deriving from early breakthroughs in the field of NN, such as the perceptron or adaline networks, this is the oldest type of architecture. Nonetheless, they are still very capable computational tools to perform regression tasks and tend to be easier to fine-tune in comparison with more recent and complex architectures. Over the years, some tweaks in the architecture, connections, activation functions, and training methods were developed to refine its accuracy. In accordance, the currently most relevant feedforward type architectures are presented in Figure 5.17, and a concise overlook of the method's key theory/implementation details is presented in the subsections.

Deep Feedforward Neural Networks (DFNNs) are perhaps the most ordinary type of DNN architecture, which in the same fashion as its shallow equivalent feeds the information forwardly between the layers (from the input to the output layer). As we saw already, the "deep" prefix accounts for a sizeable increase in complexity, i.e., in the number of hidden layers/neurons, and ultimately in the weights and bias terms. As a member of the broader class of ML methods, these networks achieve representation learning, given their more abstract and high-dimensional features mapping from the provided input/target training pairs, embodied by the assigned weights in each connected node. This results in an enhanced ability to handle and extract non-linearities from the data.

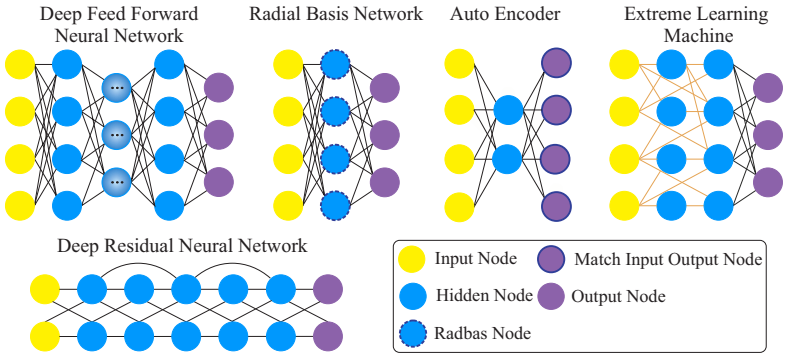


Figure 5.17: Commonly used Feed Forward type architectures: Generic overview of connections and nodes.

Radial Basis Function Neural Networks (RBFNNs) were inspired by the radial basis functions robust interpolation capabilities and their multilayered architecture belongs to the class of feed forward neural networks [432]. These RBFs rely on the distance from a center vector (centroids) that are radially symmetric to perform the fit. As such, the centroids and widths of the radbas function in each hidden node

are the major control parameters that allow the fitting to the data. Besides the conventional input and output layers, RBFNNs have a single hidden layer composed of several hidden nodes, as illustrated in Figure 5.17, known as radbas neurons that are activated using a Gaussian activation, $f\,(p) = e^{-p^2}$, which has non-negative response for all the activation inputs p.

With p denoting the Euclidian distance between the weight vector and the different inputs x (input layer), thus contrasting with the way the activation cost is computed in a DFNN. In terms of regression capabilities it is vital to calibrate the radial basis functions parameters. Avoiding too sparse or too overlapping functions, which lead to a loss of performance.

Auto Enconder (AE) as another FFNN multilayered type architecture, AE shares some resemblance with Principal Components Analysis and is trained in an unsupervised manner (target values set equal to the inputs). A stacked AE is used to transform/reconstruct the data, by employing single or multiple hidden layers, operating in two sequential stages. The encoding stage consists in mapping the low-dimensional representation space, i.e., both linear and nonlinear transformation and extracting relevant features. The decoding stage reverses this process, and it returns an output in the original space [433]. So, both stages encompass the learning process, where the aim is to minimize the reconstruction error (loss function) between the input and output information. Additionally, the input and output layers of this architecture share the same number of nodes.

Several AE variants exist, namely the sparse, denoising and undercomplete [434]. Moreover, as seen before for other architectures, the deep AE provides a better framework to handle complex, data reconstruction, dimensionality reduction, image denoising and compression, feature learning, etc.

Deep Residual Neural Networks (ResNets) are a very particular type of FFNN, closely related to CNNs, that mimic the behavior in the cerebral cortex, where some forward connections are skipped over some layers. As such, ResNets consist of a series of residual layers [435], rather than the traditional plain ones. Bypass connections are established between non-consecutive layers, usually with a jump of two or three layers (double or triple layer skips), as shown in Figure 5.17. This implies that the (ResNet) output response of an arbitrary i^{th} layer, Y^i, which is not only connected to the $i^{\text{th}}-1$ layer, but also to a $i^{\text{th}}-m$ layer through a bypass connection, can be generally written as: $Y^i = f^i\left(a^i + Y^{i-m}\right) = f^i\left(W^{i-1,i} \times Y^{i-1} + B^i + Y^{i-m}\right)$. Where f^i is the activation function in layer i; a^i denotes the i^{th} biased weighted sum vector; $W^{i-1,i}$ is the weights matrix for the plain neuron connections between layers $i-1$ and i; while the vector of biases in the i^{th} layer is given by B^i.

With these skip connections, the ResNet is able to mitigate the (main) problems of degradation and vanishing gradients that arise with increasingly deeper plain architectures [436]. The first problem concerns the fact that after a certain degree of

architecture complexity, accuracy saturates and the training error starts to increase, while the second reflects the tendency for a closer to zero gradient, as the error is backpropagated trough the layers, i.e., as we move further and further from the output layer. So, it is easy to understand how ResNets lessen the vanishing gradient problem, since the feedback error now "travels" from previous layers via shortcut connections rather than consecutively through all the layers. In addition, it is also a common feature to employ batch normalization layers and the rectified linear unit layer (ReLU) as the chosen (non-linear) activation function, which further help mitigate the mentioned problems.

Extreme Learning Machine (ELM) unlike the previously listed methods, ELM is not a different ANN architecture, instead it is a different learning algorithm (supervised), employed with shallow FFNN architectures, i.e., typically a single hidden layer is sufficient. The main concept is to transform the neural network learning into a least squares problem, which can be solved easier with less computation effort. The same is achieved by randomly initializing the hidden layer weights and biases, which are then kept constant, contrary to what occurs in classical backpropagation learning algorithms, where these weights are updated iteratively. That is why in Figure 5.17 these invariable hidden connections are displayed in a different color (orange) versus the adjustable (learnable) output connections (in black).

It is then up to the output weights, β (vector form), the burden of minimizing the training error. So, using the same DFNNs forward processing notation, with an output node without a bias term, for a training set of input/desired pairs, (x_i, d_i), with a total of N_{train} samples, the objective is to determine each output weight, β_j, such as: $y_i{=}\sum_{i=1}^{n} \beta_j \left(\omega_{ij} x_i{+}b_j\right)$, where $y_i{=}d_i$. This formulation can be simplified into a matrix form as: $H\beta{=}D$, where H denotes the hidden layer output matrix and D represents the set of desired outputs. Typically, to solve this linear system a least-squares method is employed [437].

5.3.3.2 Recurrent type architectures (RNNs)

RNNs is the other leading class of deep learning architectures designed to handle sequence data. As the name suggests, this class of networks establishes not only feed forward connections, but also loop connection between nodes, forming a recurrent graph along a temporal sequence. The RNN "state/memory" is a key feature that differentiates them from feedforward neural networks, and it implies a dependency on previous states to determine the outputs of the current inputs. Therefore, considering a simple RNN architecture, the output of the recurrent nodes for a certain time-step t, y_t, is not only a function of the input vector, x_t, but also of its own state, $h_t = \sigma_h \left(U_h h_{t-1} + W_h x_t + b_h\right)$, i.e., $y_t = \sigma_y \left(W_y h_t + b_y\right)$. Where σ_h and σ_y are the activation functions (typically non-linear) and U_h, W_h, b_h, W_y and b_y are the respective weights and biases matrices. The dynamic temporal behavior means that the same set of weights are applied recursively over a differentiable (unfolded) graph

architecture for the selected input sequence [438]. An illustration of this type of architecture is provided in Figure 5.18, where it is evident the existence of a recurrency (loop) in the nodes, in clear contrast with the architectures seen in Figure 5.17.
Additionally, RNNs can also be Bidirectional, i.e., the unfolded graph structure can be processed in the opposite direction, thus, not only using information from previous input data to update the current state, but also from later data. These traits make this class of networks a preferred choice when it comes to language translation, speech recognition, language processing and image captioning problems. Elman, Jordan, and Hopfield networks were among the first and simplest types of RNNs. Though, given its training accuracy limitations, most of these architectures are no longer a trend in the forecasting field. In opposition, the most commonly employed RNN type architectures are presented below:

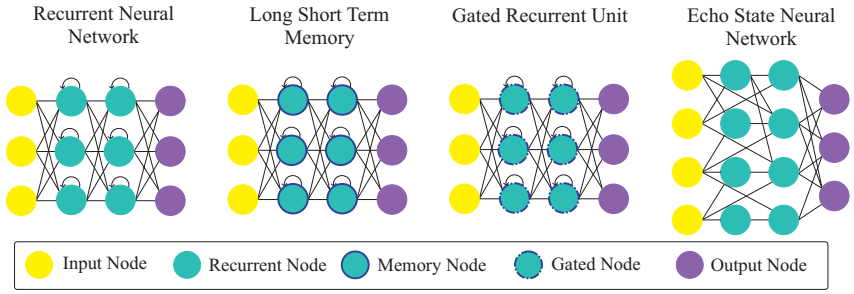


Figure 5.18: Commonly used recurrent type architectures: Generic overview of connections and nodes.

Long Short Term Memory (LSTMs) are a specific type of RNN, which have arisen to tackle a sensitive problem of the classic vanilla RNNs. This is especially true when dealing with long-term dependencies, where the network gradient-based training process would ultimately lead to a situation of vanishing or exploding gradient.

LSTM networks overcome this shortcoming with its interconnected three multiplicative gated cells in the hidden layers of the network, i.e., an input gate, a forget gate and output gates, thus expanding the memory access to both short and long-term training dependencies. This is particularly due to the distinctive connection of the forget cell with the previous internal state [439]. An LSTM unit represents the interconnected structure of the different cells with the current input sequence and the previous state information, which includes a hidden state (short-term memory related) and a cell state (long-term memory related). Typically, an LSTM layer is comprised of several those units and this hidden layer is followed by a standard feedforward output layer.

Gated Recurrent Units (GRUs) are also a very robust type of RNNs and, just like LSTMs, they base their functioning in gating mechanisms to control the flow of information between cells [440]. In other words, GRUs gated cells enable it to adaptively capture time-dependencies from sizeable sequences of input data, without the need to dis-

card early information. Thus, addressing the short-term memory problem of vanilla RNNs.

Yet, unlike LSTMs, GRUs do not have a cell state, only the hidden state. Also, the three gated cells in the LSTM unit give place to only two in the GRU unit, i.e., a reset gate and an update gate, which likewise in the LSTM will be responsible for the short-to-long memory retention. This simpler architecture, with less elements and consequently multiplicative interconnections, implies that GRU networks are usually quicker to train.

Echo State Neural Network (ESNN): Drawing inspiration from reservoir computing, and following the conclusion that in traditional RNN, the major changes occur in the output weights. ESNN can be considered a type of RNN with supervised learning, where the connections in the hidden layer are very sparse, fixed and randomly assigned. This working principle is shared by Liquid State Machines [441].

The input signals are fed to the hidden layer, which in turn trigger a nonlinear neuron response within reservoir network, which implies a sharing of the hidden neuron's states with its neighbors (the internal layer is assigned with random weights). These responses (contribution of each neuron to the output) are then fed to a trainable output layer, where only the weights of output neurons need to be learned, which through a weighted linear combination produces the ESNN output. So, the training process only consists in modifying the synaptic weights that connect the hidden neurons to output neurons, rather than the entire network weights. This fact means that ESNNs are easier to train and very flexible [442], in a clear opposition to traditional RNN, which tend to be impractical to use, given the slower nature of the training process, uncertainty regarding convergence and its susceptibility to suffer from vanishing or exploding gradients, when performing the backpropagation through time.

5.3.3.3 Convolutional type architectures (CNNs)

CNNs have been inspired by the connectivity patterns in the brain's visual cortex; this renowned class of DL architecture is primarily designed to handle image related data [443]. Once, the use of old-fashioned FFNN or RNN is unattainable in the presence of 3D-input datasets (width $\times$ height $\times$ color channels), as is the case of RGB images, due to necessary vast number of connections that are established between each individual input feature and a correspondent individual node in the input layer.
As the name itself indicates, these multilayered networks, as depicted in Figure 5.19, employ the convolution operation rather than the multiplication in the key hidden layer, called convolution layer. This type of hidden (middle) layer employs a set of learnable filters via convolution, with a local connectivity in terms of spatial components, but full along the depth (channels) dimension, to the inbound input volume, which results in different activation (feature) maps. In addition, pooling layers, also known as subsam-

pling layers, are another distinctive element of a CNN, which reduces the layer's input volume (activation maps) by downsampling the spatial dimensions. This results in a reduced number of learning parameters, which ultimately lowers the computation burden when training the network. Max and Average pooling are the common ways to perform this spatial dimensionality reduction.

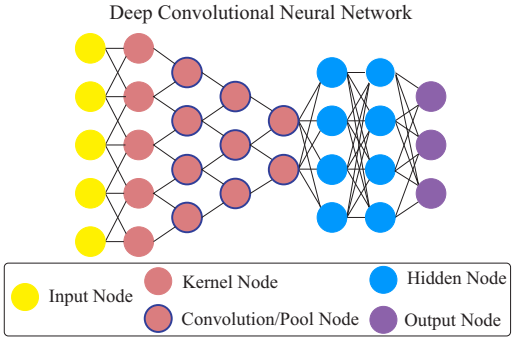


Figure 5.19: CNN architecture: Generic overview of connections and nodes.

By these means, the resulting output volume, when using these two types of hidden layers, will depend on the input volume size and then on the user-defined (layer) hyperparameters: the filter size, the stride (the way the filter shifts) and the padding (spatially added cushion). Besides, normalization layers and fully connected layers are other commonly used layers in CNNs. The application of this type of network has been rapidly growing beyond the traditional fields of image recognition, object detection, text, and video classification segmentation, etc.

5.4 Other ML models/architectures

Besides the conventional FNNs, RNNs and CNNs classes of architectures, there are a few other noteworthy ML methods, both supervised and unsupervised, that are widely used in regression problems, worth mentioning.

5.4.1 Deep Belief Neural Network

A deep belief neural network (DBNN) is a well-regarded architecture in the field of DL, consisting in a stack of restricted Boltzmann machines (RBM). In turn, these machines are a generative stochastic type of unsupervised DL model, designed to learn a probability distribution over an input dataset, where a full unconstrained (bidirectional) connectivity is established between the nodes, which then make binary decisions. A RBM layer is a generative stochastic ANN that can learn a probability distribution from its input datasets, and provide a feed forward flow of information, as such DBNNs are made from several of these singular layers as seen in Figure 5.20. The DBNN training process consists of two main stages, firstly a layer-by-layer unsupervised training of the RBM layers (Greedy learning algorithm), where the top-down degenerative weights are learned.

Secondly, through the use of error back-propagation (supervised) algorithms to refine the DBNN parameters [444].

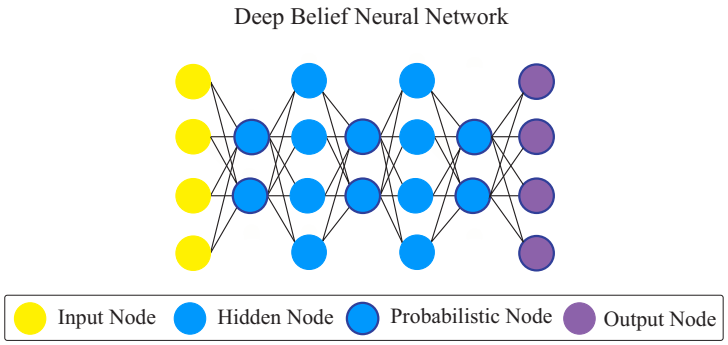


Figure 5.20: DBNN architecture: Generic overview of connections and nodes.

Its ”deep traits” makes them a tailored choice to handle deep patterns, and enables it to capture the difference between normal data and faulty data [445]. DBNNs have revealed a higher accuracy versus other regarded DL methods in the tasks of speech recognition, image recognition, and classification/regression problems.

5.4.2 Adaptive Neuro-Fuzzy Inference System

Taking advantage of both ANN non-linear learning capabilities and fuzzy systems properties, particularly the Takagi–Sugeno fuzzy inference system, Adaptive Neuro-Fuzzy Inference System (ANFIS) is a noteworthy adaptive neuro-fuzzy model, where the network weights behave similarly to the fuzzy system parameters. These weights are updated according to a hybrid learning rule, combining gradient descent backpropagation and a least-squares algorithm. In terms of architecture, represented in Figure 5.21, ANFIS consists of five different layers: i) a ”fuzzification” layer constituted by adaptative nodes, dealing with a fuzzy set of values/descriptors, using common membership functions with different membership degrees (defined by the premise parameters), such as a Gaussian or generalized bell curve. These functions are used to activate the node in a different fashion to the common sigmoid transfer function in vanilla ANNs; ii) a product/rule layer, where the output reveals the rules’ firing strength; iii) a normalized layer, which normalizes the (input) firing strength of each rule; iv) a ”defuzzification” layer, also composed by adaptive nodes, responsible for evaluating the contribution of each normalized firing strength to the overall output, and accordingly returning the fuzzy inference results into crisp outputs (based on the premise parameters); v) a single node output layer, which computes the sum of all the rules’ outputs [446].

5.4.3 Support Vector Machine

A support vector machine (SVM) is a well-regarded supervised ML method that tries to find the optimal hyperplane that constitutes a decision boundary, ideally one that achieves the largest margin between two classes, i.e., the hyperplane that maximizes the margin

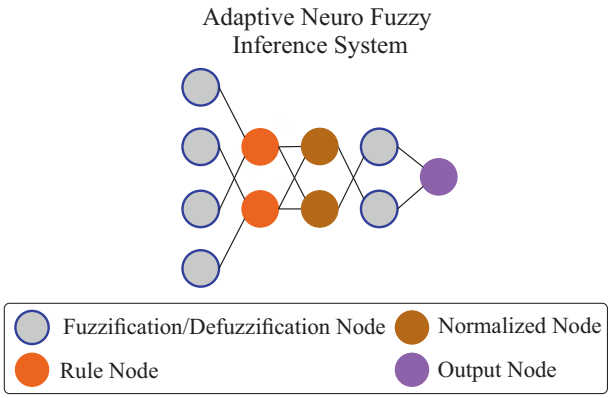


Figure 5.21: ANFIS architecture: Generic overview of connections and nodes.

to the nearest classifier. These hyperplanes are built by using support vectors, i.e., input data points that are closest to the decision boundary (hard to classify). Therefore, influencing the position and orientation of the hyperplane. For instance, a simple Linear SVM for linearly separable patterns defines a hyperplane that satisfies the condition: $w^{\mathrm{T}}x - b_{\text{offset}} = 0$, where x denotes the set of input training data, w represents the normal vector and b_{offset} denotes an offset. The SVM is able to non-linearly map the training data in a low-dimensional plane to high-dimensional space. The kernel function extends the high dimensional mapping of the original input data to non-linear separable patterns. These attributes make the SVM an effective method for regression, classification, and mining tasks. This justifies its application to forecasting problems, since it avoids the weak classification ability and over-fitting shortcomings of traditional prediction models [439].

5.4.4 Random Forest

A Random Forest (RF) can be categorized as an ensemble (aggregation) ML method [446], which is broadly applied to solve regression and classification problems. Its non-linear modeling functioning is centered around the use of numerous decision trees, as highlighted in Figure 5.22, that work as classifiers, hence, the designation of "forest". A decision tree is formed through a recursive partitioning of the predictor variables space, with the attribution of its respective response variable value (within each partition). The RF method differentiates itself from the decision tree algorithm, by randomly establishing the root and segregating nodes. The ensemble strategies of bagging (bootstrap aggregating) and boosting are used to train the RF. This method can predict the response variable values given the predictor variables values, by calculating the mean output of the several various trees. An increase in the number of trees is directly connected with the precision of the outcome. In addition, RF alleviates the overfitting problem present in decision trees. This method can predict the response variable values given the predictor variables values, by calculating the mean output of the several various trees. An increase in the number of trees is directly connected with the precision of the outcome. In addition, RF alleviates the overfitting problem present in decision trees.

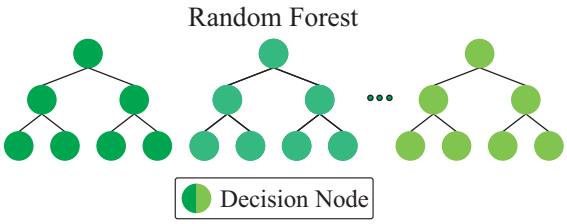


Figure 5.22: RF architecture: Generic overview of connections and nodes.

5.5 A qualitative assessment of ML models

The previous sections have highlighted ML methods are the leading trend in solving EPS problems, particularly in forecasting ones (the focus of later chapters in this thesis). Still, given the wide variety of methods and the intrinsic differences between them, it is very difficult to compare the model's attributes and draw meaningful inferences from such a wide range of literature, when so many factors can determine sizable performance swings. Nevertheless, it is still possible to underline some generic characteristics of the different methods that can help identify some of the strengths and weaknesses. Classic ANNs can be hard to fine-tune given the amount of model (both architecture and train related) hyperparameters and can also be trapped in local optimum solutions. On the other hand, DNNs have a higher representative learning potential, thus improving the abstract learning and feature extraction capabilities. Though, it is also sensitive in terms of hyperparameters configuration and together with ANN it can be slow in terms of training speed [447]. In turn, RBFNNs with its simpler learning algorithm, avoids the costly error backpropagation process; however, they tend to show a lower accuracy in comparison with FFNN, especially in the presence of complex explanatory variables, given the difficulty to adjust the radial basis function parameters so that the linear combination produces a generalized output, i.e., with no overfitting [448]. ELM is a supervised learning algorithm, which is commonly employed with a single hidden layer FFNN. This simpler structure, with less hyperparameters, where the hidden weights are randomly assigned and kept constant, and where the output weights are updated by using a least square method rather than through a lengthy backpropagation process, means that they are faster to train and show sufficient generalization capacity. However, this comes with a caveat, since the random assignment of weight can result in a poor stability and in the presence of ill-conditioned data (significant number of outliers), and thus, robustness and generalization capacity are greatly diminished [449]. AEs is another FF type of NN, tailored for feature extraction, and its speed and compactness are among the advantages. In terms of disadvantages, AEs can suffer from a narrow bottleneck layer, problems in decoding, overfitting, and convergence speed when many hidden nodes are added (to improve the training accuracy). Deep residual networks are known for its generalization capabilities but are also known to suffer from vanishing/exploding gradient problems [450].

Moving on to RNNs, the emergence of gated cells and auto differentiation techniques have successfully tackled convergence and training problems of conventional architectures. In

this regard, LSTM and GRU networks are known for its robustness and ability to avoid local optimum problems, making them a suitable tool for time-series prediction, with relatively long-time dependences [439,451]. This means that they usually tend to outperform other architectures, including ESNN, which often require more computational time and training samples to achieve a similar accuracy. Still, their structure centers around sequential mechanisms rather than parallel computing, which can result in a considerable computational cost (time-consuming); in addition, when dealing with long sequences, the limited memory capacity means that some information can be lost during the training process [452]. Despite being designed for image related problems, CNNs, including its 1D models, have shown promising results in sequence processing problems [451], including in forecasting [453], given its ability to automatically assimilate features from the input patterns [454]. Still, most of the time they need to be used in conjunction with LSTMs (more able to handle sequential data) or dense layers in order to achieve competitive results.

Concerning the other noteworthy ML methods, RFs are known to be fast, with few hyperparameters, but less accurate, black-box type methods, particularly when dealing with noisy regression problems [439]. In opposition, its deep counterpart (deep forest regression), notwithstanding the greater computational cost, has been shown to outperform well established methods in the forecasting problems, while avoiding the hyperparameter depedence of DNNs. On the other hand, SVMs are more robust by nature, but can be time-consuming, and prone to a large error when a clear margin of separation between classes is not present, i.e., noisy data. Moreover, it is not suitable for large datasets. Similarly, to FFNNs, ANFIS architectures with its knowledge representation achieve good nonlinear mapping of the input data, thus, typically leading to a smaller error for a comparable complexity. Yet, the convergence tends to be slower (more time-consuming) and it can be hard to adjust the membership functions. Lastly, Deep belief networks reveal a good data prediction ability, but they can be hard to fine-tune its parameters [455]. Nevertheless, more than a qualitative assessment, which is good for a first screening the most suitable architectures/methods, the real ”tie-breaker” relies on quantitative assessment using real case scenarios.

Chapter 6

Power Flow Studies

The integration of vast amounts of renewables into the power generation mix is paramount for the green transition. These intermittent resources bring promise of a cleaner and less expensive generation portfolio. Yet, this comes with a cost, not only in terms of the required investments but also due to the planning and operational complexity that they bring. With this regard, optimal power flow (OPF) studies have been a key tool to efficiently solve the steady-state operation of AC interconnected power systems. An optimization task that was traditionally addressed using classical numerical methods. However, in recent years, metaheuristic approaches have gained prominence due to their ability to effectively handle high-dimensional, nonlinear, and multimodal problems, as discussed in Chapter 4. As such, and before proceeding with some meaningful tests, this chapter revisits its theoretical foundations, with a uniform take on the circuit modeling of some of the common system elements, which then establishes a more grounded and uniform link with the optimal power flow theoretical basis. Namely in terms of the admittance matrix considerations, a full derivation both in terms of polar and rectangular forms, of the active and reactive power flows, associated constraints, penalty method and its solution by means of iterative methods. Rooted in a solid theoretical foundation, this chapter concludes by proposing a modified Harris Hawks Optimization (HHO) algorithm to solve the classic IEEE-30 bus system optimal power flow (OPF) constrained optimization problem, achieving operational costs that align closely with those reported in the existing literature.

6.1 Introduction

In an ever-changing environment where the power systems' load and generation backbone are constantly being modernized to meet new demands and accomplish the "green" transformation [456]. The expected large-scale integration of renewables poses widespread challenges in terms of network and resource adequacy, voltage and frequency stability, energy quality, reverse power flow, among others [457]. Which in turn, will require a new approach to forecasting, scheduling, dispatch, balancing and operational control tasks [458].

Originally formulated by Carpentier in 1962 [459], OPF or load flow studies intend to find the optimal settings in terms of a set of control variables, to meet the problem constraints, while optimizing certain operational aspects, as is the case of the costs for a single time-period. OPF problem can be described as highly constrained, nonlinear, and non-convex optimization problem, where the nodal power balance equations and the constraints of state and control variables constitute the equality and inequality problem constraints, re-

spectively. The first set of variables includes the generators' reactive power, load bus voltages and branch flows, while the second set includes the generator's reactive power and its voltage magnitude, transformer tap ratios and the reactive power contribution from system VAR sources [460].

A notable and closely related optimization task regarding power systems operation is the Economic Dispatch (ED) problem. With its origin dating back to the 1920s, when the operator faced the decision of scheduling the power output of different generation units to meet a certain load demand while minimizing the operational costs [461]. As such, it is important to emphasize right from the get go, that this optimization problem is actually a predecessor of the OPF formulation, which in plain terms involves fulfilling the total load with a cost-effective and/or emission-efficient generation allocation for a snapshot in time [462]. This, while adhering to generator constraints, but disregarding transmission constraints [463]. On the contrary, OPF has extended this formulation to include power flow limits between buses, i.e., replicating the transmission network capacities (physical limitations). And so it constitutes a more complex problem, where ED and power flow problems are solved simultaneously [464]. That is to say, ED is only one of the many scalar OPF objectives [465]. Another cardinal aspect regarding ED is that it occurs after (solving) the Unit Commitment (UC) problem. As hinted by the name, this preceding task, involves the decision of which generation units will be committed (available) for dispatch, usually on a daily or weekly basis, in order to meet the expected load plus reserve requirements [466]. An optimization task that is mostly based on the fuel-cost curves in ($/h) that reflect startup costs and ramping limits, which in turn relate with the marginal cost of the units (derivative of the first with respect to power output), as given by the incremental-cost curves ($/MW). But also and among a multitude of others, heat rate curves, stockpile managements and maintenance needs.

Nowadays, the classic OPF problem is solved in different forms, for different horizons, i.e., for the continuous operation of power systems and for long-term investments, the OPF needs to cover multiple time-periods [467], with current assumptions being directly connected with the way electricity power markets are structured [468], and with an added level of randomness and complexity being brought by the increasing share of renewables [469]. Put differently, the presence of reasonably accurate predictions regarding expected generation and demand levels, as a result of steadier assumptions and more foreseeable changes in these variables, was a key premise of old-fashioned EPSs. Hence, contributing to fairly reliable solution of this NP-hard problem, at each upcoming time-period [470]. As such, not surprisingly, it has remained a core/trendy subject in the EPS literature, given its implications in terms of future planning, operation, and scheduling of electrical power utilities. And as the name itself suggests, the objective is to ensure a reliable and reasonably priced flow of power between production (generating stations) and consumption (loads), through the different branches of the electrical grid [471]. In other words, determining the "optimal" active and reactive flow of power between adjacent buses, through transmission lines, and the respective nodal voltage parameters, often in

its polar form, for steady-state conditions [472]. Hence, constituting a classic minimization problem, where control variables are iteratively adjusted, while satisfying a variety of power flow constraints in the form of equality and inequality constraints [473]. As more non-dispatchable generation enters the market, this premise no longer holds true, given the discrepancies between predicted and realized values. This fact (if left un-dealt), can lead to constraint violations, and depending on the severity of the mismatch, it may result in safety problems. In accordance, the solutions given by standard deterministic OPF method need further correction, and so justifying the use of ”real-time” OPF approaches [474].

One very important distinction to start-off is to understand that the power or load flow problem (also known as power flow analysis) involves solving the system of nonlinear power flow equations for one time-period, i.e., determining the (OPF) state variables for a given set of load and generation conditions (taking into account a given set of control variables/settings). For example, in the case of PQ buses, this means finding the voltage magnitudes and respective angles (in the polar form). And so, it differs from the OPF, which aims to optimize an objective function (or multiple ones), while satisfying all the constraints (including these load flow equations) in the same horizon, and so it constitutes a higher hierarchy problem.

For the lower level task, numerical analysis iterative methods, such as Gauss-Seidel (GS), Newton-Raphson (NR) and later Fast Decoupled Load Flow (FDLF) are amongst the traditional options in OPF literature. Nonetheless, not without its shortcomings. In the case of GS, it suffers from slow convergence, when dealing with a vast number of buses in a radially distributed system [475], diverging in the presence of compensated lines (negative branch reactances) and by treating differently the slack bus choice. As for the NR, is regarded as the most powerful numerical method in terms of solution convergence, but this comes with a hefty computational cost for very large systems (computation of the Jacobian matrix and its derivatives) and compared with GS a different treatment of the slack bus (initial parameters) [476]. Recognizing these limitations, FDLF works as a simplified version of the NR method. Exploring the fact that bus admittance and Jacobian matrices tend to be extremely sparse for large-scale systems [477], decoupled constant matrices, can replace the power flow Jacobian matrix. This, coupled with sparse (linearized) system solving techniques, results in diminished memory requirements and significantly faster convergence times [471] without degradation in terms of solution accuracy. In accordance, FDLF methods are widely employed by power utilities to carry out power flow analysis [478]. A representation of these features but also their inherent shortcomings is given in [479], where a combination of NR and FDPF procedures was proposed, namely by resorting to the initial assumptions for the FDPF formulation to improve the convergence.

As for the higher level OPF problem, conventional techniques, include linear, quadratic and non-linear programming, the Newton’s method, the interior point method, gradient and cumulant-based methods, among others [480, 481]. Yet, these approaches, which are mostly deterministic in nature, have some limitations, namely when dealing with the

non-linearity and non-convexity traits of OPF problems, the potential of being trapped in local optimums, particularly for very complex (large) power systems, with an inherent burden posed by the handling of numerous complex constraints, which underscores issues of scalability [482,483]. In the search for alternatives, authors have found in the advent of metaheuristic algorithms (predominance of evolutionary and swarm-based ones), neural networks, fuzzy-logic, and other techniques, a great portfolio of practical and more scalable search space tools, to conduct this constrained optimization in a more consistent manner. These approaches, labeled as ”trendy”, are generally known to better cope with an inflated number of control variables and problem constraints, circumvent the multimodal behavior of the objective function, and more easily be adapted to multi-objective OPF problem formulations [484]. A brief hierarchy diagram of the main OPF methods, with a particular subcategory for the metaheuristic-based ones, is depicted in Figure 6.1.

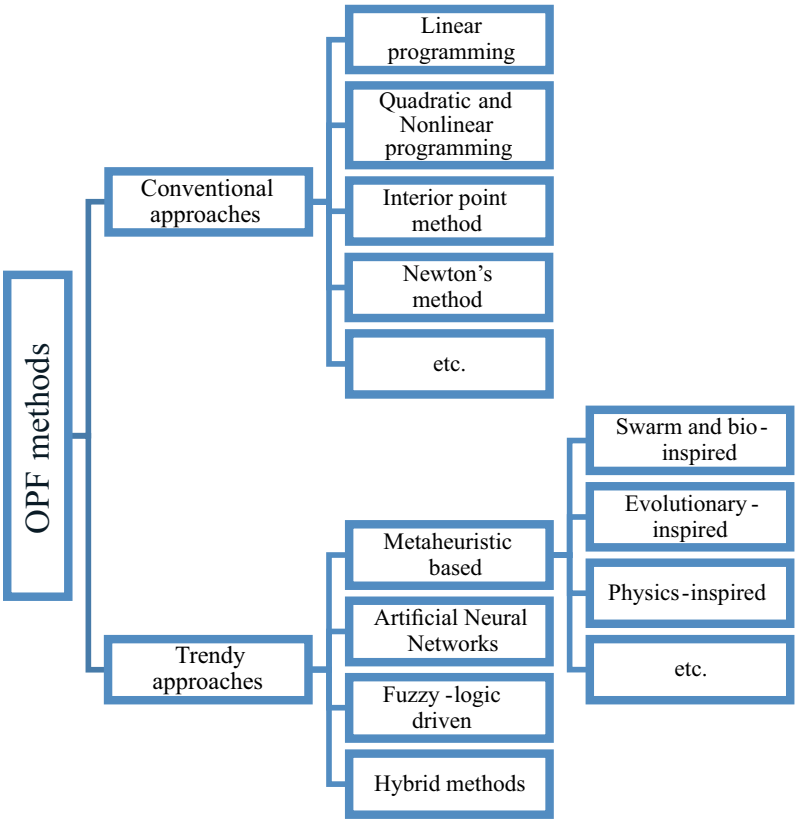


Figure 6.1: OPF literature: conventional vs trendy approaches.

Typically, before being employed, metaheuristics require an unconstrained OPF problem formulation, which in turn, is enforced using a penalty or bundle method to ensure that infeasible solutions are discarded. The same is confirmed by the apparatus on the OPF subject. For instance, in [485], several OPF problems are solved using the Differential Evolution (DE) as the basic search space algorithm, while also assessing 3-constraint handling techniques. While, in [486], a Tree-Seed algorithm is used for solving large-scale OPF problems. To accelerate the global optimum search and reduce the computation time, a metamodel-based global optimization with limited data is proposed in [487]. Hybrid

approaches are another trend, a good illustration is given in [488], where Moth Swarm Algorithm (MSA) and Gravitational Search Algorithm (GSA) are coupled to solve the OPF problem with and without wind energy sources. The interior point method is considered in [489] to surpass some weaknesses of the standalone metaheuristics in large-scale systems. A similar approach was followed by authors in [490], where the gradient pyramid mechanism and Nelder-Mead simplex are enhanced by Colony Predation algorithm for the optimal flow solution. In sum, these works exemplify the broad set of optimization methods and different problem interpretations considered in this field.

6.2 Fundamentals

As was briefly hinted in the introductory section (6.1), optimal power or load flow studies are a crucial task in the planning of future grid expansions or the restructuring of the existing ones, which invariably involves the addition of power injection and transmission sites to meet the growing load demands. OPF consists in solving a set of nonlinear equations for a steady-state condition, i.e., find the nodal voltages within certain tolerance levels to then compute real and reactive power injection at all the buses. Subsequently, line losses can also be found. Ideally, the system operation must be kept below overload conditions, in favor of system stability and longevity of the infrastructure. The same is done via a set of control variables, while complying with several operational constraints with the traditional goal of minimizing the generation costs. The assumption of a quasi-steady state means that an equilibrium was reached, i.e., no major changes occur in terms of system voltages, currents, etc. Nevertheless, this does not imply a stationary behavior. In fact, power systems are subject to transient phenomena and system disturbances. Examples of that include system contingencies, like faulty equipment, noise and electromagnetic interference, voltage surges or voltage sags, power outages, load variability, and renewable intermittency, among others.
The expected outline, where a high penetration of renewable energy sources is targeted, a fresh set of operational considerations for the problem is implicit, given the more atomized and distributed generation capacity, the phase out of base load thermal plants (e.g. coal-fired), the electrification of more human activities (e.g. transports), and the (indispensable) strengthening of the existing transmission and distribution networks. Factors that reinforce the need for a more flexible grid, with the appropriate load flow studies conducted. Therefore, for a complete comprehension of the problem, it is essential to first understand the key concepts and elements around the power flow problem.

6.2.1 Types of Buses and Transmission System Devices

The single-line diagram representation is a standardized schematic notation to represent the main elements of the EPS and its connections, namely: generators, transformers, busbars, conductors, capacitor banks, etc. As shown in previous sections, transmission networks are mostly comprised of three-phase circuits. However, under balanced conditions

(equal currents in terms of magnitude and phases separated by $2\pi/3$), the three-phase system can be analyzed using a per-phase equivalent circuit (per-unit circuit).
In load flow analyses, circuit breakers are commonly omitted from this representation. Transmission lines connect nodes – points in the system that have a non-zero impedance between them. For sufficiently large systems, these nodes denote what is referred to as busbars or more simply buses. In other words, a point from which power is either drawn/absorbed (consumption) and/or injected (generation). For the different types of buses, their representations are shown in Figure 6.2, while Table 6.1 resumes the specified and unspecified variables associated with each one; these variables include nodal complex voltages and powers, or in a decomposed manner $|V|, \phi, P$ and Q. This fact implies that, for an N-bus system, the total number of state variables is $4N$, of which $2N$ are unknowns—two per bus.

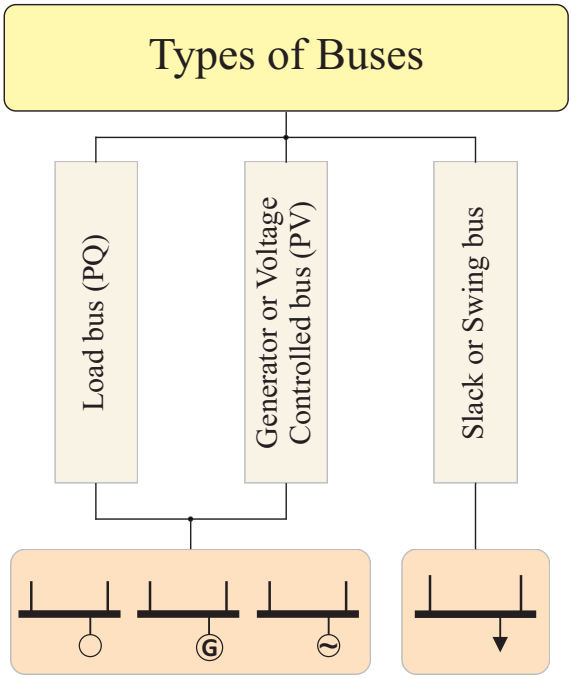


Figure 6.2: Types of Buses: name and most common representations.

Table 6.1: Power System Bus Types: Classification

Bus Type	Net Active Power (MW)	Net Reactive Power (MVAr)	Voltage (V)	
			Magnitude	***Angle***
Load (PQ)	Specified	Specified	Unspecified	Unspecified
Generator (PV)	Specified	Unspecified	Specified	Unspecified
Slack	Unspecified	Unspecified	Specified	Specified

When referring to the complex current (analogous to power) being drawn or injected by each type of bus, it is important to notice that we are interested in the net power (a nodal quantity), i.e., $S_{net} = S_G - S_{Ld}$, where the subscript G stands for generation, while Ld denotes a load. At each bus this net complex power must match the sum of the power flows from all the connected branches (when there is a transmission line). The same notation

is used below, where the additional index i associates the considered variable with the attributed bus number. In practical terms, active power and consequently power losses are primarily dependent on the differences between buses' voltage phase angles (flowing from higher to lower angles), whereas for reactive power the same is true but with respect to voltage magnitudes [491, 492].

A more in-depth explanation of these buses' role together with other common system devices, is provided in the subsections below.

6.2.1.1 Load (PQ) bus

Representing the different electric loads, it constitutes the bulk of the bus-bars in a typical power system. As the name indicates there is a load drawn (consumption), $P_{Ld,i}$ and $Q_{Ld,i}$, by the ith (load) bus, with no associated generation $(P_{G,i} = Q_{G,i} = 0)$. So, the target of the OPF (as part of the lower level power flow analysis) is to find both the voltage magnitude, $|V_i|$, and the phase angle, φ_i, while the complex power absorbed by this bus is known, $S_{net,i} = P_{Ld,i} + jQ_{Ld,i}$. Hence, the abbreviation "PQ". Conventionally, power flows drawn from these buses are defined as negative [493].

6.2.1.2 Generator (PV) or Voltage-Controlled bus

In these buses the real (generated) power, $P_{G,i}$, and the voltage magnitude, $|V_i|$, are kept constant, , with no associated load $(P_{Ld,i} = Q_{Ld,i} = 0)$ which again explains the used "abbreviation PV". Conventionally, a synchronous machine is the generator option in an EPS, where the output voltage is regulated for the different load conditions, by adjusting the DC current (exciter) flowing through the generator field winding (via an automatic voltage regulator). Operation in full load capacity tends to be more efficient. OPF will then serve to determine the unknown phase angle, φ_i, which subsequently allows us to calculate the reactive power, $Q_{G,i}$. Contrariwise to PQ buses, real power is injected by PV buses, so, conventionally $P_{G,i}$ is defined as positive. As for the reactive power, as a problem unknown, it can be either positive, null (unity power factor), or negative, depending on φ_i, if it is positive (lagging), then $Q_{G,i} > 0$ (feeding inductive loads). Whereas for a negative (leading), φ_i the bus is absorbing reactive power, and so $Q_{G,i} < 0$. The term "PV bus" is often used interchangeably with "voltage-controlled bus" or "regulated bus". However, in some specific contexts, these two last terminologies may refer to a broader category where voltage regulation is targeted (which includes slack buses as well). When solving the power flow system of equations, we may end-up with a situation, where the required reactive power, ends up outside the reactive power limits and so, this bus is effectively treated as a PQ bus, where $Q_{G,i}$ is fixed alongside $P_{G,i}$, and the $|V_i|$ is relaxed (varies). Meaning that the power flow equation is solved for φ_i and $|V_i|$, instead.

6.2.1.3 Slack or Swing or Reference bus

With its fixed voltage polar coordinates, this special type of bus sets a reference the phase angle from which the other buses phases are then measured, together with a constant magnitude, and so the voltage is commonly set to $1\angle 0°$ per units. In other words, the nominal voltage of the swing bus is the chosen base. Typically, for most load flow studies only one of these buses is required, and it is why is usually the numbered 1 bus in the considered EPS. The reactive and real power are the unknown variables, and therefore, the slack bus function is to balance the system by bridging the mismatch between the total scheduled generation and the total load including losses. Given its generation capabilities, the common representation for this type of bus is the same as for PV buses, as shown in Figure 6.2. Besides, differences in nodal phase angles impact synchronization between generators and voltage stability [494].

6.2.1.4 Transformers

Transformers are a critical device for the power system's functioning, establishing successive power transfer "bridges" between different voltage levels, thus, connecting generation plants to the consumption centers. Typically, the same is done with tap-changing transformers, whose function is to regulate the secondary voltage levels, subject to fluctuations in the primary side. By adjusting the connection point (taps) along the transformer windings (usually on the primary side or high voltage side), a variable turns ratio is achieved. Conventionally, the rated winding is selected at 100% tap, meaning a unitary nominal tap ratio of $t_{PS} = 1\,(\text{p.u.})$, which is directly tied with the transformer rated voltage. Several positive and negative taps, i.e., above and below this unitary nominal tap ratio, respectively, are used to adjust the secondary voltage V_S. Commonly with t_{PS} bounded between 0.9 and 1.1, and with taps arranged in 4 or 16 incremental steps in each direction, i.e., $\Delta t_{PS} = \pm 0.025$ or $\Delta t_{PS} = \pm 0.00625$ (for a more precise discrete control), respectively [495]. Nameplates on these transformers, do not always give the tapping range, as steps from the nominal value, or as percentage adjustment, instead different tap voltages in the primary side, V_P, are provided for different physical connections, which in turn, leads us to the mentioned variable turns ratio. Importantly, we see that the term "tap ratio" despite being related to the specific tap's settings, is often times used interchangeably with the term "turns ratio", and its associated variable, a.

The output voltage is governed by the turns ratio expression, $a = -N_P \frac{d\Phi}{dt} / -N_S \frac{d\Phi}{dt} \quad \Leftrightarrow a = \frac{N_P}{N_S}$, as derived from the Faraday's law of induction, where $\frac{d\Phi}{dt}$ is the time derivative of the magnetic flux, while N_P and $N_S$are the number of turns in the primary and secondary windings, respectively. So, in tap-changing transformers, the relation between the number of turns and the induced electromotive force 6.1 is altered, by adjusting the number of primary turns, while N_S is fixed:

$$V_S = V_P \frac{N_S}{N_P} \quad (6.1)$$

where a lower N_P and consequently, value of a, implies a higher V_S and vice versa. The possible deviation from the unitary nominal tap ratio (voltage), explains why these are known as off-nominal turns ratio transformers. The tap settings change, Δt_{PS}, is commonly done under load conditions (uninterrupted supply), and so they are also referred to as on-load tap-changing (OLTCs) transformers. These are usually equipped with automatic voltage and reactive power control. Phase-shifting is another feature of other modern OLTCs, which as the name suggests, automatically varies the phase shift angle to control the real power flow through a transmission line [496], and constituting an added control variable when performing the OPF, as we will latter explore.

The single-line diagram representation of an EPS branch with one of these off–nominal turns ratio transformers, is shown in Figure 6.3 a) together with the relevant bus-side voltages and current injections, as well as the outflow current and voltage, $I_S^{'}$ and $V_S^{'}$ [497]. The same is modeled on the tap side by an ideal transformer with a turns ratio, $t_{PS} : 1$, while on the non-tap side, there is an admittance in series, Y_{PS}. To understand the line flows through the transformer in an OPF problem, it is useful to revisit its equivalent models. The Steinmetz's model is considered the conventional T-equivalent circuit for a single-phase two-winding transformer, with two series impedance branches (capturing the copper losses and inductance leakages in both the primary and secondary windings) and a shunt impedance magnetization branch on the primary side (capturing core losses and its magnetic energization, i.e., even with the secondary winding in open circuit, the primary still draws a small magnetizing current) [498]. A very complete but complex model, that contrasts, with the most simplistic equivalent circuit, solely based on a (single) series (short-circuit) admittance, which fails to capture the intricacies of tap-changing actions and its voltage control effects, and so neither of them are deemed as adequate for modeling off-nominal turns ratio transformers (either to avoid an excess of computational cost or for simply lacking the accuracy) [499].

Alternatively, considering the transformer's branch representation shown in Figure 6.3 a), the line current and bus voltage equations can be derived as:

$$\begin{cases} t_{PS} = \frac{-I_S^{'}}{I_P} \\ I_S = Y_{PS}\left(V_S - V_S^{'}\right) \end{cases} \Leftrightarrow \begin{cases} I_P = \frac{-I_S}{t_{PS}} \\ I_S = Y_{PS}\left(V_S - \frac{V_P}{t_{PS}}\right) \end{cases} \quad (6.2)$$

Notice, how on the turns ratio equation, with $t_{PS} > 0$, the inverse current ratio has a negative numerator, given that arbitrarily, the current direction, I_S is flowing in the opposite direction of I_P, again as defined in Figure 6.3 a). Then, by substitution, the expressions in 6.2 can be rewritten into a familiar system of equations 6.3. Adequately translating the nonlinear dependence of these nodal relationships on the (real) off-nominal tap ratio

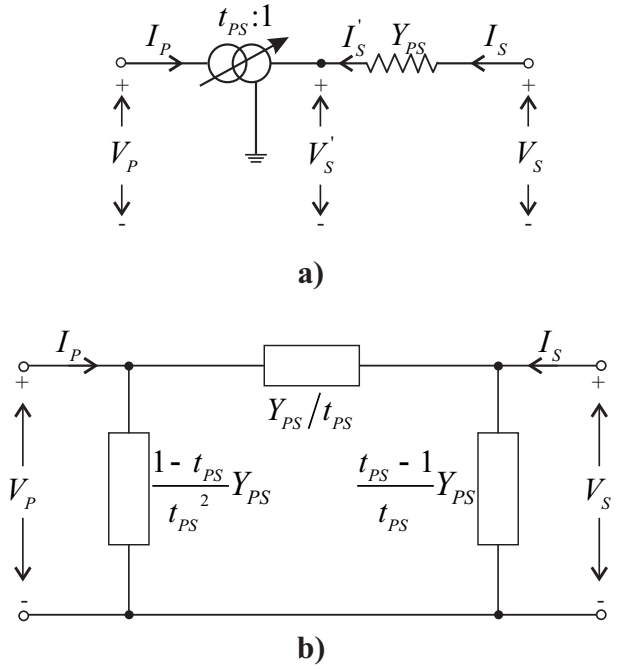


Figure 6.3: Tap-changing transformer: a) EPS branch representation; b) simplified π-type equivalent circuit.

settings, t_{PS}, or more broadly speaking the variable turns ratio, a.

$$\underbrace{\begin{bmatrix} Y_{PS}/{t_{PS}}^2 & -Y_{PS}/t_{PS} \\ -Y_{PS}/t_{PS} & Y_{PS} \end{bmatrix}}_{\text{admittances}} \begin{bmatrix} V_P \\ V_S \end{bmatrix} = \begin{bmatrix} I_P \\ I_S \end{bmatrix} \quad (6.3)$$

From 6.3 the π-equivalent circuit of the transformer emerges with two shunt branches and one series branch, as illustrated in Figure 6.3 b), much like the transmission line modeling (as we saw in sub-section 1.1.1), provides a simplified representation, with sufficient accuracy. From a flow study standpoint, the reduced number of parameters of this model speeds up calculations (bearable computational overhead), which is particularly useful for system-wide studies.

Also, observe how the admittance components are symmetric, a circumstance that does not hold for phase shifters, even though the simplified branch representation showed in Figure 6.3 a) is still suitable [500]. The same can be established provided that the turns ratio is now complex, $t_{PS} = |t_{PS}| \angle\chi$, and by deriving the nodal equations, similar to 6.2, but taking into account that with the phase-shifting, the current relation is now dependent on the complex conjugate of t_{PS}, as given by 6.4.

$$\begin{aligned} V_P I_P^* = V_S\left(-I_S^*\right) \Leftrightarrow I_S^* = \frac{V_P}{V_S} I_P^* \Leftrightarrow \\ -I_S^* = t_{PS} I_P^* \Leftrightarrow t_{PS}^* = \frac{-I_S}{I_P} \end{aligned} \tag{6.4}$$

So, after some manipulation, it leads us to the system of equations given in 6.5, where the admittance matrix is no longer symmetric, and so it no longer grants a π-equivalent circuit.a

$$\begin{bmatrix} Y_{PS}/t_{PS}t_{PS}^* & -Y_{PS}/t_{PS}^* \\ -Y_{PS}/t_{PS} & Y_{PS} \end{bmatrix} \begin{bmatrix} V_P \\ V_S \end{bmatrix} = \begin{bmatrix} I_P \\ I_S \end{bmatrix} \tag{6.5}$$

So, if we consider the polar form (coordinates) of the other nodal quantities as well:$Y_{PS} = |Y_{PS}|\angle\theta_{PS}$, $V_P = |V_P|\angle\varphi_P$ and $V_S = |V_S|\angle\varphi_S$, one can write the power flow (injection), $S_P^* = V_P^* I_P$, through a transformer as:

$$\begin{aligned} S_P^* = \frac{Y_{PS}V_P V_P^*}{t_{PS}t_{PS}^*} - \frac{Y_{PS}V_S V_P^*}{t_{PS}} \Leftrightarrow \\ S_P^* = \frac{|Y_{PS}|\angle\theta_{PS}|V_P|^2}{|t_{PS}|^2} - \frac{|Y_{PS}|\angle\theta_{PS}|V_S|\angle\varphi_S|V_P|\angle-\varphi_P}{|t_{PS}|\angle\chi} \Leftrightarrow \\ S_P^* = \frac{|Y_{PS}||V_P|^2}{|t_{PS}|^2}\angle\theta_{PS} - \frac{|Y_{PS}||V_S||V_P|}{|t_{PS}|}\angle\left(\theta_{PS}+\varphi_S-\varphi_P-\chi\right) \Leftrightarrow \\ S_P^* = \frac{|Y_{PS}||V_P|}{|t_{PS}|^2}\angle\theta_{PS}\left(|V_P| - |V_S||t_{PS}|\angle\left(\varphi_S-\varphi_P-\chi\right)\right) \end{aligned} \tag{6.6}$$

Or, in terms of active, P_P, and reactive, Q_P, power flows as:

$$\begin{cases} P_P = \frac{|Y_{PS}||V_P|}{|t_{PS}|^2}\cos\left(\theta_{PS}\right)\left(|V_P| - |V_S||t_{PS}|\cos\left(\varphi_P+\chi-\varphi_S\right)\right) \\ Q_P = -\frac{|Y_{PS}||V_P|}{|t_{PS}|^2}\sin\left(\theta_{PS}\right)\left(|V_P| - |V_S||t_{PS}|\sin\left(\varphi_P+\chi-\varphi_S\right)\right) \end{cases} \tag{6.7}$$

6.2.1.5 Other transmission system devices

As discussed priorly, voltage stability is key for a smooth EPS operation, with a steadfast load flow that minimizes equipment stress and prevents system-wide disruptions. To this intent, reactive power, unlike active power, does not perform useful work (e.g., mechanical, heating, or light outputs), yet it still plays a crucial role in supporting the power transmission (energy transfer). Providing the necessary energy flow to establish the magnetic fields required, for loads like AC motors or transformers, to properly function. An unbalance of these inductive loads, i.e., an insufficient flow of reactive power, originates a back electromotive force, that resists the rise in current, which in turn causes voltage

to sag. Without the proper correction, and to maintain the supplied power, the current increases, causing the system to consume more reactive power, which in turn cascades into a snowball effect with further voltage drops [501].
The transition to a new EPS paradigm has the potential to aggravate this problem, and so the (reactive power) voltage support capacity must be closely monitored [502]. Traditionally, VAR compensators are used to regulate the reactive power flow, and its dual functionality allows them to either inject or absorb reactive power, contributing to voltage stability, an improved load power factor, with reduced losses (otherwise inevitable due to an increased reactive power flow), thus, ultimately resulting in an augmented power transmission capability [503]. Capacitor banks, Synchronous Condensers, Static VAR Compensators (SVCs) and Static VAR Compensators (STATCOMs) are the most common types of VAR compensators, with STATCOMs being considered the faster and more reliable form of Shunt VAR compensation [504]. Both STATCOMs and SVCs are part of larger family of flexible AC transmission system (FACTS) devices [505], which are becoming more commonplace in advanced EPS, providing dynamic voltage support and increase the power transfer capability of the transmission network. This explains why, another OPF literature tendency is to determine the optimal placement and sizing of FACTS devices [506].

6.2.2 Power flow equations and the admittance matrix

Derived from the nodal analysis at each bus, power flow system of equations relates current injections at each bus (node) with bus voltages, and with admittances as constants of proportionality 6.8, as illustrated in Figure 6.4.
For a N bus system, theoretically composed of buses, transmission lines, transformers, passive impedances and series elements, and shunts [507], this relation is given by the generalized Ohm's law:

$$I_i = \sum_{k=1}^{N} Y_{ik} V_k \quad (6.8)$$

where I_i is the current injection at bus i; k is the shifting index to rotate every nodal variable of the multi-bus system; Y_{ik} is an element of the bus admittance matrix, Y_{bus}, as presented in 6.9 and it stands for the admittance between the ith and kth bus; V_k is the correspondent kth bus voltage.

$$Y_{bus} = \begin{bmatrix} Y_{11} & \cdots & Y_{1N} \\ \vdots & \ddots & \vdots \\ Y_{N1} & \cdots & Y_{NN} \end{bmatrix} \quad (6.9)$$

Faced with a case-study network where the connection between buses is exclusively formed by transmission lines, the following admittance matrix properties can be emphasized: i) it

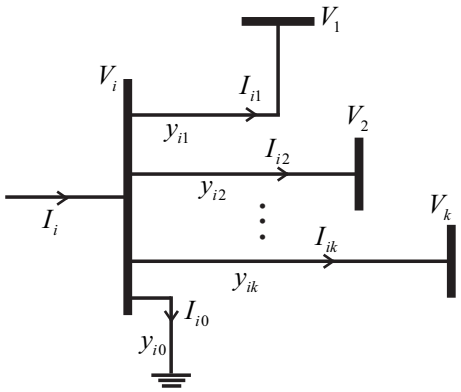


Figure 6.4: Nodal current injections from interconnected buses (power system illustration).

is symmetric, $Y_{ik} = Y_{ki}$, for all $i, k \in [1, N]$; ii) diagonal elements of Y_{bus}, are referred to as self-admittance or driving-point admittance of a system, representing the total admittance of the ith bus, i.e. $Y_{ii} = \sum_{k=0, k \neq i}^{N} y_{ik}$; iii) off-diagonal matrix elements are known as transfer or mutual admittances, and these are computed as the negative of the admittance connecting buses i and k, $Y_{ik} = -y_{ik}$; iv) this transfer admittance, y_{ik}, is computed as given in 6.10, for any two buses, connected by a number of c_Y transmission lines (providing that more than one line can be affixed from the ith to the kth bus); iv) for unconnected branches, $y_{ik} = 0$, which means that for large EPSs, Y_{bus} is a sparse matrix (granted that in real cases the majority of the buses are not physically connected).

$$\begin{cases} y_{ik} = \sum_{c_Y} \dfrac{1}{R_{ik,c_Y} + jX_{ik,c_Y}} \Leftrightarrow \\ y_{ik} = \sum_{c_Y} \overbrace{\left(\dfrac{R_{ik,c_Y}}{R_{ik,c_Y}^2 + X_{ik,c_Y}^2} \right)}^{G_{ik,c_Y}} + j \overbrace{\left(\dfrac{-X_{ik,c_Y}}{R_{ik,c_Y}^2 + X_{ik,c_Y}^2} \right)}^{B_{ik,c_Y}} \end{cases} \quad (6.10)$$

Where R_{ik,c_Y}, X_{ik,c_Y}, G_{ik,c_Y} and B_{ik,c_Y} are the series resistance, reactance, conductance and susceptance, respectively, of a given c_Y transmission line from i to k.

Conversely, the ith (nodal) current injection 6.8, can also be derived from Kirchhoff's current law for the circuit in Figure 6.4, considering the bus voltage, V_i and its adjacent (connected) bus voltages and admittances, i.e., $V_1, V_2, \ldots, V_k$ and $y_{i0}, y_{i1}, y_{i2}, \ldots, y_{ik}$, respectively, as given below:

$$I_i = y_{i0}\,(V_i - 0) + (V_i - V_1)\,y_{i1} + (V_i - V_2)\,y_{i2} + \ldots + (V_i - V_k)\,y_{ik} \quad (6.11)$$

Then, by rearranging the terms as a function of the bus(es) voltages, 6.11 becomes:

$$I_i = \sum_{k=0,k\neq i} y_{ik}V_i - \sum_{k=1,k\neq i} y_{ik}V_k \quad (6.12)$$

As a result of the inferred Y_{bus} properties, equation 6.12 can be rewritten into a more familiar notation for power flow studies, that is:

$$I_i = Y_{ii}V_i + \sum_{k=1,k\neq i} Y_{ik}V_k = \sum_{k=1} Y_{ik}V_k \quad (6.13)$$

Nevertheless, as we saw in section 6.2.1, besides the usual transmission lines connecting the different system buses, studied systems often involve the use of tap-changing and phase-shifting transformers. Similarly to transmission lines, these establish a branch between nodes i and k, and are characterized by the complex off-nominal tap ratio, $t_{PS} = |t_{PS}|\left(\cos\left(\chi\right) + j\sin\left(\chi\right)\right)$. In load-studies, and for the sake of simplicity, only a single transformer forms the branch, and no transmission lines (admittances) are jointly contemplated. So, the off-diagonal elements of Y_{bus}, that concern these transformers branches, are calculated differently, in accordance with 6.5, i.e., $Y_{ik} = -\frac{y_{ik}}{t^*_{PS}}$ and $Y_{ki} = -\frac{y_{ik}}{t_{PS}}$, which as we saw in the case of phase-shifters, breaks the symmetry of Y_{bus}. As for the diagonal terms, the branches with transformers, entail an added term to the already computed self-admittances, again in accordance with 6.5, i.e., $Y^{new}_{ii} = Y_{ii} + \frac{y_{ik}}{t_{PS}t^*_{PS}}$ and $Y^{new}_{kk} = Y_{kk} + y_{ik}$. Besides, as previously mentioned, shunt elements are only reflected in the diagonal admittance terms, and they may include VAR compensators, transmission line/transformer shunt admittances and line charging susceptance. So, the complete equation for diagonal admittance terms, Y^{total}_{ii}, is computed as follows:

$$Y^{total}_{ii} = Y^{new}_{ii} + \underbrace{\left(G^{SC}_i + jB^{SC}_i\right)}_{\text{Shunt Compensator}} + \sum_{c_Y}\left(\underbrace{G^{Y}_{i,c_Y} + jB^{Y}_{i,c_Y}}_{\text{Branch Admittance}} + \underbrace{j0.5B^{ch}_{i,c_Y}}_{\text{Branch Charging}}\right) \quad (6.14)$$

where G^{SC}_i and B^{SC}_i denote the (hypothetical) shunt compensator conductance and susceptance at bus i; c_Y is again a given index for the transmission line/transformer from i to k; and so G^{Y}_{i,c_Y} and B^{Y}_{i,c_Y} represent a given c_Y branch conductance and susceptance contribution, with a connection to the ith node; while $0.5B^{ch}_{i,c_Y}$ stands for half of the line charging (susceptance) at the same c_Y branch. To sum up, Y_{bus} is an ubiquitous matrix[1] in power system modelling, and in realistic, more complex systems, is a structurally symmetric sparse matrix, meaning that if many of the non-diagonal elements are null (a bus is only connected to a few others), and if $Y_{ik} \neq 0$, then $Y_{ki} \neq 0$. Consequently, in flow and stability studies, where Y_{bus} is used to determine nodal injections, or the Jacobian matrix, that contains the partial derivatives of these injections, sparse matrix techniques

[1] In other words, meaning that the nodal admittance matrix is constantly encountered.

can be applied to reduce the number of arithmetic operations and the memory requirements [508].
Considering the per-unit complex power for the ith bus, $S_i = V_i I_i^*$, where I_i^* is the complex conjugate of the current, we can write $I_i^* = \frac{P_i + jQ_i}{V_i} \Rightarrow P_i - jQ_i = V_i^* I_i$, and so replacing the current term with 6.8, the (complex conjugate) power supplied to the system can be computed as follows:

$$P_i - jQ_i = V_i^* \sum_{k=1}^{N} Y_{ik} V_k \quad (6.15)$$

where V_i^* is the complex conjugate of the ith voltage. Then, if we go a step further and define $Y_{ik} = |Y_{ik}| \angle \theta_{ik}$, $V_i^* = |V_i| \angle -\varphi_i$ and $V_k = |V_k| \angle \varphi_k$, the active and reactive power components can be formulated as:

$$\begin{cases} P_i = \sum\limits_{k=1}^{N} |V_i| \, |Y_{ik}| \, |V_k| \cos\left(\theta_{ik} + \varphi_k - \varphi_i\right) \\ Q_i = -\sum\limits_{k=1}^{N} |V_i| \, |Y_{ik}| \, |V_k| \sin\left(\theta_{ik} + \varphi_k - \varphi_i\right) \end{cases}, \forall\, i \in [1, N] \quad (6.16)$$

6.3 The Optimal Power Flow Problem

Having established the unknown variables associated with each system bus and having specified the power flow equations that govern them - by mapping a set of complex voltage phasors to a set of complex nodal power injections [509] - OPF emerges as a constrained optimization problem, involving the minimization of a nonlinear objective function, $OF(x, u)$. Traditionally, this function translates the total generation costs of the system 6.19, while subject to a series of equality, $f(x, u) = 0$ 6.24 or 6.25, and inequality constraints, $g(x, u) \leq 0$ 6.29- 6.33; dynamic variables, x and u represent state and control variables, respectively. State (dependent) variables characterize the system and are continuous type parameters [510]. They evolve for changing steady-state conditions (different loads, generation and also set points for the control variables). Typically, state variables include active power from the slack bus, reactive power injections from slack and PV buses, voltage magnitude of all the load buses and the transmission line power flows.[2] Key to system stability, bus voltage phase angles (except for the slack bus, where its normally imposed, φ_1), effectively drive the active power flow along the transmission network, and despite rarely being made explicit in the formulation commonly used to represent u, are indeed state variables of the OPF problem as well [470,511]. And so the same is reflected in the vector form of u, as given in 6.17. While control (independent) variables, given by 6.18, are actively adjusted by the operators (control actions) to simultaneously

[2] That is to say, the state variables vector is basically formed by all the bus unknown variables, as defined in Table 6.1, and potentially other secondary variables.

minimize the defined objective function and satisfy the problem constraints. These can be either continuous (as is the case of real power generation and the voltage magnitude of PV buses) or discrete (as is the case of transformer tap settings and switched shunt reactive components) [510, 511]. Additional control variables may include phase-shift angles, load shedding settings, renewable curtailment settings, energy storage systems SoC, currents instead of power injections, among several others [510–514].
Importantly to better distinguish between the different ith bus variables, besides the already mentioned subscript denoting a generation or load type variable, which in turns indicate us the type of bus, each of these elements has also its own/associated indexes to identify the system component being analyzed. Where the subscripts v, q,p, s, and h, are associated with a given generator, load, transformer, compensator and transmission line, respectively. Providing a useful index distinction for the problem's constraints and the objective functions, as distinguished in this section formulation. To exemplify, the notation, $P_{G,4}$, denotes the active power from the 4th generator of the considered power system, nevertheless, this numbering, $v = 4$, does not have to match the attributed bus number i, i.e., this generator may actually be part of the 9th bus (according to the predefined test case numeration), and so $i \neq v$. Yet, in other instances as is the case of the slack bus variables, given that by convention $i = 1$, then $i = v$.

$$x^T = \begin{bmatrix} P_{G,1}, |V_{Ld,1}|, ..., |V_{Ld,q}|, ..., |V_{Ld,N_{PQ}}|, \\ \varphi_2, ..., \varphi_i, ..., \varphi_N, Q_{G,1}, ..., Q_{G,v}, \\ ..., Q_{G,N_{PV}}, S_{Ld,1}, ..., S_{Ld,h}, ..., S_{Ld,N_L} \end{bmatrix} \quad (6.17)$$

$$u^T = \left[P_{G,2}, ..., P_{G,N_{PV}}, |V_{G,1}|, ..., |V_{G,N_{PV}}|, \varphi_1, Q_{C,1}, ..., Q_{C,N_C}, T_1, ..., T_{N_T} \right] \quad (6.18)$$

Where $P_{G,v}$ denotes the output power from a certain v generator source, from a total of N_{PV} generators (PV buses), excluding the known slack active power, $P_{G,1}$; similarly,$Q_{G,v}$ denotes the reactive power at each v generator; conversely, $V_{Ld,q}$ denotes the voltage magnitude at each q PQ bus, from a total of N_{PQ} loads; φ_i stands for ith bus (voltage) phase angle in a N bus power system; $S_{Ld,h}$ is the apparent power flow at each h line, from a total of N_L transmission lines; $Q_{C,s}$ is the s reactive power compensation, from a total of N_C shunt VAR compensators (or shunt element reactances); lastly, T_t stands for the tap settings of the t regulating transformer(s) (OLTCs), from a total of N_T tap changers [515].
In real terms, it is not easy for system operators to judge (qualify) certain control actions as more impactful than others, and so limiting its number. That is, prioritize adjustments to specific elements of 6.18, in detriment of others. Given that almost all of the control variables contribute in a non-separable way - which is not even strongly tied with their magnitude and individual changes, Δu_i - to not only satisfy the constraints (subsections 6.3.2 and 6.3.3), but also to minimize the objective function (subsection 6.3.1) [512].
Assuming a vector u with length N_{cv} and considering "precise" individual adjustments

Δu_i, a theoretical relation between the objective value, $OF(x,u)$, and the number of control actions can be established, highlighting the trade-off between a certain number of control actions and the expected OPF objective function minimization. The same is shown in Figure 6.5, where first, one can see that a minimum number of control actions is required to ensure problem feasibility (when none of the constraints are violated and so with no penalty factors), and from this point onwards, the chosen $OF(x,u)$ tends to decline quite rapidly up until a point of (near) stagnation, when the number of control actions approaches N_{cv}. Hence, translating the need for sensitivity analysis to be conducted, in order to, virtually, case by case, identify the most effective control variables, as done by authors in [516] while considering a PSO algorithm for the optimization. A proper identification of the "dominant" control variables can lead to a reduced problem dimensionality, and consequently, a diminished computational burden for the optimization algorithm, which is particularly useful for large systems.

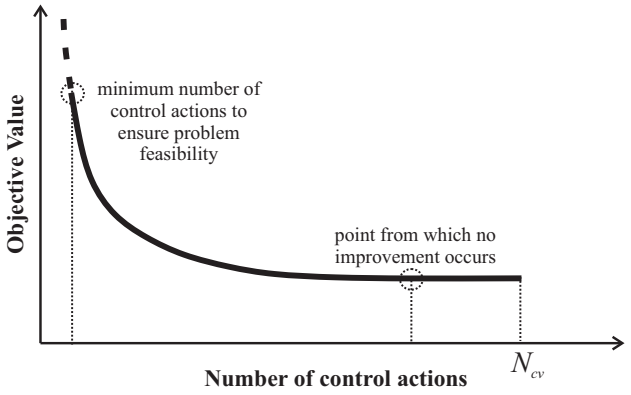


Figure 6.5: OPF objective function value vs number of control actions.

6.3.1 Objective Function(s)

When performing AC power flow optimization, there are distinct goals that can be targeted, codified through different objective functions, which are generically represented by the variable $OF\,(x,u)$ in this work. These are not always complementary, and so it may imply trade-offs between goals. The most common are presented in 6.19- 6.23. Considering a quadratic fuel-cost characteristic for each generator, OF_1 is the classic OPF objective function, aggregating all the thermal generation fuel costs, and it can be calculated as:

$$OF_1(x,u) = \sum_{i=1}^{N} \left(a_i + b_i P_{G,i} + c_i {P_{G,i}}^2\right) \quad (6.19)$$

where a_i is the fixed/maintenance costs; b_i is the running or marginal cost; c_i represents the rate at which cost increases, as the real power output, $P_{G,i}$, also increases; together they form the cost coefficients of the ith generator [517].

Nevertheless, the real functioning of a multi-valve steam turbine leads to a ripple behavior in the heat rate curve of the generator power output, in what is known as the "valve-point effect" as illustrated in Figure 6.6. Therefore, to capture this non-convex fuel-cost characteristic of thermal units, a non-smooth rectified sine term [518] is added to the fuel input–power output cost function, requiring two additional cost coefficients, d_i and e_i (6.20).

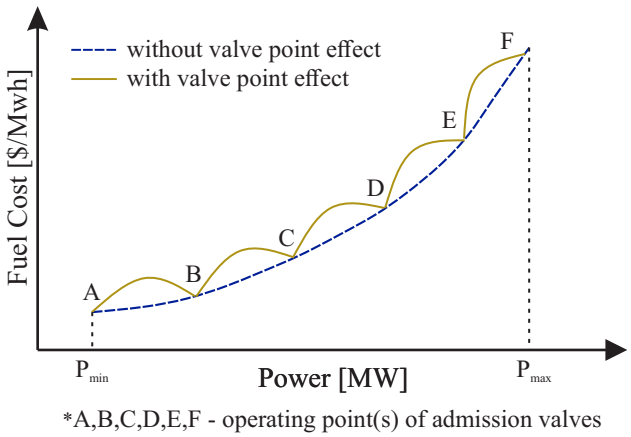


Figure 6.6: Valve point loading effect.

$$OF_2(x,u) = \sum_{i=1}^{N} \left(a_i + b_i P_{G,i} + c_i {P_{G,i}}^2\right) + \left|d_i \sin\left(e_i \left(P_{G,i}^{\min} - P_{G,i}\right)\right)\right| \quad (6.20)$$

Minimization of the active power or total transmission line losses is another well-known OPF objective function, and is expressed as follows:

$$OF_3(x,u) = \sum_{h=1}^{N_L} G_h \left[V_i^2 + V_k^2 - 2V_i V_k \cos\left(\varphi_i - \varphi_j\right)\right] \quad (6.21)$$

where N_L is the total number of transmission lines, G_h is the transfer conductance of the kth transmission line that links buses i and k.
Thermal generation impact in terms of (pollutant) gas emissions is major topic in today's "green" transformation, as such the associated total emissions [ton/h] are often minimized as well [[488], are given by:

$$OF_4(x,u) = \sum_{i=1}^{N} \left(\alpha_i + \beta_i P_{G,i} + \gamma_i {P_{G,i}}^2\right) + \mu_i e^{\xi_i \times P_{G,i}} \quad (6.22)$$

where α_i, β_i, γ_i, μ_i and ξ_i are the generator's fuel emission coefficients [519].
Recognizing the importance of voltage stability for a reliable power systems operation, the magnitude of the voltage deviation from the nominal voltage (at certain bus bars), can

also be minimized, accordingly:

$$OF_5(x,u) = \sum_{q=1}^{N_{PQ}} |V_q - V_{ref}| = \sum_{q=1}^{N_{PQ}} |V_i - 1| \quad (6.23)$$

the same is done for load buses, considering that voltage magnitude is one of its control variables, and so N_{PQ} is the number of PQ buses in the system and V_{ref} is the nominal voltage magnitude, usually expressed in per units, with a unitary value, as defined for the slack bus [488].

Beside these five, many more OPF objective functions can be found in the literature, and they are broadly categorized as: 1) generation-cost related; 2) power generation related; 3) demand related; 4) operational security related; 5) system efficiency related; 6) control devices related; 7) social/environmental related; 8) ancillary service related [514]. For instance, with regards to the aforementioned issue of reducing the number of control actions (subsection 6.3), one can find an objective function with this purpose in the control devices related OPF objectives. Thus, connecting the dots on how through several approaches, from sensitivity analysis, to specific objective formulations, and ultimately to, a targeted faster optimization convergence, system operators, recognize the progressive degradation and maintenance costs of executing numerous control actions.

6.3.2 Equality Constraints

Active and reactive power balance is an essential premise for the stability of the power system's operation. As such, the solution of the nonlinear equations given by 6.24, ensures that system loads are adequately served by the different power generation sources. So, from 6.16, these equality constraints take the following explicit form:

$$\begin{cases} P_{G,i} - P_{Ld,i} = \sum_{k=1}^{N} |V_i|\,|Y_{ik}|\,|V_k| \cos\left(\theta_{ik} + \varphi_k - \varphi_i\right) \\ Q_{G,i} + Q_{C,i} - Q_{Ld,i} = -\sum_{k=1}^{N} |V_i|\,|Y_{ik}|\,|V_k| \sin\left(\theta_{ik} + \varphi_k - \varphi_i\right) \end{cases}, \forall\, i \in [1, N] \quad (6.24)$$

where $Q_{C,i}$ is the shunt VAR compensator associated with the ith bus. Alternatively, by replacing the line series admittances between buses i and k 6.16, by its rectangular form, $Y_{ik} = G_{ik} + jB_{ik}$(more common than the polar form), using the Euler's formula[3] and considering that $\cos(x)$ and $\sin(x)$, are respectively, even and odd functions[4], we can rewrite 6.24, to factor the real and imaginary parts of the admittances, i.e., the conductance, G_{ik}, and susceptance, B_{ik}, terms, as shown below:

[3] $e^{j\theta} = \cos(\theta) + j\sin(\theta)$

[4] $\cos(-\theta) = \cos(\theta)$ and $\sin(-\theta) = -\sin(\theta)$

$$\begin{cases} P_{G,i} - P_{Ld,i} = \sum\limits_{k=1}^{N} |V_i|\,|V_k|\,(G_{ik}\cos(\varphi_i - \varphi_k) + \\ \qquad B_{ik}\sin(\varphi_i - \varphi_k)) \\ Q_{G,i} + Q_{C,i} - Q_{Ld,i} = \sum\limits_{k=1}^{N} |V_i|\,|V_k|\,(G_{ik}\sin(\varphi_i - \varphi_k) - \\ \qquad B_{ik}\cos(\varphi_i - \varphi_k)) \end{cases}, \forall\, i \in [1, N] \quad (6.25)$$

with the full derivation provided below:

$$\begin{aligned} P_i - jQ_i = V_i^* \sum_{k=1}^{N} Y_{ik} V_k \Leftrightarrow P_i - jQ_i = V_i^* \sum_{k=1}^{N} (G_{ik} + jB_{ik})\, V_k \Leftrightarrow \\ P_i - jQ_i = |V_i|\, e^{-j\varphi_i} \sum_{k=1}^{N} |V_k|\, e^{j\varphi_k}\, (G_{ik} + jB_{ik}) \Leftrightarrow \\ P_i - jQ_i = \sum_{k=1}^{N} |V_i|\,|V_k|\, e^{j(\varphi_k - \varphi_i)}\, (G_{ik} + jB_{ik}) \Leftrightarrow \\ P_i - jQ_i = \sum_{k=1}^{N} |V_i|\,|V_k|\,(\cos(\varphi_k - \varphi_i) + j\sin(\varphi_k - \varphi_i))\,(G_{ik} + jB_{ik}) \Leftrightarrow \\ P_i - jQ_i = \sum_{k=1}^{N} |V_i|\,|V_k|\cos(\varphi_k - \varphi_i)\, G_{ik} + |V_i|\,|V_k|\cos(\varphi_k - \varphi_i)\, jB_{ik} + \\ \sum_{k=1}^{N} |V_i|\,|V_k|\, j\sin(\varphi_k - \varphi_i)\, G_{ik} + |V_i|\,|V_k|\, j\sin(\varphi_k - \varphi_i)\, jB_{ik} \end{aligned} \quad (6.26)$$

$$\begin{cases} P_i = \sum\limits_{k=1}^{N} |V_i|\,|V_k|\,(G_{ik}\cos(\varphi_k - \varphi_i) - B_{ik}\sin(\varphi_k - \varphi_i)) \\ -Q_i = \sum\limits_{k=1}^{N} |V_i|\,|V_k|\,(G_{ik}\sin(\varphi_k - \varphi_i) + B_{ik}\cos(\varphi_k - \varphi_i)) \end{cases} \Leftrightarrow$$

$$\begin{cases} P_i = \sum\limits_{k=1}^{N} |V_i|\,|V_k|\,(G_{ik}\cos(\varphi_i - \varphi_k) + B_{ik}\sin(\varphi_i - \varphi_k)) \\ Q_i = \sum\limits_{k=1}^{N} |V_i|\,|V_k|\,(G_{ik}\sin(\varphi_i - \varphi_k) - B_{ik}\cos(\varphi_i - \varphi_k)) \end{cases}$$

Having established the power balance equations 6.26, the net line flows can also be written in terms of the control bus voltage magnitudes and its phase-angles, off-nominal tap settings, and the susceptance values of VAR compensated systems [520]. Hence, the active and reactive power flow equations over the branch connecting buses i and k is given by 6.27, assuming that the connecting branches are modeled as the π-equivalent [521].

$$\begin{cases} P_{ik} = |V_i|^2 \left(\frac{G_{ik}}{T_i^2} + G_i\right) - \frac{|V_i|\,|V_k|}{T_i}\,(G_{ik}\cos(\varphi_i - \varphi_k) + \\ \qquad B_{ik}\sin(\varphi_i - \varphi_k)) \\ Q_{ik} = -|V_i|^2 \left(\frac{B_{ik}}{T_i^2} + B_i\right) - \frac{|V_i|\,|V_k|}{T_i}\,(G_{ik}\sin(\varphi_i - \varphi_k) - \\ \qquad B_{ik}\cos(\varphi_i - \varphi_k)) \end{cases}, \forall\, i, k \in [1, N] \quad (6.27)$$

Where besides the already defined variables, T_i is the off-nominal tap setting at a given

bus i (similar to t_{PS}), this denominator term is only considered in the case of transformer branches, while for line branches $T_i = 1$; $G_i + jB_i$ is the sum of the shunt admittance, the half-line charging susceptance and the VAR compensation (external shunt admittance) if any, as given in 6.28, considering the formulated conductance and susceptance terms in 6.14.

$$\begin{cases} G_i = G_i^{SC} + \sum\limits_{c_Y} G_{i,c_Y}^{Y} \\ B_i = B_i^{SC} + \sum\limits_{c_Y} B_{i,c_Y}^{Y} + 0.5 B_{i,c_Y}^{ch} \end{cases} \quad (6.28)$$

Having established the power flows, one can compute the net branch current flow, as $I_{ik} = \frac{\sqrt{P_{ik}^2 + Q_{ik}^2}}{V_i}$.

6.3.3 Inequality Constraints

In turn, these constraints translate the operational limits of the different power system components. Starting with PV buses and its associated generators, the following restrictions apply:

$$\begin{aligned} & P_{G,v}^{\min} \leq P_{G,v} \leq P_{G,v}^{\max} \\ & Q_{G,v}^{\min} \leq Q_{G,v} \leq Q_{G,v}^{\max}, \ \forall\, v \in [1, N_{PV}] \\ & V_{G,v}^{\min} \leq V_{G,v} \leq V_{G,v}^{\max} \end{aligned} \quad (6.29)$$

where $P_{G,v}^{\min}$, $P_{G,v}^{\max}$, $Q_{G,v}^{\min}$, $Q_{G,v}^{\max}$, $V_{G,v}^{\min}$ and $V_{G,v}^{\max}$ denote the minimum and maximum active and reactive powers, as well as the voltage magnitude lower and upper limits.
With regards to the regulating OLTC transformers, the tap changer settings must be within the lower and upper limits 6.30, $T_t^{\min}$ and $T_t^{\max}$, respectively.

$$T_t^{\min} \leq T_t \leq T_t^{\max}, \ \forall\, t \in [1, N_T] \quad (6.30)$$

When these transformers have also a phase shifting ability, an added constraint 6.31, concerning the lower, $\delta_t^{\min}$, and upper, $\delta_t^{\max}$, phase shift angles, is required.

$$\delta_t^{\min} \leq \delta_t \leq \delta_t^{\max}, \ \forall\, t \in [1, N_T] \quad (6.31)$$

Shunt VAR compensators have also an operational range, and so the following restriction emerges:

$$Q_{C,s}^{\min} \leq Q_{C,s} \leq Q_{C,s}^{\max}, \ \forall\, s \in [1, N_C] \quad (6.32)$$

where $Q_{C,s}^{\min}$ and $Q_{C,s}^{\max}$ are the minimum and maximum reactive capacity that these compensators can source.
Last but not least, security or contingency constraints 6.33, typically concern the transmission capacity limits (safely carried), i.e., the lower, $S_{Ld,h}^{\min}$, and upper, $S_{Ld,h}^{\max}$, apparent power limits (MVA); as well as the minimum and maximum load voltages at the different PQ buses, $V_{Ld,q}^{\min}$ and $V_{Ld,q}^{\max}$.

$$\begin{array}{c} S_{Ld,h}^{\min} \leq S_{Ld,h} \leq S_{Ld,h}^{\max}, \forall\, h \in [1, N_L] \\ V_{Ld,q}^{\min} \leq V_{Ld,q} \leq V_{Ld,q}^{\max}, \forall\, q \in [1, N_{PQ}] \end{array} \tag{6.33}$$

Solution suboptimality cannot be discarded, given the nonlinear and non-convex nature of the AC OPF problem. Moreover, when the optimization does not satisfy all the constraints simultaneously, it yields an infeasible solution [522]. Control variables directly, and state variables indirectly, have an associated sensitivity value, i.e., the expected change in the objective function in response to a small variation in each of these variables, $\delta OF\,(x,u)/\delta u$. Ideally, a control variable should have zero sensitivity, thus, promoting system stability and robustness (less affected by disturbances and uncertainties) [523].

6.3.4 Iterative Methods for the power flow solution

As we saw in section 6.3.2 the OPF problem requires the solution of a nonlinear system of equations, concerning the nodal power balance of the system. Meaning that the roots of this system cannot be directly and exactly computed, without an explicit (closed-form) solution [524]. A trait that rendered iterative methods instrumental for the field development [482]. Particularly, given its inner ability to handle the nonlinearity and scalability of the problem (potentially a very large system), and so constituting a well-suited path (even nowadays) to solve the power flow analysis, that is the low-level hierarchy problem (as previously mentioned). Starting with an initial estimate/guess, and as the name suggests, these methods' principle is to iteratively refine the solution and gradually converge towards the optimum. In the subsections below we will explore the most popular iterative methods and how are they specifically adapted to find the power flow unknowns (state variables of the OPF problem).

6.3.4.1 Gauss-Seidel method

Based on Gauss's method, GS or the successive displacements method, is the simplest fixed-point iterative method to solve the OPF, and it follows the following steps [8]: i) rearrange the system of equations into the form $x = g\,(x)$; ii) establish an initial estimation of the solution (control variables), x^0, and compute $g\left(x^0\right)$; iii) update the solution estimate, $x^1 = g\left(x^0\right)$, or more broadly (for further iterations), $x^{iters+1} = g\left(x^{iters}\right)$; iv) repeat this iterative process until the Euclidian norm between the current solution and previous solution is within an acceptable tolerance, ε, i.e., $\left\|x^{iters+1} - x^{iters}\right\| \leq \varepsilon$. A stopping cri-

terion which is also referred to as absolute tolerance. Others include the residual norm, Max norm, backward error among others [525]. Besides, to improve the convergence rate, acceleration factors based on a step-size parameter, α, are often considered [526]. This involves recalculating the solution update, $x^{iters+1} = x^{iters} + \alpha \left[x^{iters+1} - x^{iters}\right]$ [472].

In the specific context of the power flow problem, both voltage parameters are unknown for PQ buses (Table 6.1), as such considering 6.13 and $P_i - jQ_i = V_i^* I_i$, we can rewrite the balance equation as an iteratively function of the voltage:

$$V_i^{iters+1} = \frac{1}{Y_{ii}} \left(\frac{P_i^{sch} - jQ_i^{sch}}{\left(V_i^{iters}\right)^*} - \sum_{k=1, k\neq i} Y_{ik} V_k^{iters} \right) \quad (6.34)$$

where besides the already defined variables, P_i^{sch} and Q_i^{sch} are the scheduled real and reactive power terms, respectively. For this type of bus, these known variables assume negative values, meaning a power absorption, as we saw in section 6.2.1. Usually, for the first iteration, the initial guess considers $V_i^0 = 1\angle 0° = 1 + j0$ p.u. Often, to accelerate convergence, the already updated iterative voltage parameters for the kth bus, when available $(i > k)$, are considered, and so V_k^{iters} is replaced by $V_k^{iters+1}$. Moreover, to reduce the number of calculations, the invariant quantities for the optimization, can be computed at the beginning and used for every iteration, and so 6.34 is often presented as:

$$V_i^{iters+1} = \frac{C_i}{\left(V_i^{iters}\right)^*} - \left(\sum_{k=1}^{i-1} D_{ik} V_k^{iters+1} + \sum_{k=i}^{N} D_{ik} V_k^{iters} \right) \quad (6.35)$$

where $C_i = \frac{P_i^{sch} - jQ_i^{sch}}{Y_{ii}}$ and $D_{ik} = \frac{Y_{ik}}{Y_{ii}}$.

In turn, PV buses present different assumptions, given that the voltage magnitude and the scheduled active power injection (positive value) are already known (Table 6.1), and so equation 6.34 or 6.35 is only used to determine the respective angle, φ_i. But first, considering 6.15, the net reactive power emerges as a function of the iterative voltages, i.e., $Q_i^{iters+1} = -\operatorname{Im}\left\{ \left(V_i^{iters}\right)^* \sum_{k=1}^{N} Y_{ik} V_k^{iters} \right\}$, and by the same token, the rationale regarding the already known voltage applies to the iteratively updated reactive power as well. Meaning that V_k^{iters} is replaced by $V_k^{iters+1}$, or in in its polar form, $\left|V_k^{iters}\right|$ and φ_k^{iters} are replaced by $\left|V_k^{iters+1}\right|$ and $\varphi_k^{iters+1}$, respectively, for $(i > k)$. Therefore, $Q_i^{iters+1}$ is computed as follows:

$$Q_i^{iters+1} = -\operatorname{Im}\left\{\left(V_i^{iters}\right)^* \left(\sum_{k=1}^{i-1} Y_{ik} V_k^{iters+1} + \sum_{k=i}^{N} Y_{ik} V_k^{iters}\right)\right\} \Leftrightarrow$$

$$Q_i^{iters+1} = -|V_i| \left(\begin{array}{l} \sum_{k=1}^{i-1} |Y_{ik}| \left|V_k^{iters+1}\right| \sin\left(\theta_{ik} + \varphi_k^{iters+1} - \varphi_i^{iters}\right) + \\ \sum_{k=i}^{N} |Y_{ik}| \left|V_k^{iters}\right| \sin\left(\theta_{ik} + \varphi_k^{iters} - \varphi_i^{iters}\right) \end{array} \right) \quad (6.36)$$

where $|V_i| = \left|V_i^{iters}\right|$, and for the first iteration, $\varphi_i^0 = 0°,\ \forall\, i \in [1, N]$.
Once $Q_i^{iters+1}$ is determined, and if $Q_{i,\min} \leq Q_i^{iters+1} \leq Q_{i,\max}$, we proceed to compute $V_i^{iters+1}$. Importantly, when updating the voltage for the ith PV bus using 6.34 or 6.35, we constantly end up with a different voltage magnitude (not in agreement with the specified value), given that for this type of buses, voltage is controlled, as such, when considering the polar form representation, the obtained $\left|V_i^{iters+1}\right|$ is discarded, i.e., $\left|V_i^{iters+1}\right| = \left|V_i^0\right| = |V_i|$, while holding the voltage phase angle, φ_i^{iters}. Instead, when using the rectangular form representation, the real part of the obtained iterative voltage is discarded and is then recalculated in order to meet the specified voltage magnitude, i.e., $\operatorname{Re}\left(V_i^{iters+1}\right) = \sqrt{|V_i|^2 - \operatorname{Im}\left(V_i^{iters+1}\right)}$. Alternatively, we can recalculate both the real and imaginary parts to satisfy the required magnitude (polar radius). The same is done by multiplying the computed voltage, by a ratio between the desired voltage magnitude and the computed one, $V_i^{iters+1} = \frac{|V_i|}{\left|V_i^{iters+1}\right|} V_i^{iters+1}$. As previously stated, an acceleration factor is often considered to improve the convergence rate of GS, meaning that for both type of buses, $V_i^{iters+1}$ is (yet) again corrected based on its current and prior value , as follows: $V_i^{iters+1} = V_i^{iters} + \alpha\left(V_i^{iters+1} - V_i^{iters}\right)$. Despite being case-specific, several empirical studies pointed to values of α, ranging between 1 and 1.8, as the most suitable [527].
After updating all the estimates for the respective buses' unknowns, the convergence is checked at the end of each iteration, i.e., $\left|V_i^{iters+1} - V_i^{iters}\right| \leq \varepsilon$, in an iterative process, that is repeated until this condition is met, and so the optimization is stopped. After that, the swing bus apparent power, line flows and losses can be computed using the nodal voltages.

6.3.4.2 Newton-Raphson method

Coined after the notorious Newton's work on fluxions and the independent Raphson's approach to approximating the roots of an equation, NR method, also known as method of successive approximations, is a widely used iterative approach for solving nonlinear systems of equations [475]. Starting with an initial guess (benefiting in terms of convergence if it is an "informed" choice), the NR method involves the use of several tangent lines, in what is a very intuitively geometric premise. These lines are drawn from successive root approximation points, which in turn derive from the intersection with the x-axis in a

Cartesian coordinate system, revealing consecutively closer root approximations (candidate solutions).

For continuous and infinitely differentiable functions, Taylor's Expansion makes use of a series of increasing derivative orders and powers to approximate an arbitrary (complicated) function, $g(x)$, around a specific point. So, if x^p denotes the point where the derivative is taken, and $x^p + \Delta x$ denotes a neighboring point, we wish to approximate, then one can write [528]:

$$g(x^p + \Delta x) = \sum_n \frac{1}{n!} g^{n'}(x^p)(\Delta x)^n, \; n \in \mathbb{N}_0 \quad (6.37)$$

With this in mind, NR method works by approximating a set of nonlinear equations to a set of linear ones, using Taylor's expansion truncated to the first order approximation [471]. In other terms, $n = 1$ implies $g(x^p + \Delta x) \approx g(x^p) + g'(x^p)\Delta x$. A choice that strikes a balance between accuracy and simplicity since higher order approximations would require more function evaluations, which is not justified for smaller Δx, when the higher order terms yield neglectable contributions, and so they can be safely disregarded. The objective is then to find a Δx (in practice several ones, through a sequence of iterative steps), that satisfies for the first order approximation, $g(x^p + \Delta x) \approx 0$, so considering that $\Delta x = x^{iters+1} - x^{iters}$, we find that the next guess(es) are iteratively computed as:

$$x^{iters+1} = x^{iters} - \frac{g\left(x^{iters}\right)}{g'\left(x^{iters}\right)} \quad (6.38)$$

Meaning, it leverages the information regarding the function local behavior, to find the function roots. Overall, a small ratio is an indication of a zero-crossing approximation, given that $g\left(x^{iters}\right) \ll g'\left(x^{iters}\right)$, and so $g\left(x^{iters}\right) \to 0$. On the contrary if $g'\left(x^{iters}\right) \to 0$, that may cause the method to diverge or oscillate around critical points or saddle points, and a new initial guess maybe required [529]. Absolute tolerance, $\left|x^{iters+1} - x^{iters}\right| \leq \varepsilon$, establishes a threshold for the absolute difference between successive iterations, and when it indicates minimal changes, e.g., $\varepsilon = 10^{-6}$, a good root approximation was achieved. Other criteria include residual and relative tolerances, different norms, or a combination of those. For instance, in [530] both the residual and absolute tolerance are considered.

Transposing this into the load flow problem, where 6.16 translates the system of nonlinear equations, and as we saw for the GS method, the unknowns are the voltage parameters for PQ buses, while for PV buses, the unknowns are the voltage angle and associated bus reactive power (Table 6.1). As such, according to 6.38, the linearized relationship between changes in the independent variables, i.e., $\Delta\left|V_i^{iters}\right|$ and $\Delta\varphi_i^{iters}$, and the changes in the real and reactive power components, ΔP_i^{iters} and ΔQ_i^{iters}, as functions of these voltage parameters, can be formalized in matrix form, for a $N = N_{PQ} + N_{PV} + 1$ bus system, as given by 6.39. With the power residuals (mismatch between the scheduled and the com-

puted power quantities) are given as: $\Delta P_i^{iters} = P_i^{sch} - P_i^{iters}$ and $\Delta Q_i^{iters} = Q_i^{sch} - Q_i^{iters}$. Besides, notice how in 6.39, the swing bus variables, which as we saw in subsection 6.2.1.3 are usually numbered as 1, are omitted from the matrices (and are computed afterwards).

$$\underbrace{\begin{bmatrix} \Delta P_2^{iters} \\ \vdots \\ \Delta P_N^{iters} \\ \Delta Q_2^{iters} \\ \vdots \\ \Delta Q_{N_{PQ}+1}^{iters} \end{bmatrix}}_{(N+N_{PQ}-1\times 1)} = \underbrace{\begin{bmatrix} \overbrace{\begin{bmatrix} \frac{\delta P_2^{iters}}{\delta \varphi_2^{iters}} & \cdots & \frac{\delta P_2^{iters}}{\delta \varphi_N^{iters}} \\ \vdots & \ddots & \vdots \\ \frac{\delta P_N^{iters}}{\delta \varphi_2^{iters}} & \cdots & \frac{\delta P_N^{iters}}{\delta \varphi_N^{iters}} \end{bmatrix}}^{J_{P\varphi}^{iters}} & \overbrace{\begin{bmatrix} \frac{\delta P_2^{iters}}{\delta \left|V_2^{iters}\right|} & \cdots & \frac{\delta P_2^{iters}}{\delta \left|V_{N_{PQ}+1}^{iters}\right|} \\ \vdots & \ddots & \vdots \\ \frac{\delta P_N^{iters}}{\delta \left|V_2^{iters}\right|} & \cdots & \frac{\delta P_N^{iters}}{\delta \left|V_{N_{PQ}+1}^{iters}\right|} \end{bmatrix}}^{J_{P|V|}^{iters}} \\ \overbrace{\begin{bmatrix} \frac{\delta Q_2^{iters}}{\delta \varphi_2^{iters}} & \cdots & \frac{\delta Q_2^{iters}}{\delta \varphi_N^{iters}} \\ \vdots & \ddots & \vdots \\ \frac{\delta Q_{N_{PQ}+1}^{iters}}{\delta \varphi_2^{iters}} & \cdots & \frac{\delta Q_{N_{PQ}+1}^{iters}}{\delta \varphi_N^{iters}} \end{bmatrix}}^{J_{Q\varphi}^{iters}} & \overbrace{\begin{bmatrix} \frac{\delta Q_2^{iters}}{\delta \left|V_2^{iters}\right|} & \cdots & \frac{\delta Q_2^{iters}}{\delta \left|V_{N_{PQ}+1}^{iters}\right|} \\ \vdots & \ddots & \vdots \\ \frac{\delta Q_{N_{PQ}+1}^{iters}}{\delta \left|V_2^{iters}\right|} & \cdots & \frac{\delta Q_{N_{PQ}+1}^{iters}}{\delta \left|V_{N_{PQ}+1}^{iters}\right|} \end{bmatrix}}^{J_{Q|V|}^{iters}} \end{bmatrix}}_{(N+N_{PQ}-1\times N+N_{PQ}-1)} \underbrace{\begin{bmatrix} \Delta \varphi_2^{iters} \\ \vdots \\ \Delta \varphi_N^{iters} \\ \Delta \left|V_2^{iters}\right| \\ \vdots \\ \Delta \left|V_N^{iters}\right| \end{bmatrix}}_{(N+N_{PQ}-1\times 1)} \tag{6.39}$$

The matrix of partial derivatives, which is known as the Jacobian matrix, a $N + N_{PQ} - 1 = 2N_{PQ} + N_{PV}$ square and invertible matrix, establishes a linearized relationship between the changes in the power components and the voltage parameters. Often, for conciseness, the NR power flow relation can simply be put as:

$$\begin{bmatrix} \Delta P^{iters} \\ \Delta Q^{iters} \end{bmatrix} = \begin{bmatrix} J_{P\varphi}^{iters} & J_{P|V|}^{iters} \\ J_{Q\varphi}^{iters} & J_{Q|V|}^{iters} \end{bmatrix} \begin{bmatrix} \Delta \varphi^{iters} \\ \Delta \left|V^{iters}\right| \end{bmatrix} \tag{6.40}$$

where, the partial derivatives form 4 submatrices, as delimited in 6.39, which are denoted as (with sizes): $J_{P\varphi}^{iters}\ (N-1\times N-1) = (N_{PV} + N_{PQ} \times N_{PQ} + N_{PQ})$, $J_{P|V|}^{iters}\ (N-1\times N_{PQ})$, $J_{Q\varphi}^{iters}\ (N_{PQ}\times N-1)$and $J_{Q|V|}^{iters}\ (N_{PQ}\times N_{PQ})$.
Then, by computing the inverse of the Jacobian, the voltage deviations are computed as:

$$\begin{bmatrix} \Delta \varphi^{iters} \\ \Delta \left|V^{iters}\right| \end{bmatrix} = \begin{bmatrix} J_{P\varphi}^{iters} & J_{P|V|}^{iters} \\ J_{Q\varphi}^{iters} & J_{Q|V|}^{iters} \end{bmatrix}^{-1} \begin{bmatrix} \Delta P^{iters} \\ \Delta Q^{iters} \end{bmatrix} \tag{6.41}$$

Thus, the new (iterative) approximation for the voltage parameters takes the following form:

$$\begin{bmatrix} \varphi^{iters+1} \\ \left|V^{iters+1}\right| \end{bmatrix} = \begin{bmatrix} \varphi^{iters} \\ \left|V^{iters}\right| \end{bmatrix} \begin{bmatrix} \Delta \varphi^{iters} \\ \Delta \left|V^{iters}\right| \end{bmatrix} \tag{6.42}$$

Lastly, a residual tolerance can be used as a stopping criterion, such as: $\begin{bmatrix} \Delta P^{iters} \\ \Delta Q^{iters} \end{bmatrix} \leq \varepsilon$.

6.3.4.3 Fast decoupled method

Considering the primary transmission EPS, and its high voltage levels, with values around $|V_i| \approx 1$ p.u.; its meshed topology; the presence of high X/R ratios [478], i.e., dominance of inductive effects on the lines, $B_{ik} \gg G_{ik}$; the relatively small angular differences between adjacent buses [531], often referred to as the power angle, usually $\varphi_i - \varphi_k \leq 10°$. A small difference that is crucial to ensure system stability [532]. Moreover, authors in [492] showed that for real-time applications involving the solution of the OPF problem, voltage phase angle difference of connected buses may be approximated by 0, under the assumption of HV or EHV operation together with a relatively low, power loss to line flow ratio. This approximation yielded a sizeable reduction in computational cost, with a faster convergence, while presenting a marginal impact on the solution accuracy. Realizing these inherent steady-state characteristics of many of the system cases and the computation burdens of the NR method for large interconnected systems, Stott and Alsaç, taking advantage of the $P\varphi - QV$ decoupling, presented a simplified version of the AC power flow, the FDPF method [533]. In other words, for adjacent buses, $P\varphi$ denotes the strong coupling between real power and the voltage angle (real power flows in the direction of decreasing voltage angles), while QV, expresses the coupling between reactive power and the voltage magnitude (reactive power flows in the direction of decreasing voltage magnitudes), a trait already noted at the end of subsection 6.2.1.

While on the contrary, there is a weak coupling between reactive power and voltage angles, and real power and voltage magnitude. For this reason, and as the name suggests, real and reactive power calculations, can be decoupled [472], which means that the off-diagonal Jacobian elements can be neglected (as consequence of the high X/R ratios), effectively meaning that $\Delta P^{iters} \not\propto \Delta \left|V^{iters}\right|$ and $\Delta Q^{iters} \not\propto \Delta \varphi^{iters}$, and so 6.40 takes the following form:

$$\begin{bmatrix} \Delta P^{iters} \\ \Delta Q^{iters} \end{bmatrix} = \begin{bmatrix} J_{P\varphi}^{iters} & 0 \\ 0 & J_{Q|V|}^{iters} \end{bmatrix} \begin{bmatrix} \Delta \varphi^{iters} \\ \Delta \left|V^{iters}\right| \end{bmatrix} \quad (6.43)$$

where in practice the decoupling translates two separate systems of equations, i.e., $\left[\Delta P^{iters}\right] = \left[J_{P\varphi}^{iters}\right] \left[\Delta \varphi^{iters}\right]$ and $\left[\Delta Q^{iters}\right] = \left[J_{Q|V|}^{iters}\right] \left[\Delta \left|V^{iters}\right|\right]$, and where the Jacobian elements, $J_{P\varphi}^{iters}$ and $J_{Q|V|}^{iters}$, while ignoring the shunt connected elements for simplicity purposes, can be derived as detailed in 6.39, considering the power expressions in 6.26, for any given iteration when $k \neq i$, as:

$$
\begin{cases} J_{P\varphi}^{iters} = \dfrac{\delta P_i^{iters}}{\delta \varphi_k^{iters}} \\ J_{Q|V|}^{iters} = \dfrac{\delta Q_i^{iters}}{\delta \left|V_k^{iters}\right|} \end{cases} \Leftrightarrow
$$

$$
\begin{cases} J_{P\varphi}^{iters} = \dfrac{\delta}{\delta \varphi_k^{iters}} \left(\displaystyle\sum_{k=1}^{N} \left|V_i^{iters}\right| \left|V_k^{iters}\right| \begin{pmatrix} G_{ik} \cos\left(\varphi_i^{iters} - \varphi_k^{iters}\right) + \\ B_{ik} \sin\left(\varphi_i^{iters} - \varphi_k^{iters}\right) \end{pmatrix} \right) \\ J_{Q|V|}^{iters} = \dfrac{\delta}{\delta \left|V_k^{iters}\right|} \left(\displaystyle\sum_{k=1}^{N} \left|V_i^{iters}\right| \left|V_k^{iters}\right| \begin{pmatrix} G_{ik} \sin\left(\varphi_i^{iters} - \varphi_k^{iters}\right) - \\ B_{ik} \cos\left(\varphi_i^{iters} - \varphi_k^{iters}\right) \end{pmatrix} \right) \end{cases} \Leftrightarrow \tag{6.44}
$$

$$
\begin{cases} J_{P\varphi}^{iters} = \left|V_i^{iters}\right| \left|V_k^{iters}\right| \left(G_{ik} \sin\left(\varphi_i^{iters} - \varphi_k^{iters}\right) - B_{ik} \cos\left(\varphi_i^{iters} - \varphi_k^{iters}\right)\right) \\ J_{Q|V|}^{iters} = \left|V_i^{iters}\right| \left(G_{ik} \sin\left(\varphi_i^{iters} - \varphi_k^{iters}\right) - B_{ik} \cos\left(\varphi_i^{iters} - \varphi_k^{iters}\right)\right) \end{cases}
$$

If we now take into account the following practical assumptions, which are mostly based on the steady-state conditions for any AC load flow study (as introduced in the beginning of this subsection (6.3.4.3), $\varphi_k^{iters} - \varphi_i^{iters} \approx 0$, hence $\sin\left(\varphi_i^{iters} - \varphi_k^{iters}\right) \approx 0$ and $\cos\left(\varphi_i^{iters} - \varphi_k^{iters}\right) \approx 1$, which leads us to $B_{ik} \cos\left(\varphi_i^{iters} - \varphi_k^{iters}\right) \gg G_{ik} \overbrace{\sin\left(\varphi_i^{iters} - \varphi_k^{iters}\right)}^{\approx 0}$; in addition, given that $J_{P\varphi}^{iters}$ differs from $J_{Q|V|}^{iters}$ only by a factor $\left|V_k^{iters}\right|$, in 6.43 we can divide the voltage deviation by its own magnitude, i.e., $\frac{\Delta\left|V^{iters}\right|}{\left|V^{iters}\right|}$, meaning that to uphold the consistency in terms of the equations, the partial derivatives with respect to this variable, need to be multiplied by the same factor, and so overall both Jacobian submatrices 6.44 can be simplified as follows:

$$
\begin{cases} J_{P\varphi}^{iters} = -\left|V_i^{iters}\right| \left|V_k^{iters}\right| B_{ik} \\ J_{Q|V|}^{iters} = \dfrac{\delta Q_i^{iters}}{\delta \left|V_k^{iters}\right|} \left|V_k^{iters}\right| = -\left|V_i^{iters}\right| \left|V_k^{iters}\right| B_{ik} \end{cases} \tag{6.45}
$$

Besides, it is important to notice that the choice of the alternative power expression, considering the polar form of the admittance 6.16, instead of 6.26, would as expected yield equal expressions for the Jacobian elements, given that the partial derivatives are:

$$
\begin{cases} \dfrac{\delta P_i^{iters}}{\delta \varphi_k^{iters}} = -\left|V_i^{iters}\right| \left|Y_{ik}\right| \left|V_k^{iters}\right| \sin\left(\theta_{ik} + \varphi_k^{iters} - \varphi_i^{iters}\right) \\ \dfrac{\delta Q_i^{iters}}{\delta \left|V_k^{iters}\right|} \left|V_k^{iters}\right| = -\left|V_i^{iters}\right| \left|Y_{ik}\right| \left|V_k^{iters}\right| \sin\left(\theta_{ik} + \varphi_k^{iters} - \varphi_i^{iters}\right) \end{cases} \tag{6.46}
$$

and so after taking into consideration the same simplifications, i.e., $\varphi_k^{iters} - \varphi_i^{iters} \approx 0$ and $|Y_{ik}| \approx B_{ik}$, which in turn imply that $\sin(\theta_{ik}) \approx 1$ (an almost purely imaginary admittance), we get the same equations as given in 6.45.

Now, following the same rationale (all the assumptions above) for the Jacobian elements,

where $k = i$, we obtain:

$$\begin{cases} J_{P\varphi}^{iters} = \dfrac{\delta P_i^{iters}}{\delta \varphi_i^{iters}} \\ J_{Q|V|}^{iters} = \dfrac{\delta Q_i^{iters}}{\delta \left|V_i^{iters}\right|} \left|V_i^{iters}\right| \end{cases} \Leftrightarrow \tag{6.47}$$

$$\begin{cases} J_{P\varphi}^{iters} = \displaystyle\sum_{\substack{k=1 \\ k\neq i}}^{N} \left|V_i^{iters}\right| \left|V_k^{iters}\right| \begin{pmatrix} -G_{ik}\sin\left(\varphi_i^{iters} - \varphi_k^{iters}\right) + \\ B_{ik}\cos\left(\varphi_i^{iters} - \varphi_k^{iters}\right) \end{pmatrix} \\ J_{Q|V|}^{iters} = \left|V_i^{iters}\right| \overbrace{\left(2\left|V_i^{iters}\right| \left(G_{ik}\sin\left(\varphi_i^{iters} - \varphi_k^{iters}\right) - B_{ik}\cos\left(\varphi_i^{iters} - \varphi_k^{iters}\right)\right)\right)}^{k=i} + \\ \qquad \left|V_i^{iters}\right| \left(\displaystyle\sum_{\substack{k=1 \\ k\neq i}}^{N} \left|V_k^{iters}\right| \left(G_{ik}\sin\left(\varphi_i^{iters} - \varphi_k^{iters}\right) - B_{ik}\cos\left(\varphi_i^{iters} - \varphi_k^{iters}\right)\right)\right) \end{cases} \Leftrightarrow$$

$$\begin{cases} J_{P\varphi}^{iters} = -\left(\displaystyle\sum_{k=1}^{N} \left|V_i^{iters}\right| \left|V_k^{iters}\right| \begin{pmatrix} G_{ik}\sin\left(\varphi_i^{iters} - \varphi_k^{iters}\right) - \\ B_{ik}\cos\left(\varphi_i^{iters} - \varphi_k^{iters}\right) \end{pmatrix}\right) - \overbrace{\left|V_i^{iters}\right|^2 B_{ii}}^{k=i} \\ J_{Q|V|}^{iters} = \overbrace{2\left|V_i^{iters}\right|^2 \left(G_{ii}\sin\left(\varphi_i^{iters} - \varphi_k^{iters}\right) - B_{ii}\cos\left(\varphi_i^{iters} - \varphi_k^{iters}\right)\right)}^{k=i} + \\ \qquad \left(\displaystyle\sum_{k=1}^{N} \left|V_i^{iters}\right| \left|V_k^{iters}\right| \begin{pmatrix} G_{ik}\sin\left(\varphi_i^{iters} - \varphi_k^{iters}\right) - \\ B_{ik}\cos\left(\varphi_i^{iters} - \varphi_k^{iters}\right) \end{pmatrix}\right) - \\ \qquad \left|V_i^{iters}\right|^2 \left(G_{ik}\sin\left(\varphi_i^{iters} - \varphi_k^{iters}\right) - B_{ik}\cos\left(\varphi_i^{iters} - \varphi_k^{iters}\right)\right) \end{cases} \Leftrightarrow$$

$$\begin{cases} J_{P\varphi}^{iters} = -Q_i^{iters} - \left|V_i^{iters}\right| B_{ii} \\ J_{Q|V|}^{iters} = -\left|V_i^{iters}\right| B_{ii} + Q_i^{iters} \end{cases} \because \left|V_i^{iters}\right|^2 \approx 1 \approx \left|V_i^{iters}\right|, \varphi_k^{iters} - \varphi_i^{iters} \approx 0 \Leftrightarrow$$

$$\begin{cases} J_{P\varphi}^{iters} \approx -\left|V_i^{iters}\right| B_{ii} \\ J_{Q|V|}^{iters} \approx -\left|V_i^{iters}\right| B_{ii} \end{cases} \quad \because \left|V_i^{iters}\right|^2 B_{ik} \gg Q_i^{iters}$$

Then, by considering the same approximation as we did in 6.47 for $\left|V_i^{iters}\right|$, for the Jacobian expressions in 6.45, i.e., $\left|V_k^{iters}\right| \approx 1$, we get the same resulting expressions for both $J_{P\varphi}^{iters}$ and $J_{Q|V|}^{iters}$. These are basically functions of the susceptance terms, typically denoted as B' and B'', respectively, and the corresponding bus voltage. As such returning to the deviation notation used in NR method, as in 6.41, we can write for FDPF method:

$$\begin{cases} \left[\dfrac{\Delta P^{iters}}{\Delta \varphi^{iters}}\right] = -\left[B'\right]\left[\left|V_i^{iters}\right|\right] \\ \left[\dfrac{\Delta Q^{iters}}{\Delta \left|V^{iters}\right|}\right] = -\left[B''\right]\left[\left|V_i^{iters}\right|\right] \end{cases} \Leftrightarrow \begin{cases} \left[\Delta\varphi^{iters}\right] = -\left[B'\right]^{-1}\left[\dfrac{\Delta P^{iters}}{\left|V_i^{iters}\right|}\right] \\ \left[\Delta\left|V^{iters}\right|\right] = -\left[B''\right]^{-1}\left[\dfrac{\Delta Q^{iters}}{\left|V_i^{iters}\right|}\right] \end{cases} \tag{6.48}$$

Resulting in a single fixed matrix, thus, not requiring the expensive iterative computation

of the partial derivatives and the Jacobian matrix inversion, to find the successive solution approximations.
Lastly, in a DC power flow context, with fixed voltage magnitudes, the presented FDPF formulation can be extended with further assumptions, which result in a simpler linear problem, with only the active power coupling, and so iterative methods are no longer needed [531], rather classical elimination and substitution methods are sufficient.

6.3.5 Penalty Method

Due to the equality and inequality constraints introduced in subsections 6.3.2 and 6.3.3, establishing a feasible region Ω with n_{dims}, the OPF problem, as we commented in the introduction section 6.1, independently of the chosen objective function(s) $OF\left(x,u\right)$, constitutes a constrained optimization problem, as given below:

$$\begin{aligned} \min \quad & OF\left(x,u\right) \\ & \text{s.t.} \\ & f(x,u)=0 \\ & g(x,u)\leq 0 \\ & x,u\in\Omega \end{aligned} \tag{6.49}$$

Therefore, as explored in greater detail in Section 4.3, solving this type of problem using for example metaheuristic algorithms, requires replacing Equation 6.49 with an unconstrained problem of the form:

$$\begin{aligned} \min \quad & OF\left(x,u\right)+c\,pen\left(x,u\right) \\ & x,u\in\mathbb{R}^{n_{dims}} \end{aligned} \tag{6.50}$$

where c is a positive (sufficiently penalty parameter and $pen\left(x,u\right)$ is a continuous penalty function that maps $\mathbb{R}^{n_{dims}}\rightarrow\mathbb{R}$; with $pen\left(x,u\right)=0$ if $x,u\in\Omega$, while for $x,u\in\mathbb{R}\backslash\Omega$, $pen\left(x,u\right)>0$. In other words, in case any of the control and state variables lie outside the feasible search space region, violating any of the n_{eqc} equality constraints, $f(x,u)=0$, and/or any of the n_{ineqc}, inequality constraints, $g(x,u)\leq 0$, several quadratic penalty terms are added to the already defined objective function, resulting in a sequence of unconstrained sub-problems 6.51 [534], that should be minimized (together) by the chosen optimization algorithm.

$$\min \quad OF\left(x,u\right)+\overbrace{\frac{c}{2}\left(\sum_{i=1}^{n_{eqc}}f(x,u)^{2}+\sum_{i=1}^{n_{ineqc}}\max\left\{0,g(x,u)^{2}\right\}\right)}^{c\;pen(x,u)} \tag{6.51}$$

6.4 OPF: A Test Case

The chosen OPF test case is the standard IEEE-30 bus system, as shown in the single-line diagram in Figure 6.7, comprised of 41 branches (connection between nodes); 6 thermal generators connected to PV buses 1, 2, 5, 8, 11, and 13; 24 PQ buses; 4 transformers; and 9 compensators.

For benchmarking purposes, and drawing on the knowledge acquired in transforming constrained optimization problems into unconstrained ones—as well as the versatility of metaheuristics in addressing multidimensional, non-convex optimization problems (see chapter 4) — a modified Harris Hawks Optimization (HHO) algorithm is proposed and compared with the original HHO algorithm [535]. For a more detailed explanation of the algorithm's rationale, search space mechanisms, proposed modifications, and their motivations, please refer to Appendix B.

In terms of the optimization process, the objective is to find the optimal value for the control variables, u to minimize the objective function 6.19 while fulfilling all the system constraints. The population size and the maximum number of iterations are set to 30 and 300, respectively. For a fair comparison and to ensure meaningful statistical results (overcoming possible initial positioning bias), a total of 33 runs is executed, with the same positioning (starting hawk population) being given to both the original and modified HHO for each run.

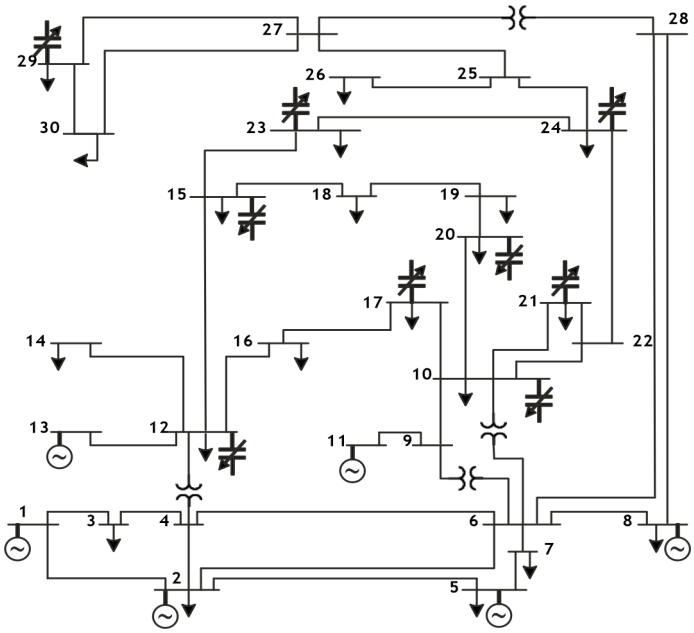


Figure 6.7: IEEE 30-bus system (test case).

In terms of the inequality constraints, and besides the active power that follows the original IEEE 30-bus system bounds (specific to each generator), the generator and bus load voltages are limited in per units between 0.95 and 1.1. The transformer tap ratios are bounded between 0.9 and 1.1 (per units); while the VAR source capacity of the compensators ranges between 0 and 5 MVAr. A summary of the case-study (IEEE-30 bus system)

main parameters, as well as the generator's quadratic fuel cost coefficients is given in Table 6.2 (as given in [536]).

Table 6.2: IEEE 30-Bus System: Main Parameters and Fuel Cost Constants.

Case-study parameters	Number of branches	41		
	N	30		
	N_{PV}	6		
	N_{PQ}	24		
	N_T	4		
	N_C	9		
	P_{Ld} (Total MW)	283.4		
	Q_{Ld} (Total MVAr)	126.2		
Generator identifier	***Bus (connected to)***	***Fuel Cost Constants* (??)**		
		a_i	b_i	c_i
G1 (Swing)	1	0	2	0.00375
G2 (PV)	2	0	1.75	0.0175
G5 (PV)	5	0	1	0.0625
G8 (PV)	8	0	3.25	0.00834
G11 (PV)	11	0	3	0.025
G13 (PV)	13	0	3	0.025

The optimization is executed on MATLAB® software version 2020a, in a computer with the following specifications: CPU running at 2.5GHz and with 8 GB of RAM. The well-known MATPOWER package, known for its implementation of steady-state AC OPF problems [537], is used not as a solver of the OPF per se but as a framework from which the control variables and operational limits (inequality constraints) are imposed, by (constantly) loading and adjusting the case study matrices (assembling the power flow model). Then, for each tested set of control variables, u, or simply put each hawk's position, X_i, the dependent variables (non-linear system of equations), x, that translate the low hierarchy load flow problem are calculated via NR's method; followed by the branch power flows. Then, the feasibility of the solutions is verified, including the possible violation of any constraint and the need for the respective penalty term, which if it is the case, is then added to the fuel cost value (computed based on all the active power outputs). The benchmark results revealed a superior performance of the mHHO on all the considered statistical metrics, with a mean value of 800.24 $/h that compares with 812.73 $/h for the HHO, a standard deviation of 0.43 $/h versus 1.73 $/h, and an average computation time of 67s that compares with 70s.

To better visualize the improved performance of the mHHO in terms of the objective function, an area plot (Figure 6.8) was constructed which comprises the convergence curves data of both algorithms, i.e., the absolute minimum and maximum values of each iteration for all the runs. While, the mean convergence curves are plotted in black. Notice, that none of these plots, accurately reflects a real convergence curve, but they are used to better illustrate the convergence behavior of the benchmarked algorithms, which as one can notice, after around a dozen to fifty iterations, mHHO refined search space mechanisms, are key to achieve a more cost-efficient operation of the studied power system.

As for the best run (in terms of objective value), the obtained control variables and the to-

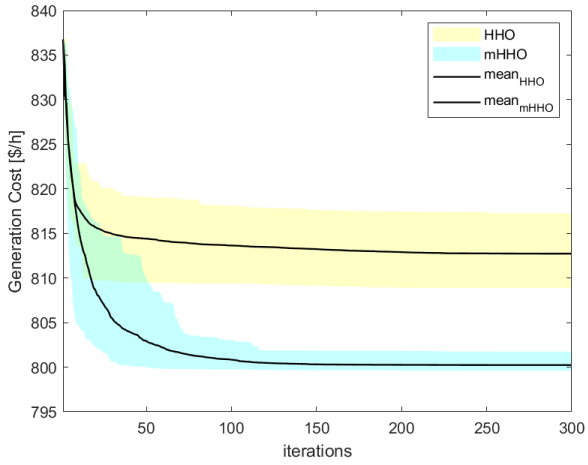


Figure 6.8: Convergence curve limits and mean value: HHO vs mHHO.

tal generation cost for both algorithms (reached in different runs) are shown in Table. 6.3. We can see, that as is the case of all the other runs, the algorithms have achieved feasible solutions, with the mHHO even being able to operate below the 800 $/h, which is generally considered a good indicator of the algorithm's ability to solve this specific OPF problem.

Table 6.3: Control variables bounds, optimized values and total cost for the best run.

Control parameters	**Bounds**		**Optimization Results**	
	Min	***Max***	***HHO***	***mHHO***
P1(MW)	50	200	177.35	177.00
P2(MW)	20	80	30.22	48.44
P5(MW)	15	50	22.60	21.91
P8(MW)	10	35	26.24	20.37
P11(MW)	10	30	14.80	12.42
P13(MW)	12	40	20.48	12.00
V1 (p.u)	0.95	1.10	1.10	1.10
V2 (p.u)	0.95	1.10	1.09	1.09
V5 (p.u)	0.95	1.10	1.07	1.06
V8 (p.u)	0.95	1.10	1.07	1.07
V11(p.u)	0.95	1.10	1.06	1.07
V13(p.u)	0.95	1.10	1.05	1.09
T11 (p.u)	0.90	1.10	1.02	1.03
T12 (p.u)	0.90	1.10	1.05	0.96
T15 (p.u)	0.90	1.10	1.04	1.10
T36 (p.u)	0.90	1.10	1.04	1.03
QC10 (MVAR)	0	5	1.43	0.10
QC12 (MVAR)	0	5	3.70	2.67
QC15 (MVAR)	0	5	2.42	3.68
QC17 (MVAR)	0	5	1.52	4.93
QC20 (MVAR)	0	5	1.25	0.33
QC21 (MVAR)	0	5	1.50	3.53
QC23 (MVAR)	0	5	1.13	4.32×10^{-4}
QC24 (MVAR)	0	5	1.35	1.97
QC29 (MVAR)	0	5	2.34	4.98
Total cost ($/h)			808.86	799.61

Chapter 7

Forecasting in the Electric Power Systems

Building upon the extensive foundations introduced in the previous chapters on optimization (chapter 4) and artificial neural networks as regression models (chapter 5), this chapter shifts gears toward a critical application area: forecasting in Electrical Power Systems (EPSs). Forecasting remains a cornerstone task in EPSs, essential for ensuring reliability, efficiency, and economic viability. In recent years, its importance has only intensified due to a confluence of both longstanding and emerging challenges, as well as increasingly dynamic system behavior. These include more decentralized and variable load profiles—driven by demand response mechanisms, self-consumption, and the rise of microgrids — greater variability in the energy generation mix, the expansion of market participants, and the integration of distributed energy resources.

Additionally, as discussed in Chapter 2, the anticipated growth in electric vehicle (EV) adoption introduces both flexibility and complexity. If not managed properly and incrementally, EV integration could overburden existing EPS infrastructure and further complicate forecasting tasks.

This chapter explores key contributions by the author in the areas of load, price, and generation forecasting, with a particular emphasis on leveraging machine learning tools—especially artificial neural networks (ANNs) — as forecasting models. It also addresses the practical intricacies of input data preparation and model fine-tuning, both of which are essential for handling the inherent complexity of EPS data and enabling more adaptive, data-driven decision-making in these critical forecasting tasks.

7.1 Introduction

The large-scale and complex nature of Power Systems Operation in a constantly changing environment poses a series of challenges for policy makers, regulators, market operators and participants (both the generation and demand sides) and transmission and distribution operators, among others [538]. Some key operational challenges — closely linked to the structure of electricity markets — include electrical power load forecasting, electricity price forecasting, and power generation forecasting tasks.

These are, in turn, crucial for the smooth operation of the power grid—which runs on a set of tight tolerance requirements—and the decision-making policies of utility companies, i.e., influencing the purchase and generation of electric power in the power market and consequently affecting operation costs, energy efficiency, load switching and infrastructural development [539]. In this regard, the increased level of variable renewable energy sources directly rises the required spinning reserve volume to offset the fluctuation of renewable generation [540]. Factors such as an expanded network of distributed energy

resources and a higher share of demand-side management further add to the complexity of this prediction task [541], which is already influenced by several uncertain factors including climate and seasonal factors [542, 543].
Given that a large part of market decisions and utilities operations have an hourly, daily or weekly basis, these types of forecasting lead times (from 1 h to 168 h) are of most interest and are known as ”short-term” forecasts.

7.2 Load Forecasting

Short-Term Load Forecasting is critical for reliable power system operation, and the search for enhanced methodologies has been a constant field of investigation, particularly in an increasingly competitive environment where the market operator and its participants need to better inform their decisions. Hence, it is important to continue advancing in terms of forecasting accuracy and consistency.
An interval that is well categorized in the electric power system literature is the Short-Term Load Forecast (STLF) [544], which has long been a very active field of research, and which is the object of renewed interest given the advent of smart grids and microgrids [545].
This has led researchers to continuously follow and adapt the developments in data filtering, clustering and modeling in order to propose load forecasting methodologies able to support the decision-making process regarding crucial operation, maintenance and planning tasks. Typically, an underestimation of load demand results in unreliable supply, with penalty costs or even shortage of supply to customers, while an overestimation causes economic waste and results in curtailment periods [546].
Various methodologies have been used for the STLF problem, with the two main groups being statistical methods and machine learning methods. The former comprises methods specially tailored to handle the inherent linear correlation between the load and its lagged values, as well as weather and temporal variables. Among the many statistical methods, one can highlight the well-regarded autoregressive integrated moving average (ARIMA) models, exponential smoothing and generalized autoregressive conditional hetero-scedasticity (GARCH) models, among others [547].
An alternative path that has been widely adopted in recent years is using ML and deep learning DL given their ability to extrapolate and learn hard non-linear relationships through complex and dynamic structures [548], which are exactly the difficult aspects for statistical methods. Among the broad spectrum of methods, recurring options include ANNs [547] and MLPs, random forests [548], fuzzy systems [549], SVMs [550], RNNs [551] and its modified counterparts, LSTMs [438, 552] and GRUs [553], ELM [554], DBNs [555], echo state networks [556], CNNs [557] and more recently deep stack of fully connected layers connected with forward and backward residual links (N-BEATS) [558].
In addition to the issue of the chosen forecaster (prediction method), critical aspects in STLF include selecting the proper input data and pre-processing (i.e., deciding which sets

of variables to include in the model and its treatment). To illustrate this point, the authors in [559] analyzed the impact of the selection of exogenous variables on an ANFIS forecaster's prediction of load in the Polish power system (using the Hellwig method). In turn, the authors in [560] reinforced the importance of data pre-processing methods, specifically by employing the Variational Mode Decomposition (VMD) method. Other common options include empirical mode decomposition [549], wavelet transform [547] and principal component analysis among others.

A common approach in recent studies is to employ hybrid or ensemble approaches that combine several methods, extracting the best prediction accuracy from each one by primarily merging methods that model both linear and nonlinear components of the time-series [561]. For instance, a combined MLP-ANN with ANFIS was used for simultaneous load and price forecasting in [562]. In [563], a hybrid method was presented, where an ARIMA model was combined with SVMs. This generic approach of hybridizing an ARIMA-ANN method has been validated for time-series forecasting in [564].

Recognizing the growing role of ML methods with respect to solving the STLF task, a fusion algorithm was presented in [565] where SVM, Random Forest and LSTM forecasters were used to produce an enhanced prediction. Similarly, an ensemble approach was published in [566] using VMD to decompose the load data and 2D-CNN to extract the features that fed the Temporal Convolutional Network working as forecaster. Using another ensemble concept, a method combining three models, namely, Light Gradient Boosting Machine, eXtreme Gradient Boosting machine (XGB) and MLP was presented in [567]. Lastly, a recent approach made use of stacked boosters (forecasters) with different horizons coupled with a deep neural network to perform the STLF [568].

Given these works and their merits, an ensemble approach will be introduced in this paper where a new enhanced method to accurately forecast the one-day ahead loads of national or sizable regional (as is the case of New England in the US) electric grids is proposed. We address the need to further develop data-driven forecasting techniques that are compact in terms of user-dependence and computationally cost-efficient. To fulfill this objective, we chose an ensemble method based on a deep neural network, which due to its scalable and non-linearity traits is a great tool for enhancing feature extraction and better mapping the input space [569]. Thus, an ensemble approach has the potential to improve generalization and increase model stability (less output variance). In turn, different ARIMA models contribute to increasing input diversity through a stacked generalization framework. The straightforward formulation and the relatively small number of hyperparameters favors its robustness to model time-series based on the auto-correlation between past samples (only requires prior data), making it one of the preferred methods to benchmark proposed STLF models.

7.2.1 Theoretical Background

This sub-section introduces the theoretical fundamentals behind each stage and building block considered in the proposed methodology.

7.2.1.1 Box-Jenkins Method

To study and model a generic time-series data X with a total of n_{length} sequentially spaced time observations, such as $X = (\chi_1, \chi_2, ..., \chi_{n_t-1}, \chi_{n_t})$, a simple and common choice is to use the Box-Jenkins method to identify and fit the most suitable Autoregressive Integrated Moving Average (ARIMA) time-series models. These statistical models constitute a good approach to tackle the forecasting task, as mentioned in section 7.3.

These models translate a linear relationship with the observed output data and rely on the assumption that future observations can be approximated based on past observations. Models that can be written using a single equation as expressed in Equation 7.1, with an autoregressive term, AR(p), where p is the autoregressive polynomial order and a moving average term, MA(q), where q is the moving average polynomial order, hence translating a linear combination of the preceding error terms up to lag q. Thus, the polynomial orders or degrees, $p, q \in \mathbb{Z}^+$, represent the maximum value for the AR and MA terms, respectively. In turn, the integrated part is commonly represented by the letter d, which stands for the differencing order. In the case of $d = 0$, the ARIMA model is simply an ARMA model [570]:

$$\chi_t = \mu + \varepsilon_t + \sum_{i=1}^{p} \varphi_i \, \chi_{t-i} + \sum_{j=1}^{q} \theta_i \, \varepsilon_{t-j}, \quad (7.1)$$

where μ denotes the constant term (vertical translation), χ_{t-i} denotes the lagged (past) time-series (independent variable) and ε_{t-j} represents the unknown lagged error terms, i.e., the random error term that accounts for the difference between the actual observation and the theoretical value given by the regression-based model, often referred as white Gaussian noise. Therefore,the actual observation, χ_t, is influenced by factors beyond the independent variable. In turn, the autoregressive and moving average coefficients are denoted by φ_i and θ_i, respectively. Typically, when using the Box–Jenkins notation, a backshift operator, B, is used to represent the lagged terms such that $B\,\chi_t = \chi_{t-1}$. Thus, the model in Equation 7.1 is re-written as demonstrated in Equation 7.2.

$$\begin{gathered} \varphi_p\,(B)\,\chi_t = \mu + \theta_q\,(B)\,\varepsilon_t \Leftrightarrow \\ (1 - \varphi_1 B - .... - \varphi_p B^p)\;\chi_t = \mu + (1 - \theta_1 B - .... - \theta_q B^q)\,\varepsilon_t \end{gathered}, \quad (7.2)$$

In addition, most of the time-series encountered in engineering problems, including traditional energy load forecasting, reveal non-stationary characteristics, meaning that the unconditional joint probability distribution changes with time (a non-constant mean and variance across observations). These traits in the time-series reveal the presence of trends and/or seasonal variations. Consequently, since the Box-Jenkins models assume stationarity, one approach is to obtain the first differences of the time-series, i.e., $\chi_t - \chi_{t-1} = (1 - B)\,\chi_t$, which results in a general version of the ARIMA(p, d, q) model with an added

generic difference term, as shown in Equation 7.3.

$$\begin{gathered}\varphi_p(B)(1-B)^d\chi_t = \mu + \theta_q(B)\varepsilon_t \Leftrightarrow \\ (1-\varphi_1 B - .... - \varphi_p B^p)\ (\varphi_1 B + ... + \varphi_p B^p)^d \chi_t = \mu + (1-\theta_1 B - .... - \theta_q B^q)\varepsilon_t\end{gathered}, \quad (7.3)$$

To determine the order at which the original time-series must be integrated to acquire stationarity, authors often resort to unit root tests to assess the stationarity hypothesis. For the vast range of time-series, when the unit root test determines that the data needs to be integrated, due to variance in the time-series, the first and second differences are usually ruled sufficient for removing the trends [547]. In this scenario, where the time-series reveals clear features of seasonality, e.g., daily, monthly and semesterly cycles, among others, it is advantageous to use a seasonal ARIMA$(p,d,q)\times(P,D,Q)$ (SARIMA) model [571]. The same letters but capitalized, P, D and Q, describe the seasonal orders of the autoregressive, moving average and differentiation terms; conversely, the lower-case letters represent the non-seasonal term orders. As such, the generic abbreviated form of the multiplicative model can be written as follows Equation 7.4:

$$\varphi_p(B)\,\Phi_P(B^s)(1-B)^d(1-B^s)^D\chi_t = \mu + \theta_q(B)\,\Theta_Q(B^s)\varepsilon_t, \quad (7.4)$$

where the superscript s distinguishes the seasonal terms from the non-seasonal, whereas the seasonal autoregressive and moving average coefficients are denoted by Φ_P and Θ_Q, respectively. All the remaining parameters were already defined in Equations 7.1 and 7.2. These statistical models are especially tailored to handle linear problems, but the same cannot be said about hard non-linear problems, i.e., they do not follow the normal distribution assumption [572].

With the groundwork notation introduced, the next step concerns the question of how to identify ARIMA(p,d,q) order levels. To this end, the typical choice is to apply the well known Box and Jenkins method, which relies on the auto-correlation function (ACF), a measure of the mutual dependence between distinct lags, and the partial auto-correlation function (PACF) to draw inferences about an adequate model [561]. This last function is a measure of the relationship between the current observation and its lags excluding the effects of intermediate observations, which is useful to specify AR terms.

7.2.1.2 Deep Neural Networks

A Deep Neural Network (DNN), as we saw in 5 is a type of machine learning approach, with the particularity of having multiple (hidden) layers between the input and output layers. The term ”deep” is used to effectively convey the more complex structure of layers and nodes, which significantly increases the number of weights and bias terms, deriving a more abstract and high-dimensional feature mapping from the provided input and output data [573, 574]. In the same fashion as its classic shallow ANN counterpart, it draws

inspiration from human brain functioning, particularly the information flow in biological neurons [575]. Moreover, its ability to cope with non-linearities renders it a suitable tool to handle a vast range of engineering problems from pattern recognition and data classification to time-series modeling, etc.

A common architecture for the DNN is the Deep Feedforward Neural Network (DFNN), forming a directed acyclic graph as represented in Figure 7.1 meaning that the flow of information only proceeds forward, simplifying the backpropagation training process. In each DFNN layer, every neuron output is evaluated by the activation function f_j, with a cost given by a biased weighted sum a_j, which works as a threshold. Hence, one can state that for any two consecutive layers $l-1$ and l, the neuron (node) response can be formulated as follows:

$$y_j = f_j\left(a_j\right) = f_j\left(\sum_{i=1}^{n} \omega_{ij}\, x_i + b_j\right), \quad (7.5)$$

where $j \in]0, m]$ is an index representing an arbitrary neuron response in layer l; and the $i \in]0, n]$ index denotes an arbitrary input node from the previous layer $l-1$ (notice the flexibility of this formulation where the previous layer can either be an input layer or an adjacent hidden layer). Accordingly, n is the total number of neurons/input nodes in layer $l-1$, and m is the number of neurons in the connecting layer l, with n and $m \in \mathbb{Z}^+$. y_j is the output for neuron j; x_i is the input signal from neuron i; b_j is the respective neuron bias; and lastly ω_{ij} is the weight of the synaptic connection between neurons i and j. This output response (Equation 7.5) is then one of the inputs for all the nodes of the proceeding layer in sequential form. Thus, one can compute the output response of all of the hidden layers in a DFNN, with an arbitrary number of L hidden layers, as a composite function g of the original input x (input layer), as denoted in Equation 7.6 [576]:

$$g\left(x\right) = g\left[a_{(L+1)}\left(h_{(L)}\left(a_{(L)}\left(\ldots\left(h_{(2)}\left(a_{(2)}\left(h_{(1)}\left(a_{(1)}\left(x\right)\right)\right)\right)\right)\right)\right)\right)\right], \quad (7.6)$$

where $h^{(l)}\left(x\right)$ denotes the output of an arbitrary hidden layer l (superscript), and $a^{(l)}\left(x\right)$ is an abbreviated form for the biased weighted sum for an arbitrary hidden layer l (pre-activation cost).

In the considered training process, the DFNN weights are updated by using an offline supervised strategy, where the weights are adjusted to minimize the error in the output layer with known target samples. The scaled conjugate gradient backpropagation algorithm was chosen to perform the training task given its balance between performance and speed, which is particularly important for deep networks.

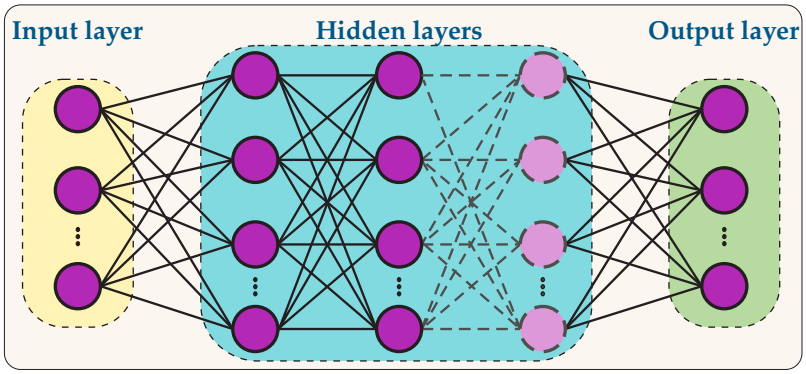


Figure 7.1: Deep Forward Neural Network (DFNN).

7.2.2 Proposed Methodology

This sub-section presents a comprehensive description of the proposed load forecasting methodology and discusses its implementation details. First, the proposed methodology is guided by the ensemble methods (learning) concept, where an improved predictive performance is expected when employing/combining multiple methods versus any of the individual methods. More specifically, this work is centered on a stacking ensemble type approach, which in a broad sense involves training a machine learning method to merge the predictions of some other methods [577]. To reinforce this point, the literature in the forecasting field documents report that when dealing with an unstable and varying data pattern, it is convenient to use dissimilar models to improve time-series forecasting accuracy [578]. Much of the reasoning behind the wide-adoption (as shown the Introduction section) of ensemble methods resides in the increased diversity among the base classifiers/forecasters. When using a stacking strategy, the meta-learning model (assembler) is used to integrate the output of base models [579].

Following this line of thought, a stacking ensemble approach is proposed in this work where a DFNN combines of the Box-Jenkins predictors. For better context and for learning purposes, additional input data from the load time-series and relevant related exogenous variables are also fed to the DFNN. An overview of the most relevant stages and main blocks that constitute the proposed methodology is given in Figure 7.2. A more thorough account of these stages and blocks is provided in the lower level sub-sections below. Additionally, the pseudo code of the main implementation stages is given in Appendix C.

7.2.2.1 Prediction Models: ARIMA Forecasters

Given this ensemble approach to perform STLF, ARIMA models were chosen to serve as reference inputs for the combiner DFNN. This approach of coupling soft computing methods with ARIMA model has proven to increase accuracy, as reported in [572, 578]. While ARMA models are unable to capture the non-linearity of the series, they can capture the linear traits of the time-series, which is worthy information to be fed as inputs to the assembler (non-linear) DFNN [580].

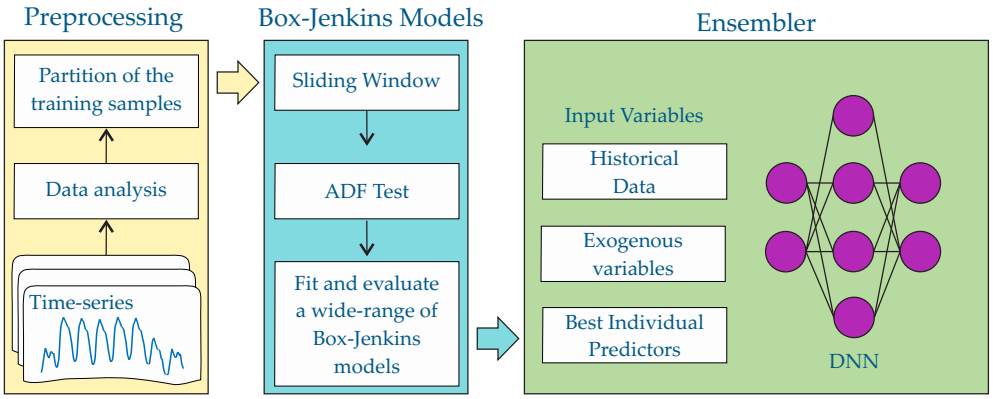


Figure 7.2: Proposed methodology: overview.

Before identifying and fitting the ARIMA models, a significant question concerns the suitable number of observations to estimate the models. In this regard, a common-sense rule is to have more observations than model parameters. This implies that, for example, in a SARIMA model as defined in Equation 7.4, a number of $p+d+q+P+D+Q+1$ observations is required to estimate the model. However, when dealing with data exhibiting substantial random variation, this minimum requirement is insufficient, and observations must exceed the parameters several fold [581]. Hence, a hard rule can often be misleading given the underlying variability of the data. As such, instead of following a narrow approach, in this work, a conscious decision was made to assemble a pool of ARIMA models.

These differed from each other not only in parameters but also in the training sample window (number of observations used to fit the model). This procedure results in very different ARIMA models, where some will better capture the most recent trends (when the training window is shorter), while others with a longer training window (more samples) will provide a fuller picture of the past time-series behavior. Hence, in this work, five different training windows (in months) were selected, i.e., the number of past observations (training samples) used to fit the model was given by a specific number of prior months defined in the variable $\delta = [1, 2, 3, 6, 12]$.

For each selected training window, stationarity was evaluated through the Unit Root hypothesis using the Augmented Dickey Fuller (ADF) test. Based on the test's p-value, the methodology decides the order of integration. In our case study, the first differences proved to be sufficient, but with a 12-month training window, in some instances, the unit-root null hypothesis was rejected in favor of the alternative hypothesis (ARMA model is suitable). Having sorted the integration order d, the ACF and PACF functions were built, proceeding far as 15 days prior, and a list of the most promising AR and MA terms was established. With this task completed, all the parameter combinations were tested, and the respective ARIMA models were assembled together with their mirroring SARIMA models, where the seasonal P and Q orders depict daily and weekly seasonalities. This process of automatic model selection, based on the information provided by Unit Root Tests and correlation analysis, is frequently referred as "Automatic ARIMA" [582].

Having identified the several models identified, a Maximum likelihood estimation (MLE)

algorithm was used to estimate the model coefficients, i.e., the coefficients that maximize the log-likelihood function. Subsequently, the Bayesian Information Criterion (BIC) was used to select the three best models for each training window (for a total of 15 ARIMA forecasters, by determining the set of model parameters $[\varphi_p, \Phi_P, \mu, \theta_q, \Theta_q]$ that minimize the BIC value, as expressed in Equation 7.7:

$$BIC = k \ln(n_{length}) - 2 \ln\left(\frac{RSS}{n_{length}}\right), \quad (7.7)$$

where k denotes the number of model parameters, n_{length} is the number of observations and RSS is the residual sum of square, i.e., a measure of the variance in the error term of the model [583]. Notice that while all combinations were tested, this does not imply that if, for example, $p = 24$, all AR terms between 1 and p have non-null coefficients. On the contrary, since the model complexity is penalized, as observed in Equation 7.7, this reinforces the need to select models with strong correlations.

This offline process is performed iteratively in an automatic manner, and its different stages are illustrated in Figure 7.3. The selected training window is only valid for testing in the subsequent week, and the forecasting performance is stored upon test completion. Thereafter, the most recent data (testing data) are added to the training window (batch), while the oldest data are removed, similarly to a Last In First Out procedure. Therefore, it ensures a training window (batch) with a fixed size. This windowing process is carried out until all testing data are evaluated and the respective ARIMA Forecasters are stored. The process concludes after analyzing all the training windows and their correspondent prediction models, producing the key prediction inputs for the ensemble method.

7.2.2.2 Ensembler DFNN

As previously discussed in section 7.2 and sub-section 7.2.1.2, DNNs are a great solution for solving a wide range of problems given their ability to explore and model deep nonlinear relationships, as such encompassing prediction tasks as well.

With respect to the DFNN architecture, its input layer was built with 26 nodes that account for the following: 15 ARIMA forecasters (three best forecasts for each of the five training windows), $\{\hat{y1}_t{}^{\delta_1}, ..., \hat{y1}_t{}^{\delta_5}, \hat{y2}_t{}^{\delta_1}, ..., \hat{y2}_t{}^{\delta_5}, \hat{y3}_t{}^{\delta_1}, ..., \hat{y3}_t{}^{\delta_5}\}$, which are stacked and fed to the ensemble method providing information about the expected linear trends; the six past lagged values of the actual load data $(x_{t-24}, x_{t-25}, x_{t-26}, x_{t-32}, x_{t-48}, x_{t-168})$; four date/time related exogenous variables, namely hour, weekday, binary holiday indicator and season of the year; information about the time-sample of the 24 h ahead target load $(exog_t{}^{hour}, exog_t{}^{weekday}, exog_t{}^{holiday}, exog_t{}^{season})$; and finally, the lagged relative humidity information was also fed to the DFNN $(exog_{t-24}{}^{relHum})$. The addition of exogenous variables to forecast the load in the ensemble method is of crucial importance since electricity demand does not depend exclusively on the auto-correlation of its own data (endogenous). In fact, a comprehensive assortment of exogenous factors has an impact on the load [580].

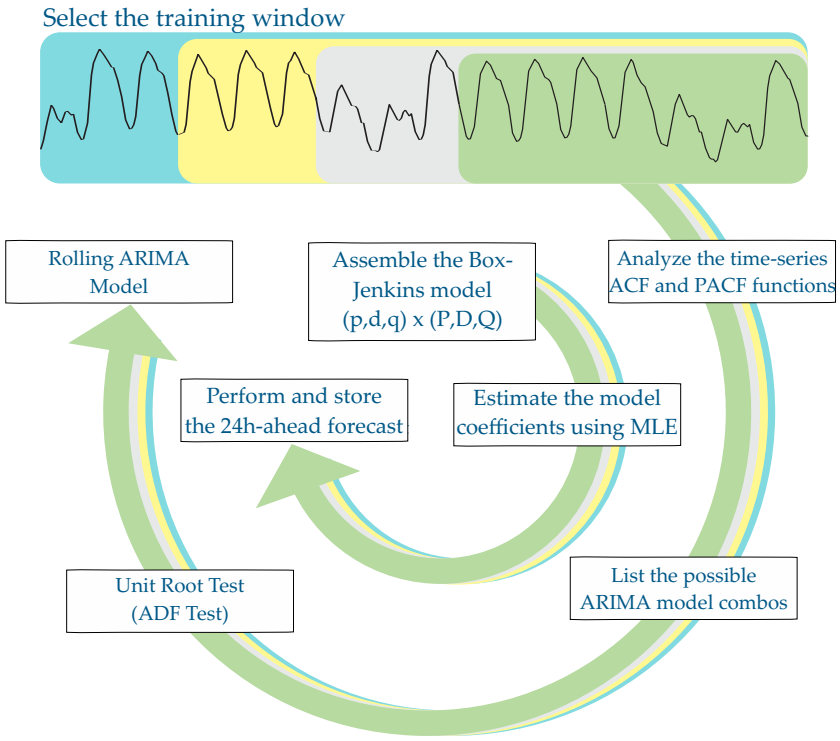


Figure 7.3: Predictions models: outline of the different stages from data selection to ARIMA Forecasters.

Concerning the hidden layer(s) configuration and after performing a few trial simulations, it was found that a three hidden layer DFNN (meaning a total of five layers, including the output and input layers), with 20, 10 and 5 neurons in each hidden layer, respectively, was found to be a good fit. These hidden sizes ensured a good level of complexity and abstraction without using too many neurons and, consequently, weights, which can often result in overfitting. In this regard, note that an optimal paradigm for the definition of the best model does not exist and likewise for the number of neurons in the hidden layer(s). As such, the conventional path is to rely on past experiences and trial-and-error procedures.

The hyperbolic tangent sigmoid transfer function, $\sigma\left(x\right) = \frac{2}{1+\exp(-2x)} - 1$, was chosen as the activation function for the hidden layers. The DFNN flow ends in the output layer, with a single neuron, producing the 24 h ahead forecasting, $\hat{y}_t$. In addition, a linear transfer function, $out(x) = \alpha x$, was used in this layer, which is a very common choice for approximation and regression tasks.

With the architecture defined, the next stage involves pre-processing (Equation 7.8) the input and target data, ensuring that the data always fell within the range $[-1, 1]$ by using the expression below:

$$x_{norm} = -1 + 2\left(\frac{x - x_{\min}}{x_{\max} - x_{\min}}\right), \quad (7.8)$$

All the training samples are employed to form the learning batch; as such, a Batch Gradi-

ent Descent paradigm is considered in this work. This implies that all samples contribute to training and validation errors before updating DFNN weights and biases. The input and target data spanned over the period of 1 year prior to the testing of the trained ensemble methodology in the subsequent month. These datasets were divided randomly into training set and validation set by using a typical 70% to 30% ratio in order to improve the generalization capabilities.

This validation set provides a separate (unseen) set of representative data-samples that are used to perform an unbiased evaluation of the model fit (validation error), different from the training error, during the training phase (weights and biases update). When the validation error starts to decouple from the training error, i.e., does not decrease for several iterations, further continuation of the training process may invariably result in overfitting. This decoupling, therefore, constituted the main stopping criterion for DNN training even though a maximum number of learning iterations was also defined. Next, the essential training process of the ensemble DFNN was carried out. The scaled conjugate gradient (SCG) algorithm [584] was chosen to perform this supervised learning task given that no critical user-dependent parameters are needed, making it a fully automated algorithm. In addition, its step size scaling mechanism improves the speed of convergence by avoiding a time-consuming line search per iteration, which is a common calculation of other second order algorithms [547, 585].

Furthermore, in order to improve generalization during the training process, a modified version was given by Equation 7.9, $\mathrm{modPerfF}$, from the standard performance function using the mean squared error (MSE), and it was chosen by adding the mean of the sum of squares of the network weights and biases (MSW). This forces the network to have smaller weights (as we saw in sub-section 5.3.2.2):

$$\mathrm{modPerfF} = (1-\eta)\,\mathrm{MSE} + \eta\,\mathrm{MSW} = (1-\eta)\,\frac{\sum_{i=1}^{N_S}(y_i-\hat{y}_i)^2}{N_S} + \eta\frac{\sum_{j=1}^{N_W} w_j^2}{N_W}, \quad (7.9)$$

where η is the generalization ratio (a good compromise was achieved with its value set to 0.3), N_S is the number of training samples, N_w is the number of weights in the network, $\hat{y}_i$ is the DFNN output and y_i is the desired (target) output response.

7.2.3 Case Study: New England Load

In order to evaluate the proposed methodology's forecasting accuracy, the well known real case study from the New England region system is considered. This region in Northeastern US spreads across six states, namely Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island and Vermont. The hourly system load samples are provided by the New England Independent System Operator (ISO-NE) and are available in [586]. In this work, data from 2014 to 2015 were selected for training different forecasters. The data from 2016 were applied to test the models, and Figure 7.4 illustrates its hourly and monthly behavior.

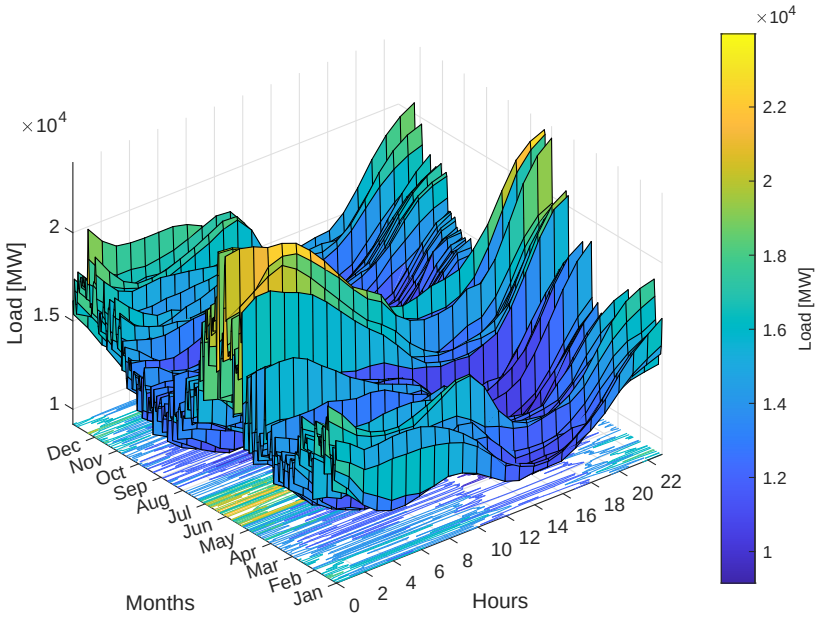


Figure 7.4: New England region (US) — Electrical Load (MW) 2016.

A brief data analysis revealed a mean load value of 13.504 GW and significant amount of standard deviation, 2.426 GW (∼18% of the mean value), with a moderately positively skewed and Leptokurtic distribution (skewness = 0.779 and kurtosis = 3.870). This implies a distribution with longer tail and more observations located to the right (higher) of the mean value. In addition, as illustrated in Figure 7.4, the summer months registered not only higher load values but also greater variabilities largely due to the cooling related loads. By contrast, the spring months had a smaller mean and standard deviation. Accordingly, the month of June had the highest mean and standard values, 16.700 GW and 2.9269 GW, respectively, as well as the maximum recorded load value of 23.973 GW, which took place on 3 June. The month of August recorded the lowest mean value of 11.480 GW and also the minimum observed load value of 9.149 GW (7 August), while the minimum standard deviation occurred in the spring month of April with a value of 1.3383 GW.

7.2.3.1 Day-ahead Load Forecasting Results

In order to demonstrate the proposed methodology's behavior in a representative manner for the entire year, four different scenarios (out-of-sample scenarios) were considered, comprising four months corresponding to four different meteorological seasons, including holidays (typically similar to weekends in terms of load profile).The following months were selected to serve as testing data from the New England region: (i) winter month (February 2016); (ii) spring month (April 2016); (iii) summer month (August 2016); and (iv) fall month (October 2016). In order to ensure the consistency, the results of the ensemble DFNN were the mean results of chosen error metrics over 50 simulations, thereby avoiding skewed inferences based on specific random data divisions and random DFNN

weight initialization.

In order to gauge the forecasting accuracy of the 24 ahead predictions, the common scale-invariant error metrics MAPE(%) (mean absolute percentage error) and RMSE (MW) (root mean square error) were used, respectively, Equations 7.10) and 7.11, which are defined as follows:

$$MAPE\ (\%) = \frac{100}{N_{test}} \sum_{i=1}^{N_{test}} \left| \frac{load_i - \hat{load_i}}{load_i} \right|, \quad (7.10)$$

$$RMSE = \sqrt[2]{\frac{100}{N_{test}} \sum_{i=1}^{N_{test}} \left(load_i - \hat{load_i} \right)^2}, \quad (7.11)$$

where N_{test} is the number of testing samples (length of the target and forecast vector); and $load_i$ and $\hat{load_i}$ are the actual and predicted load value at each time-step i, respectively.

The obtained load forecasting error metrics are displayed in Tables 7.1 and 7.2. Notably, in terms of MAPE (%) (Table 7.1), the ensemble method was able to outperform the ARIMA forecasters (input Predictions Models) for all the testing periods. This translates the sizeable monthly accuracy gains of MAPE (%) in comparison with the mean accuracy of all the ARIMA forecasters of ∼23%, ∼14%, ∼12% and ∼26% for the months of February, April, August and October, respectively, and accuracy gains of ∼16%, ∼2%, ∼8% and ∼14% when comparing the ensemble methodology versus the best ARIMA Forecaster in the same testing months. Table 7.2 indicates a similar pattern for the RMSE, where one can see monthly error improvements of ∼22%, ∼11%, ∼13% and ∼28% in comparison with the mean accuracy of all the ARIMA forecasters and RMSE accuracy gains of 12%, ∼1%, ∼10% and ∼18%, respectively, for the four different testing months. In addition, we can oberve that the methodology's accuracy is not specifically dependent on the time-series scale and variance characteristics, i.e., the largest MAPE and RMSE values took place in the month of April, which as observed previously is not the month with the highest mean or standard deviation.

Table 7.1: One day-ahead NE Load Forecasting results: MAPE (%).

Prediction Model/Test Month	February 16	April 16	August 16	October 16
ARIMA Forecaster 1	4.36	6.34	5.26	4.37
ARIMA Forecaster 2	4.79	7.28	5.25	5.37
ARIMA Forecaster 3	4.87	7.40	5.51	5.32
ARIMA Forecaster 4	4.93	7.49	5.73	5.20
ARIMA Forecaster 5	4.93	7.55	5.48	5.14
Ensemble DFNN	**3.67**	**6.19**	**4.82**	**3.75**

Figure 7.5 illustrates the forecasting performance of the proposed ensemble methodology in comparison with the selected ARIMA forecasters (inputs) for two indiscriminately selected fortnights, i.e., the period from 1 to 14 February (winter) and the period from 14 to

Table 7.2: One day-ahead NE Load Forecasting results: RMSE (MW).

Prediction Model/Test Month	February 16	April 16	August 16	October 16
ARIMA Forecaster 1	737.3	1266.6	983.7	774.5
ARIMA Forecaster 2	840.7	1443.5	978.6	937.7
ARIMA Forecaster 3	859.8	1446.4	1034.6	916.7
ARIMA Forecaster 4	857.6	1454.9	1067.2	891.7
ARIMA Forecaster 5	860.5	1454.6	1034.1	915.9
Ensemble DFNN	**649.4**	**1254.1**	**883.3**	**634.7**

27 August (summer). This figure also underlines the ensemble methodology's good performance in resembling the real load profile. More importantly, these examples reveal the ability of DFNN to generalize beyond the provided ARIMA forecasters and extrapolate further, thereby enabling the ensemble forecast to match the real load more closely. This behavior is illustrated in the zoomed secondary axis in both plots, where one can clearly observe the ensemble forecast outside of the region mapped by the 15 selected ARIMA forecasters (input prediction models). In addition, it is also clear that, as expected, the ARIMA forecasters are a suitable option to capture the linear components of the time-series but struggle to reproduce the faster changing variations of the morning and evening peak loads. This confirms the need for a stacking ensemble methodology where a deep network with nonlinear mapping abilities, fed with additional data, compensates this shortcoming and provides relevant error accuracy gains.

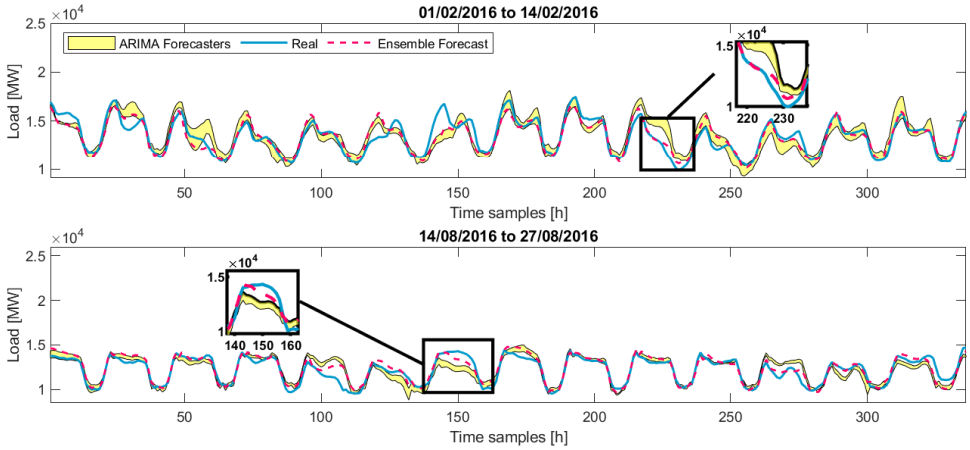


Figure 7.5: Load forecasts vs. real load (New England region), where the ARIMA forecasters area in yellow represents the region mapped by the forecasts of 15 different ARIMA models, the stacking ensemble forecast is illustrated in a dashed magenta line and the real (target) load is illustrated in a solid blue line.

To complement these results, a comparative analysis was made against a common shallow NN architecture with a single hidden layer with 35 nodes, also trained using the SCG algorithm, for benchmarking purposes and also for evaluating the effect of the added exogenous input variables in the input dataset (the same exogenous variables were fed to the ensemble methodology). Ergo, two different input datasets are considered. Likewise, a support vector machine regression with a Gaussian kernel function and input standard-

ization is used to map the input domain, building the decision boundaries to achieve the hyperplane that allow us to predict continuous load outputs. The average error metrics over the course of 50 runs are shown in Tables 7.3 and 7.4, highlighting not only the superior performance of the the ensemble method for all the testing periods but also the clear benefit of including related exogenous information in the input datasets when we compare the errors between the two NNs. A brief statistical analysis reveals gains in terms of RMSE between ∼1.9% to ∼17.5% versus the NN with exogenous inputs; between ∼10.5% to ∼32.9% versus the NN without exogenous inputs; and between ∼3.4% to ∼33.7% versus the Regression-SVM. An identical account is also depicted by the MAPE (%) values, revealing considerable monthly accuracy improvements (between ∼11.5% to ∼33.8%) for all the months versus the other three forecasting approaches, with the exception being the mean monthly MAPE (%) of August where the gain was more moderate (∼1.5%).

Table 7.3: Comparison of one day-ahead NE Load Forecasting results: MAPE (%).

Method/Test Month	February 16	April 16	August 16	October 16
R-SVM	4.71	8.15	5.39	5.67
NN wo/ Exog	4.87	6.96	5.85	5.16
NN w/ Exog	4.22	6.99	4.89	4.66
Ensemble DFNN	**3.67**	**6.19**	**4.82**	**3.75**

Table 7.4: Comparison of one day-ahead NE Load Forecasting results: RMSE (MW).

Method/Test Month	February 16	April 16	August 16	October 16
R-SVM	769.5	1492.7	914.6	958.2
NN wo/ Exog	860.1	1400.5	1066.8	945.2
NN w/ Exog	717.8	1278.6	923.5	769.2
Ensemble DFNN	**649.4**	**1254.1**	**883.3**	**634.7**

Another important feature of the proposed methodology and its comparative NN peers was the relatively low levels of standard variation across the 50 runs, as one can notice in Figure 7.6, by the narrow range of the interquartile distribution in the box-plot and the small number of outliers, both for the MAPE and RMSE error metrics. More specifically, the ensemble methodology even with a large input dataset is in line with the NN with exogenous inputs variance, and its median line ”sits” lower in the box-plots for all four testing months. In the months of February, April, August and October, the recorded values of standard deviation were, respectively, 0.132%, 0.252%, 0.217% and 0.113% in terms of MAPE (%) and 26.5 MW, 60.8 MW, 36.3 MW and 21.4 MW in terms of RMSE. These numbers validate the robustness of the presented ensemble methodology with respect to consistently performing STLF in an improved manner.

Last but not least, Figures 7.7 and 7.8 are presented to illustrate two different weeks (first week of April and third week of October, respectively) of forecasting loads versus the real load, as well as the individual hourly deviation (signed error) of each comparative method and the proposed approach below (in the form of bar plots). These bars allow a better understanding of the difference between predicted and real hourly loads, and the smaller (better) magnitudes seen in the Ensemble DNN reinforce the inferences made by not only

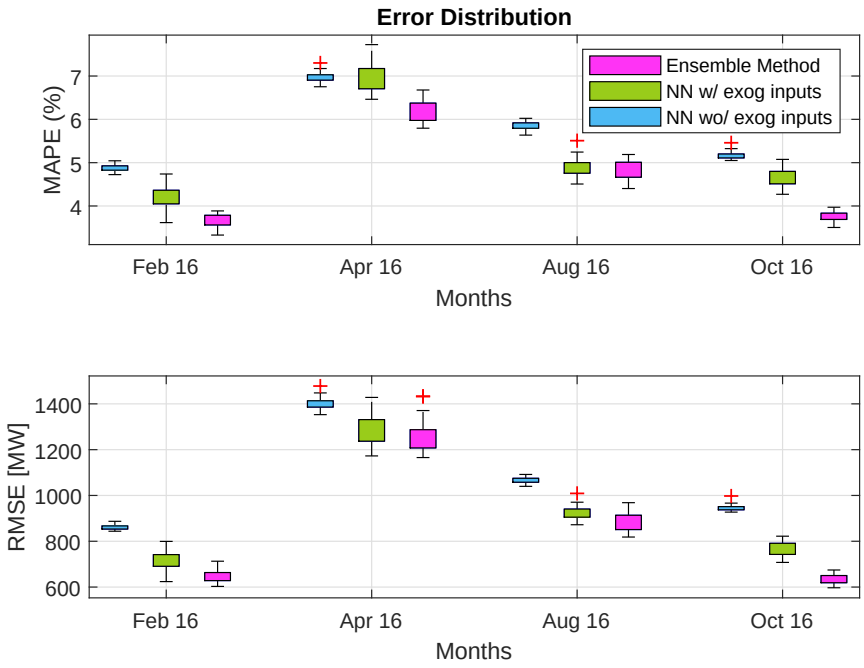


Figure 7.6: Box-plots of the MAPE (%) and RMSE (MW) error distributions (NN without exogenous inputs-blue, NN with exogenous inputs-green and the proposed ensemble methodology—magenta over the course of 50 runs).

analyzing the error tables but also the error distribution in Figure 7.6, with a slightly positive bias in terms of error signal for all the considered methods.

7.3 Electricity Price Forecasting

The liberalization of electricity markets, which occurred (mostly) during the late nineties and early thousands as we saw in sub-section 1.1.4, was a major milestone in terms of electricity pricing, with the "freedom" for producers and consumers to place their energy bids (energy trade orders) in a pool-based market. On the one hand, this framework promoted a more crowded, transparent, and horizontal playing field, known to foster price competitiveness, system efficiency and innovation in all the players involved in the EPS [587]. But on the other hand, this intricate environment, with an increasing number of intertwined "moving parts", with players of varied sizes, with different portfolios, both in terms of the end-user types and its numbers, its associated load profile, as well as in terms of generation, with an increasing penetration of renewables, translates a higher degree of uncertainty, which ultimately is reflected in terms of price volatility. As such (in a broader sense), the day-ahead price time-series is characterized by multiple seasonality effects (calendar), price spikes, periods of high frequency variance, among others [588, 589]. Moreover, as it also occurs with financial type commodities, the price is affected by market sentiment and expectations. These include planned additions of transmission and generation capacity, primary energy price variations (coal, natural gas, fuel, etc.), regulatory changes, demand growth, players' financial health, operational faults, transmission

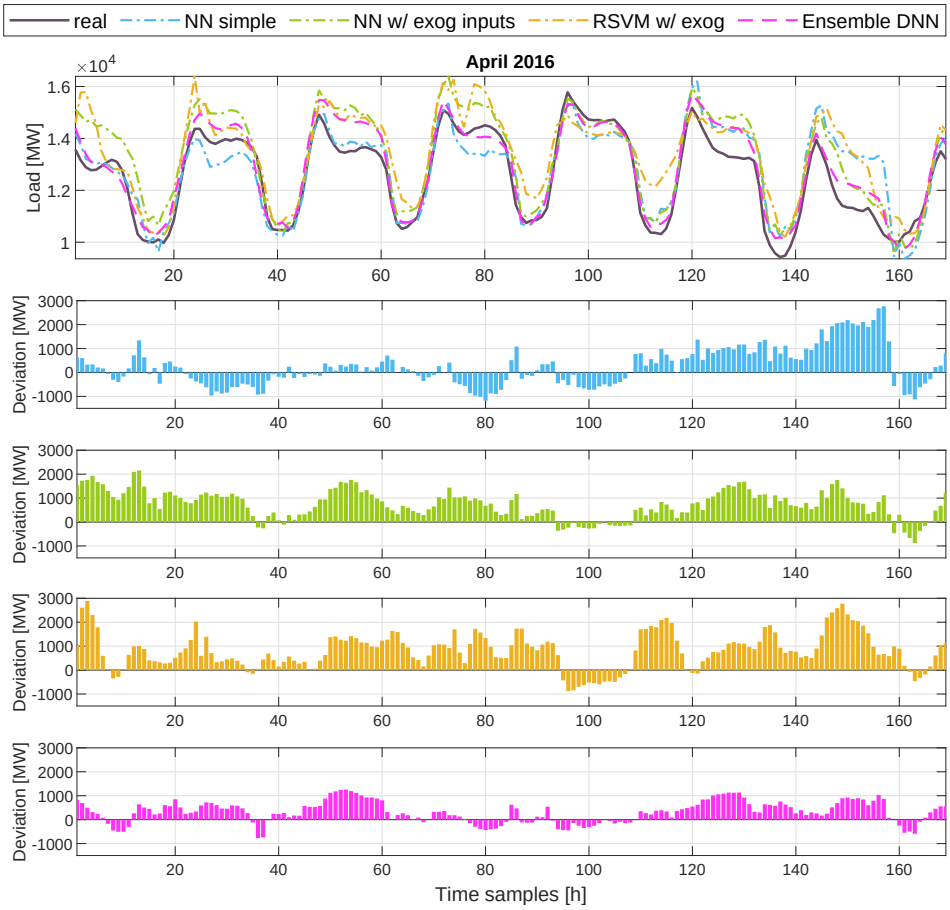


Figure 7.7: Predicted loads vs. real load and the respective mean deviation (signed error) of each method: 1st week of April 2016.

congestion, infrastructure and environmental disasters, along others [137].

All of these factors contribute to an augmented difficulty in electricity price forecasting [590]. Given that a large part of the operational and market decisions take place in lead times ranging from hours to one week, this time horizon encompasses a major portion of the research efforts and is commonly referred to as short-term price forecasting (STPF). Hence, the search for reliable STPF methods is of great interest for utilities and market players, which aim to maximize their profits. And to market, transmission, and grid operators to ensure the security of supply and EPS reliability.

A literature survey reveals a broad spectrum of STPF methods, where generally, statistical methods, ML models and hybrid methods are considered the three major classes of methods [590]. The first class consists in a variety of time series models, with both univariate and multivariate regression settings [591]. However, the emergence of ML methods has brought tremendous progress to the field, these include topologies such as: ANNs, RNNs, LSTMs, GRUs architectures, etc. They constitute some of the most popular forecasting engine choices [592, 593], given its well-known ability to handle large amounts of data, and, from these capture hard non-linear dependencies (model fitting). The last class of methods is the most prevalent nowadays and, as the name suggests, it encompasses all

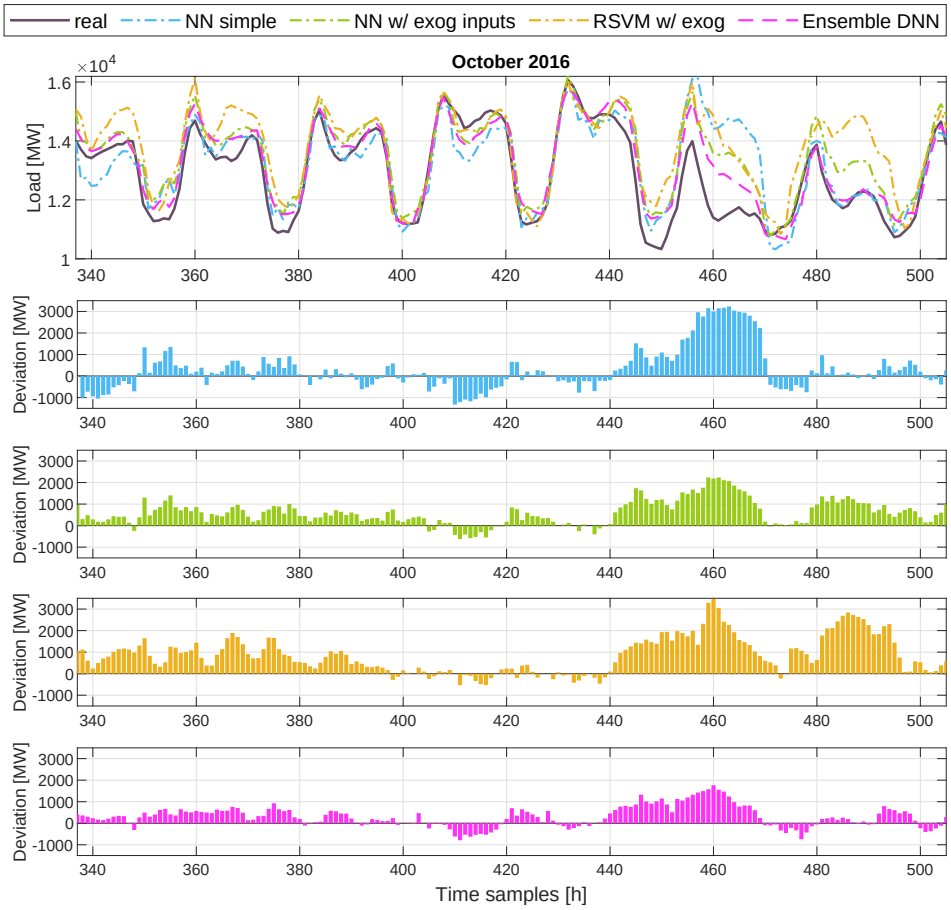


Figure 7.8: Predicted loads vs. real load and the respective mean deviation (signed error) of each method: 3rd week of April 2016.

the proposals that combine several methods, including the previous two referred classes. In this manner, they tend to achieve greater strides when compared with single methods. These methods often tackle different stages of the forecasting process, including data pre-processing, clustering, architecture optimization, design of multi-leveled deep learning architectures, parallel and ensemble methodologies, as well as forecast correction techniques, etc.

For example, in [594], the proponents have resorted to three different algorithms to tackle different issues, first Variational Mode Decomposition (VMD) decomposes the electricity price dataset into intrinsic mode functions, then a CNN is used, given its ability to perform meaningful feature extraction from the decomposed data. And lastly, a GRU is used as a forecasting engine, i.e., it will learn from the training input/target samples. In turn, authors, in [595] develop optimal kernel-based extreme learning machine models to solve the STPF problem. Zhang et al. [596] present another good instance of an adaptive hybrid model that combines methods from different classes, namely SARIMA and deep belief network, VMD for data pre-processing, and a PSO variant to optimize different parameters of the other methods. Lastly, in a more extreme example, a combined model was proposed in [597], where six individual methods/techniques are assigned to different tasks.

Singular spectrum analysis is employed to denoise the data. A Jordan ANN, an echo state network and a least squares support vector machine are used as forecasting engines and two optimization algorithms are used to optimize the combined model parameters. More recently, acknowledging the growing dominance of cutting-edge machine learning algorithms, the outputs of these models have been used as inputs to a predictive meta-model, aiming to achieve the most accurate forecasting solution—following, to some extent, the ensemble approach adopted for STLF in Sub-section 7.2.2.

7.3.1 Case Study I: Iberian electricity market

The research focus is often placed on the forecaster itself (typically a regression-type ML model) rather than on the selection of input data. Therefore, a case study aiming to address the existing gap in knowledge regarding the benefits of including various preselected exogenous variables is considered valuable.

To this end, the STPF approach will leverage the capabilities of multilayered neural networks — as introduced in sub-section 5.3.3.1 — using well-established configurations to test and evaluate the sensitivity of forecasting results, considering the Iberian day-ahead electricity prices.

7.3.1.1 Endogenous variable: day-ahead electricity prices

Data from the well-versed MIBEL [598] is used as a case-study for the proposed sensitivity analysis, more specifically the day-ahead market prices, i.e., the primary market, where prices and energy volumes are assigned for the 24h of the next day. For testing purposes, the hourly price data of 4 different months of 2019, namely February, May, August, and October (comprising different astronomical seasons of the year and different weather conditions) are used for testing the different ANN models, while the Iberian market day-ahead electricity prices from 2016 to 2018, a span of 3 years of hourly data, is used for training the models (form the training input/target samples). With that, it is important to "know" the time-series main characteristics. In accordance, Figure 7.9 is built to analyze the patterns, and its lagged inner relationships, using the auto-correlation (ACF) and partial auto-correlation (PACF) functions.

As expected, and mentioned in sub-section 7.3, the first subplot in Figure 7.9, reveals a time-series with several degrees of intra-year seasonalities, with several outliers and periods of high frequency variations. The ACF plot reveals a slowly decaying auto-correlation, which signals a non-random distribution, with clear daily, weekly, and monthly dependences, i.e., strong correlations (spikes) with adjacent samples in both PACF and ACF functions. In terms of the most meaningful statistics, the data shows an average recorded price of 49.21 €/MWh, and a standard deviation of 14.34 €/MWh, which is significant since it represents roughly 30% of the mean value. Both the maximum recorded price of 101.99 €/MWh and the minimum price of 0.03 €/MWh occurred during the winter (cold) season, in January of 2017 and during the last week of December of 2019, respec-

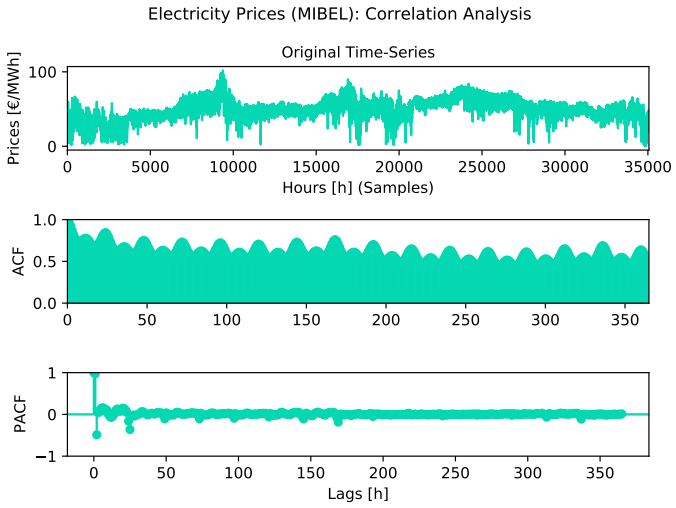


Figure 7.9: MIBEL Day-Ahead Electricity Prices Time-Series (2016-2019) and Correlation Analysis.

tively. The recorded bottom value is virtually equal to the lower price limit of 0 €/MWh, while the maximum was still far from the upper price limit (in place at the time).
All these traits help to explain the difficulty in terms of the day-ahead electricity forecasting in MIBEL and the same was confirmed in a thorough study conducted in [599], where several well-known methods, including ML-based ones, are used to predict the day-ahead MIBEL prices, in a total of 50 days of testing. The mean results, shown in Table 7.5, reveal just that, with all the mean MAPEs above 10%, RMSE values above 5 €/MWh, together with sizable standard deviation.

Table 7.5: One-day ahead electricity MIBEL price forecast (random 50 days – 2018).

Methods	MAPE (%)	RMSE [€/MWh]
ARIMA	15.21 ± 5.06	6.91 ± 2.96
LSTM	13.23 ± 4.13	5.62 ± 2.05
XGBoost	11.91 ± 4.00	5.19 ± 1.89
SVR	15.03 ± 3.64	6.61 ± 2.68

7.3.1.2 Exogenous variables: selection and statistical analysis

A common feature of ML-based STPF methods is the use of relevant exogenous variables. The reason being to provide complementary information (beyond the lagged price values), that often improves the contextualization process during the model fit, by explicitly providing influential components behind the electricity price formation and its recorded variations. Calendar information has extensively shown to greatly contribute to the model training accuracy [600]. For instance, the intra-week price seasonality is relevant, given that weekdays are different from weekends, public holidays (weekday) tend to be closer to Sunday's profile, and so on [601].
Accordingly, the following calendar ordinal variables are selected as promising: hour,

month, holiday index (binary value) and weekday index. Understanding that the market balancing is a match between load and generation, these two time-series, and interestingly its forecasted value for the same horizon as the price prediction, are also regularly chosen [602]. Following this path, this work will produce a set of secondary (intermediate) day-ahead load forecasts for both the Portuguese and the Spanish Peninsular system's load. With regards to the generation side, the day-ahead wind (onshore) and solar (PV + thermal) non-dispatchable generation forecasts for Peninsular Iberia, collected from the Independent System Operators, are also deemed worthy of assessment, given the relatively high levels of renewables integration in the market. These time-series are especially relevant, due to the favorable feed-in-tariffs and precedence agreements for these producers, which enables them to pose their market bids at very low (almost null) prices [276]. Thus, representing a solid indicator of a downward market driver. In turn, for the non-renewable sources, and given the prominence of natural gas/cogeneration thermal plants in Iberia, the day-ahead natural gas prices (D+1 in MIBGAS) are also worthy of investigation.

For the third set of exogenous variables, two original indicators, for this market, are handpicked. The first is the (measured) generation pump consumption, which provides an account of the hydropower producers' ability to use the water reservoirs, effectively as storage. This is relevant since, contrary to the other main renewable sources, hydro plants provide a degree of operational flexibility for the market players, enabling them to offset surpluses at low prices, or to determine the prices in peak periods. Secondly, the last selected exogenous variable is the available import capacity from France to the Iberian Peninsula, immediately after the matching up of the daily market. Thus, providing a contextualization of the interconnection capacity usage from France, defining different scenarios in terms of how the available Peninsular generation capacity is being used, signaling in some form the hourly import price competitiveness from a neighboring electricity market. The meaningful statistics for the handpicked selected exogenous variables are provided in Table 7.6. The displayed variance and range underscores how even some slight variation in these variables may be reflected in the day-ahead price swings.

The Pearson correlation coefficient, r, also known as bi-variate correlation, is considered a good measure of the linear correlation between two sets of data, X and Y, computed as shown in Equation 7.12. In the case of this work, these are presented in Table 7.7, to provide an early indication of the relationship between the endogenous variable, i.e., MIBEL day-ahead electricity prices and the chosen exogenous variables. Its value ranges between -1 and 1, where 0 implies no correlation, while values closer to the limits indicate a very strong correlation. Besides, Table 7.7 show that the D+1 natural gas prices, the pumping load consumption and the Load Forecasts for the entire system are the most bi-variate correlated variables. On the contrary, the holiday and weekday indexes, which help characterize the data in terms of certain seasonal behaviors (with different regularity), reveal diminished levels of correlation.

Table 7.6: Exogenous variables: statistical measures.

Selected Exogenous Variables	Mean	Range	Std
Hour	–	0,1,...,23	–
Holiday index	–	0,1	–
Month	–	1,2,...,12	–
Weekday index	–	0,1,...,6	–
Day-ahead Load Forecasts (Spain) [MW]	28563.74	19324.87–39498.41	4484.54
Day-ahead Load Forecasts (Portugal) [MW]	5729.91	3695.15–8306.92	975.94
Day-ahead Natural Gas Market price: Daily Reference Price [€/MWh]	19.49	7.89–43.0	5.52
Hourly Import Capacity from France [MW]	2387.32	379.00–3750.00	659.77
Hourly Pumping (Load) Spain [MW]	-418.12	-4509.44–80.49	772.63
Hourly Day-ahead Wind + Solar (PV and Thermal) Forecasts (Spain + Portugal) [MW]	8500.49	613.02–23456.00	4071.71

$$r = \frac{\sum_i \left(\left(X_i - \overline{X} \right) \left(Y_i - \overline{Y} \right) \right)}{\sqrt{\sum_i \left(X_i - \overline{X} \right)^2 \sum_i \left(Y_i - \overline{Y} \right)^2}} \quad (7.12)$$

Where X_i and Y_i are the corresponding samples (for the same hour/date), and $\overline{X}$ and $\overline{Y}$ represent the mean of the X and Y datasets, respectively.

Table 7.7: Pearson correlation coefficients between the electricity prices and the selected exogenous variables.

Bi-variate Correlation	Day-ahead MIBEL Electricity Prices
Hour	0.211
Holiday index	-0.066
Month	0.214
Weekday index	0.152
Day-ahead Load Forecasts (Spain + Portugal)	0.515
Day-ahead Natural Gas Market price: Daily Reference Price	0.624
Hourly Import Capacity from France	0.142
Hourly Pumping (Load) Spain	0.574
Hourly Day-ahead Wind + Solar (PV and Thermal) Forecasts (Spain + Portugal)	-0.389

Though, it is important to notice that this measure has some sensitivity issues around certain data traits such as: data polluted with heavy noise, data distribution abnormalities, e.g., heavy-tailed distributions or the heavy presence of outliers, sample size, among others [596]. In addition, this measure only translates the linear correlation, while the performed analytic assessment will rely on the forecasting results, obtained by the multilayered NN, which is able to extract and learn the non-linear relationships between the input and target datasets. Therefore, the direct forecasting results provide a full account of the error gains, as a result of the inclusion of the different exogenous variables, and in turn allow us to confirm or reject the inferences shown in Table 7.7 and from past works.

7.3.1.3 Methodology: Sensitivity assessment of the exogenous variables influence

As mentioned previously, the ANNs capabilities to perform the forecasting task, as a regression-based problem, are very handy when handling with large amounts of data, which also exhibits non-linear traits. Hence, in this work a multilayered network, with feedforward propagation between the dense layers, was the natural option to perform the assessment of the forecasting sensibility (influence), in the presence of different promising exogenous variables. An overview of the main stages involving the methodology is presented in Figure 7.10.

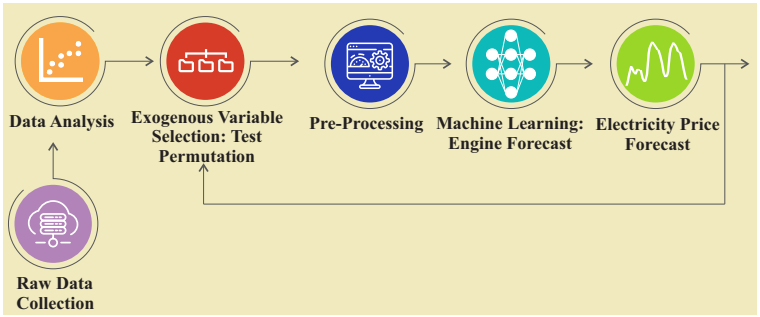


Figure 7.10: Proposed Methodology: An overview.

To achieve this purpose, all the combinations without repetition of the considered exogenous variables will be fed and tested with the multilayered FFNN. In practice, this means having different input training data samples (with different input sizes) for each combo. The evaluation of the forecasting performance will then highlight the individual and combined benefits of the inclusion of the exogenous variables. Nevertheless, and considering that combinatorial explosion occurs (due to the factorial), it was crucial to group the exogenous variables into related classes, as seen in Table 7.8.

Table 7.8: Grouped (list) of exogenous variables.

Group	Variant of exogenous input data	Group	continued
1	Calendar Variables Type 1: hours and months	4	Combined Wind + Solar Forecasts: Portugal + Spain
2	Calendar Variables Type 2: holiday index and weekday index	5	Iberian Market D+1 Nat Gas Prices
3	Combined Load Forecast (STLF): Portugal + Spain	6	Pumping Load and Import Capacity France

This grouping of the variables has allowed us to limit the total of permutations to 64, as succinctly highlighted in tree form in Figure 7.11. This is still a feasible number for the sensitivity study given the required computational effort of each simulation. In terms of the input variables size, the 33 most correlated lagged prices are a constant in all the different inputs (for all permutations), corresponding to a PACF threshold of 0.125, thus including the significant lags up to 2 weeks prior (336h).

For the D+1 natural gas prices, pumping load and import capacity from France, the 4 most important lags spanning over to 2-days prior of the lead time (forecast horizon) are selected when testing combinations that consider these variables. Lastly, both calendar type variables, groups 1 and 2, the combined Iberian system load forecasts (sum of the country values), as well as the combined Iberian wind + solar generation forecast share the time-step of the target, i.e., only one lag, since it is information regarding the same horizon as the predicted price. Further, before being fed to the multilayered FFNN, these variables are re-scaled using data normalization between -1 and 1.

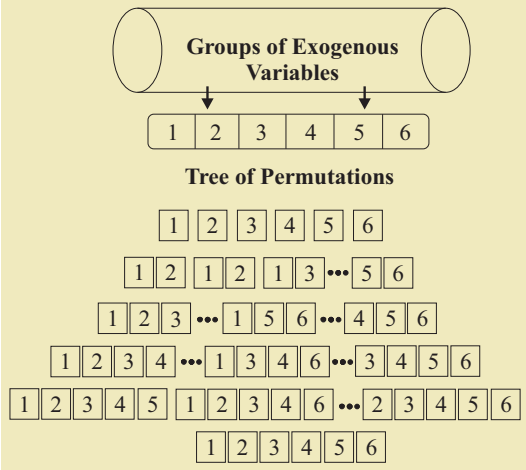


Figure 7.11: Tree of possible permutations between the selected exogenous variables, symbolized by its respective index.

In terms of the architecture, a sequential model with 3 dense layers is considered, with 25, 11 and 9 neurons in each one of these hidden layers. Hence, providing a sufficient degree of depth. Regarding the activation of each neuron, a *tanh* function is applied in the hidden layers, while a *softsign* is used in the output layer. For the learning process (model fit), our choice fell on the well-regarded Adam gradient descent algorithm, with a batch size of 6 months of data (4300 hourly samples) and the mean squared error as the loss function, MSE.

To minimize the chances of overfitting, i.e., the loss of generalization capacity, which is very prejudicial when testing the models, from the comprehensive 3-year range of training data (26304 hourly samples), 30% will be used for validation purposes. This means that these data are not used for training the model, but to ensure that during the 2250 epochs, the model stops adjusting the weights when the validation error increases for 30 consecutive epochs, i.e., with an additional stopping criterion.

7.3.1.4 Results

The entire methodology described above was implemented using Python 3.8.7 and its deep learning packages, primarily TensorFlow and Keras. To ensure a fair comparison between the different permutations of input exogenous data, 15 independent runs were simulated, thus, representing different initial conditions for the FFNN in terms of weights

initial positioning, as well as in terms of the 30% of randomly selected samples for validation. Furthermore, to measure the proposed method's prediction accuracy two common error metric were picked: MAPE (7.10) and RMSE (7.11).

As designed, the exogenous variables listed in Table 7.6 include day-ahead load forecasts for the power systems of both countries. Consequently, STLF was also performed as an intermediate objective. A brief description of the implemented approach (forecaster and input selections) and an assessment of the accuracy of these intermediate forecasting results are provided in Appendix D.

Now moving on to the main objective, the performed sensitivity analysis revealed interesting results, with different combinations, excelling in different testing months. Nevertheless, the error metrics shown in Table 7.9 confirmed that the addition of certain exogenous variables, in particular combination number 33 (2,3,4), can lead to an impressive 15% and above in error improvements.

In the same manner, illustrations of two monthly day-ahead forecasted price time-series are shown in Figures 7.12 and 7.13, revealing an added difficulty due to the periods of high frequency changes in the prices, especially in the periods where the price reaches the bottom. Others promising combinations tend to also include the natural gas price, as such a further refinement of the obtained data is required to check if there is any periodical consistence in the periods with higher errors. In addition, the error is larger for the lower prices, particularly in the presence of sharp "valleys" in the electricity price curve.

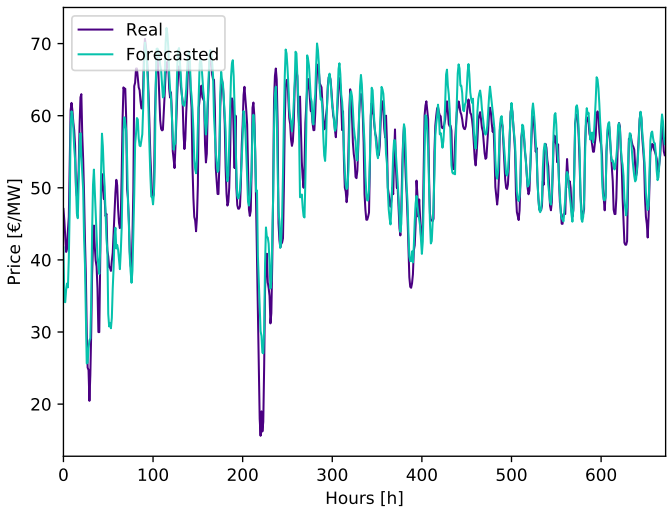


Figure 7.12: Day-ahead price forecast [€/MWh] (MIBEL) for the month of February 2019: Real Price (in violet) vs Forecasted Price (in teal green).

The results showed that traditional pre-analysis metrics, including the bi-variate correlation, should only be taken as an indicative figure, and that further sensitivity analysis with the chosen forecasting engine is recommended. Moreover, different input combinations have excelled in different testing months, underlining how can they be complementary,

Table 7.9: STPF of the MIBEL day-ahead electricity prices: error metrics analysis.

Month	Combo	MAPE (%)				RMSE [€/MWh]			
		Mean	Min	Max	Std	Mean	Min	Max	Std
Feb 2019	1	10.80	6.34	6.61	0.08	7.13	6.99	7.30	0.11
	2	11.28	7.61	9.27	0.53	7.22	6.99	7.70	0.26
	3	10.46	6.28	6.63	0.13	6.83	6.47	7.01	0.18
	4	10.51	6.74	7.26	0.17	6.74	6.60	6.87	0.11
	5	8.15	6.84	7.28	0.16	5.51	5.06	6.69	0.56
	6	11.07	6.03	6.44	0.15	7.46	7.28	7.58	0.12
	7	10.73	6.70	7.22	0.22	7.10	6.65	7.79	0.36
	33 (Best)	6.72	5.31	5.82	0.17	4.27	3.72	4.75	0.30
May 2019	1	13.45	16.16	16.74	0.21	7.42	7.24	7.61	0.12
	2	13.39	16.51	17.83	0.41	7.52	7.34	7.89	0.19
	3	12.66	16.25	16.50	0.08	7.16	6.95	7.35	0.14
	4	13.17	15.44	15.85	0.13	7.44	7.20	7.59	0.15
	5	11.32	13.34	13.98	0.21	6.30	6.13	6.50	0.12
	6	13.79	15.35	16.23	0.28	8.00	7.52	8.57	0.33
	7	13.74	15.93	16.92	0.33	7.59	7.32	7.72	0.13
	33 (Best)	9.35	10.38	11.32	0.32	5.38	5.17	5.48	0.11
Aug 2019	1	6.66	5.99	6.19	0.07	3.73	3.70	3.77	0.02
	2	6.84	5.80	6.10	0.10	3.76	3.64	3.86	0.07
	3	6.42	5.57	5.86	0.10	3.58	3.50	3.72	0.08
	4	6.34	5.45	5.56	0.04	3.54	3.50	3.58	0.03
	5	6.18	5.91	6.21	0.11	3.38	3.30	3.42	0.04
	6	8.88	7.84	10.63	0.98	4.94	4.23	6.07	0.69
	7	6.52	6.03	6.75	0.23	3.67	3.51	3.80	0.09
	33 (Best)	5.07	4.27	4.51	0.09	2.76	2.61	2.84	0.08
Oct 2019	1	8.67	10.36	10.91	0.19	5.09	4.94	5.29	0.11
	2	9.74	11.17	11.86	0.28	5.63	5.34	5.92	0.22
	3	8.94	10.44	11.18	0.26	5.22	5.10	5.34	0.10
	4	8.90	10.62	11.60	0.37	5.39	5.33	5.54	0.07
	5	8.01	8.28	8.52	0.10	4.50	4.36	4.64	0.09
	6	10.97	10.43	14.69	1.44	6.30	5.62	6.88	0.38
	7	8.88	10.39	11.40	0.34	5.08	4.79	5.31	0.17
	33 (Best)	6.41	6.35	7.34	0.32	3.73	3.62	3.83	0.08

and so, why is it necessary to tailor the input to the chosen case study data. Further work will consider the evaluation of additional exogenous variables, i.e., the inclusion of different auxiliary price forecasting inputs, as enhancing guidelines to achieve a better error accuracy, coupled with a more recent weight update strategy during the tested year.

7.3.2 Case study II: Daily operational optimization of a hybrid energy system using an STPF-based approach

An hybrid energy system (HES) is typically formed by matching different forms of renewable/non-renewable energy sources with or without storage units in order to achieve a greater bal-

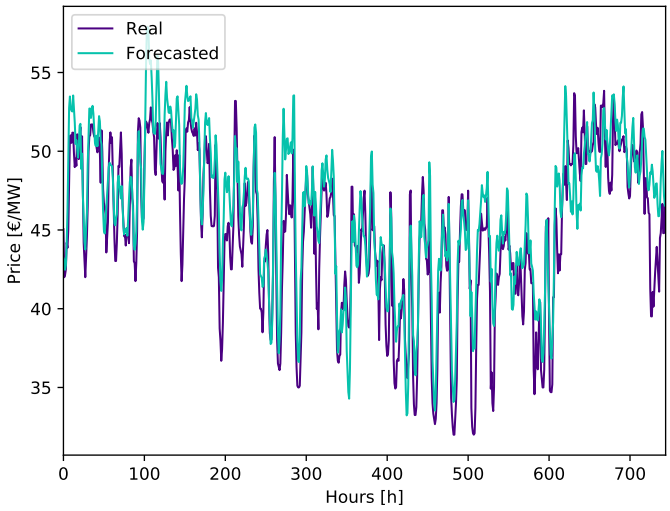


Figure 7.13: Day-ahead price forecast [€/MWh] (MIBEL) for the month of August 2019: Real Price (in violet) vs Forecasted Price (in teal green).

ance in the energy supply (mitigating its uncertain nature), a more distributed generation, as well as an increased system reliance [603, 603–605], i.e., by better exploiting the local endogenous resources, taking advantage of the combined production from different technologies, e.g., solar photovoltaic, solar thermal, wind, hydro, biomass, hydrogen, fuel cells, micro-turbine generators, diesel generators among others. As a result of the wide range of existing technologies, numerous HESs have been proposed, combining all sorts of generation technologies to deal with both generation and load uncertainty, in different topologies and operating strategies [606–609]. For instance an optimal scheduling approach for a thermal-wind-solar system with storage, considering the uncertainty in load forecast, and in wind and solar power output is proposed in [609], another example is the stand-alone wind-photovoltaic-hydrogen presented in [610]. In another instance, a maritime solution anticipating the increasingly high costs of fossil fuels and increasing restrictions on diesel usage and gas emissions from ships was proposed. Its operation is based on an optimal PV/battery/diesel scheme during the cold-ironing connection. A solar HES with diesel generators is contemplated (via Homer®) in [603] to respond to the growing demand on isolated telecommunication infrastructures in Nigeria. This type of hybrid renewable energy systems integration is an effective electrification solution, in particular for rural electrification in least developed countries. Following this path, an off-grid hybrid PV-diesel-battery appliance (with an optimization based on PSO and ε-constraint method) serving a demand of 20 households in remote Saharan regions with 93% renewable fraction is proposed in [611].

Besides, these systems can operate as “grid-connected” or “off-grid”, i.e., as stand-alone systems, constituting a suitable solution for rural and isolated areas (ranging from suitable HES for remote islands to particular alternatives for buildings) with demand and

resource availability. However, the optimized operation of these systems still poses some challenges. For all these reasons, HESs are presently a hot topic among researchers [607] covering topics ranging from feasibility analysis, optimum sizing, modeling, dispatch, control aspects and reliability issues [605, 612].
In this framework, this research proposes an optimal dispatch strategy for a simulated grid connected/stand-alone hybrid energy system, owned by an hypothetical large prosumer with abundant resources availability. Contemplating the common PV/Wind (onshore) generation in simultaneous operation mode [612], as well as a traditional energy storage unit (battery), complemented by the addition of a less common small hydro unit with pumping that, in parallel to the battery bank, will provide storage capability. This is a particular feature since few research works have explored HES with both multi-generation units and multi-storage units [613]. The operating strategy proposed in this paper, essentially, can be decomposed into two phases.

- In the first phase, a STPF is conducted for the day ahead — not as an end in itself, but as a foundation for the proposed operational strategy. Specifically, the dispatch strategy is developed based on a 24-hour electricity price forecast, integrated with the state-of-charge data from storage technologies.
- In a second phase, a real-time dispatch is performed subject to several constraints, incorporating the short-term price forecast and the current state-of-charge of the storage units, to allow the best compromise between cost and reliability of the various generation and storage technologies, i.e., satisfying the electricity demand while minimizing the operation costs.

To keep this section concise and focused, only the topics essential to the chapter — including the case study, forecasting methodology, and key results — are addressed here. In contrast, the overall operational strategy, time-series analysis, and hybrid system modeling, which go beyond the immediate scope of this chapter, are presented in the dedicated Appendix E.

7.3.2.1 Architecture and operational strategy

The chosen HES architecture, interconnected with the grid, combining the generation technologies (PV-wind-PSH-diesel) and the storage unit mentioned above, is illustrated in Figure 7.14. An advantage of this architecture is the plug-and-play operation, where wind, hydropower and diesel generation units are coupled directly to the AC Bus. The storage unit (battery bank) and the PV generation are coupled to a DC bus, through proper DC/DC converters, which is then sequentially coupled to the AC bus through a bidirectional DC/AC, thus allowing a two-way (flow between the two bus-bars) power flow. The diesel generator will play a reserve role, i.e., its primary state is off, only on shut down situations will be turned one, otherwise the grid exchanges will provide existing shortages.
The proposed HES was designed to fully satisfy a real load profile based on a simulated prosumer scenario. A detailed account of all operational assumptions, decision-making

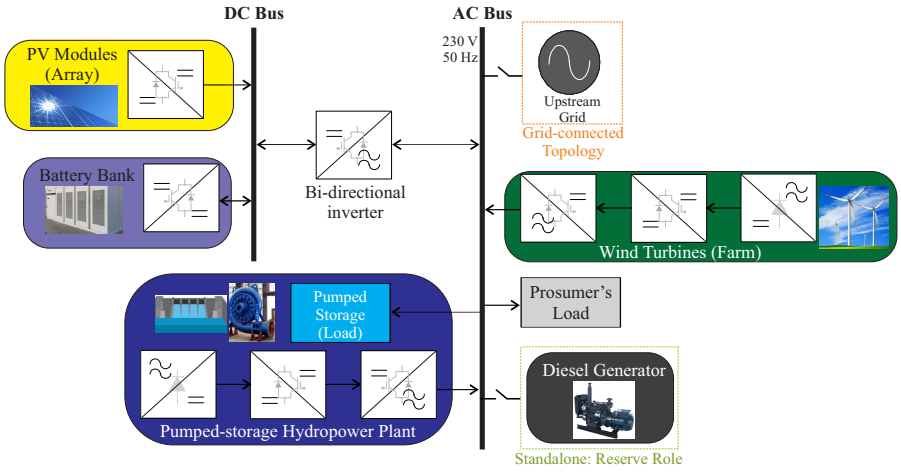


Figure 7.14: Schematic representation of the proposed hybrid system architecture.

considerations, cost equations, auxiliary variables, and the generic problem pseudo-code (Algorithm 4) is provided in section E.1.

In addition, a detailed description of the main considerations for each component of the proposed HES architecture — along with their respective modeling, including governing equations and rules (a key aspect for simulation purposes) — is presented in section E.3.

Taking advantage of all the knowledge already gathered in 5, a common FFNN with 2-hidden layers, employing dozens of neurons constitutes the forecasting engine. With SCG algorithm is then used as a learning algorithm. The performance of the FFNN is evaluated through the mean absolute error (MAE) at each training step (iteration). Additionally, a standard cross-validation technique (K-fold as described in 5.3.2.2) was used to partition the data set into 4 mutually exclusive subsets, with a distribution of 7% and 25% of training and validation data, respectively.

Besides the forecasting engine choice, another delicate issue surrounding forecasting accuracy is the proper feature selection (selected price explanatory variables) [46]. By this means, first a pre-selection stage is carried out where Correlation Analysis (CA) is performed over the available price data, i.e., we take the current day D (last available information) and measure the partial correlation between the prior lags with a span up to a total of 168 h (enabling the capture of intra-day and intra-week relationships), and by selecting the most significant lags the recent days set is formed. Then, a complementary set containing information regarding Similar Days (SD) is also formed, i.e., we take the current day D and find other historical days (SD = 3 days) with a similar daily pattern, thereby exploring the seasonal properties of the electricity price time-series. With this stage completed, then Wavelet Transform is used to decompose the signal up to three levels into its detailed and approximation coefficients, and these are also fed to the multi-layer network.

In addition, the FFNN input layer is completed with a set of exogenous variables, including calendar variables (hour, season, weekday and binary holiday indexes), lagged load and renewable generation, as well as load and renewable generation estimates (i.e., with the same time horizon for which the price forecast is made). An extended diagram describing

the followed forecasting methodology (first phase) is presented in Figure 7.15.

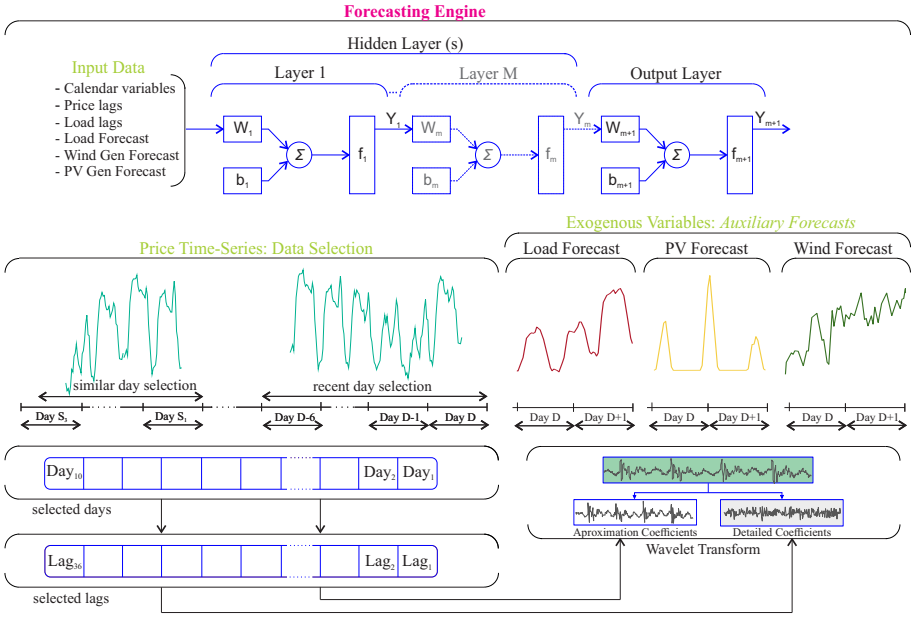


Figure 7.15: Illustration of the employed price forecasting: FFNN input data selection mechanism.

7.3.2.2 Case study: source data and adjustments

The proposed daily operation optimization strategy for a HES (dispatch strategy) is tested under the assumption that a prosumer, such as a resourceful industrial/commercial corporation, has invested in an increasingly mature energy generation portfolio, firstly with the intent of suppressing/lessening their electric loads consumption burden, but also to take advantage of the decreasing storage cost, in other words to "act" as market player on both sides. Moreover, by taking advantage of the exchanges with the upstream grid, in foreseeable profitable periods, thus he accelerates not only the payback period (shortened), but also the medium-long term profitability. For example, the operation of a storage system owned by a large consumer that purchases electricity from the Ontario's wholesale market, where the price forecasts are also applied to an optimization platform for operation scheduling of a battery energy storage system, within a grid-connected micro-grid in Ontario, is studied in [614].

For simulation purposes, it was necessary to use a nationwide database, which included not only all the parameters required to model the different HES sources, but also the vital information (explanatory/exogenous variables) for the successful implementation of a 24-h electricity prices forecasting in a real market context. To serve this purpose, the open power system data [615] was used. This online platform compiles national/system level data from several sources, not only containing common time-series data (load, prices and renewable generation and its respective forecasts, etc.), but also the important weather data (solar irradiation, temperature and windspeed). Thus, to include all the necessary

simulation variables (used both in price prediction and in HES modeling), only a handful of countries provide all the necessary data on an hourly basis. In this sense, the data regarding the Czech Republic (CZ) system for the period from 1 January 2015 until 31 December 2016 was used. Furthermore, the annual average diesel prices for Czech Republic were also used to representatively model the diesel generator operation cost [616]. Nevertheless, some processing of these raw data is still necessary, starting with the necessary electric load adjustment, since the average CZ system load value is about 7343 MW, then a reducing scale factor is applied to bring the load to an average value around 100 kW (prosumer level). Secondly, since the wind speed is the result of average anemometers measurements across the country, its average value is around 3.93 m/s, which following the Wind Energy Resource Atlas of the United States for a height of 10 m placing it in the lowest class of wind power potential (class 1) and well below the majority of the wind turbines cut-in speed, when viable wind farming sites usually present values higher than 6 m/s. So, therefore, a scaling factor of 3 is applied to the existing wind profile to simulate more favorable conditions, as is customary in wind farm sites. In the same manner, a minor adjustment is applied to the irradiance levels, in order to better replicate more consistently the existing conditions in photovoltaic parks.
The electricity prices and the adjusted prosumer load profile are shown in Figure 7.16, in turn Figure 7.17 illustrates the employed meteorological data.

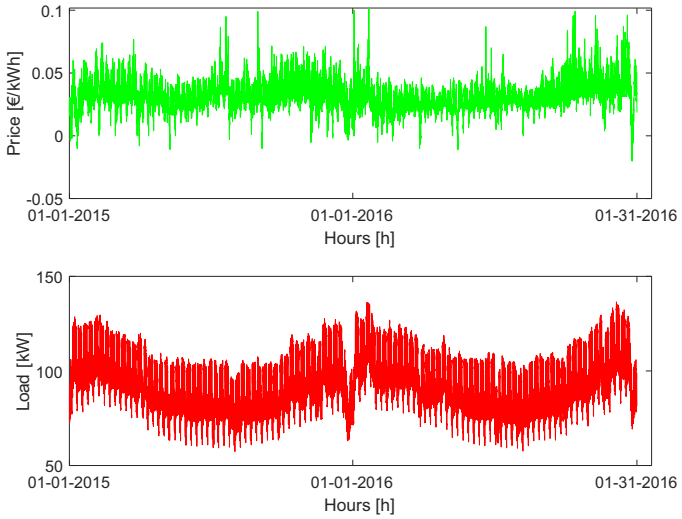


Figure 7.16: Day-ahead Electricity Prices and Prosumer's Load Profile.

Adjusted to the presented load profile, this HES totals roughly 275 kW of peak non-dispatchable renewable capacity, however, due to the unstable nature of these production sources, a traditional battery bank with an energy availability of around 130 kWh is added to the mix, accompanied by a small hydro unit which is one of the highlights of this paper, since it provides a second storage capacity and a nominal power around 150 kW. To conclude, the availability of diesel generators is also considered (60 kW), to evaluate whether or not constitutes a viable alternative to the upstream grid exchanges, thereby as a preferred

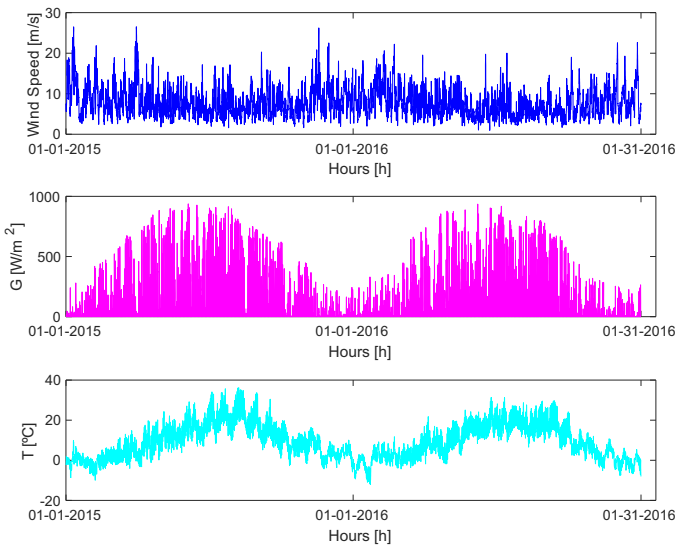


Figure 7.17: Weather Data: Wind Speed, Solar Irradiance and Temperature.

reserve source.
To complement the visual analysis of the time-series, a statistical analysis was also conducted. This helped tailor the case study and refine the characteristics of the selected input data — including decisions regarding lag selection — and is presented in Section E.2.

7.3.2.3 Results

To validate and analyze the proposed dispatch strategy, a scenario involving a grid-connected topology — which allows energy power flows (exchanges) with the upstream grid — is considered. These exchanges are governed by the presence of surpluses or deficits. The estimated cost or income resulting from grid usage is then evaluated and compared against the benefits and constraints imposed by the various technologies within the hybrid system, as defined by their respective cost functions (Equations E.1 and E.2).
The proposed HES and its operational strategy was tested in 4 different situations, which correspond to 1 week per season of the year. The chosen weeks were from February 09th to the 15th of 2016 (winter week); from April 25th to May the 01st of 2016 (spring week); from August 08th to the 14th of 2016 (summer week); finally, from November 11th to the 17th of 2016 (fall week).
Thus, not only do the load profile, prices, and renewable generation differ, but the initial SOCs — $SOC_1\,(0)$ and $SOC_2\,(0)$ for the battery bank and PSH, respectively — also define four different starting points, as can be seen in Table 7.10.
Besides the test week and the starting SOCs, the main simulation results are also given in Table 7.10, they include the average non-dispatchable power balance $\bar{\varepsilon}\,(k)$; the average power-flow with the upstream grid $\bar{P}_{\text{grid}}$; its respective imports and exports costs $Grid_{\text{buy}}$ and $Grid_{\text{sell}}$; lastly the final SOCs are presented, $SOC_1\,(168)$ and $SOC_2\,(168)$. For in-

stance, it makes perfect sense when we refer to the PSH to consider the initial SOC in a summer week ($SOC_2\,(168) = 0.50$) lower than the existing SOC in the winter week ($SOC_2\,(168) = 0.99$).

Table 7.10: HES Simulation: Main results.

Test Week	$SOC_1\,(0)$	$SOC_2\,(0)$	$\bar{\varepsilon}\,(k)$	$\bar{P}_{\text{grid}}$	$Grid_{\text{buy}}$	$Grid_{\text{sell}}$	$SOC_1\,(168)$	$SOC_2\,(168)$
12–18 Feb (Winter)	0.50	0.99	22.17	17.62	87.11	1.97	0.21	0.25
18–24 May (Spring)	0.85	0.85	−3.82	1.53	31.48	25.62	0.92	0.90
10–16 Aug (Summer)	0.70	0.50	−8.89	−1.36	24.36	38.51	0.71	0.68
11–17 Nov (Fall)	0.90	0.75	8.54	8.61	65.61	6.83	0.29	0.24

From the results, we can see the extreme of when there is a week (winter) that presents a very weak renewable generation coupled with high load levels, especially due to the absence of solar generation, as such most of the power flow is given by the upstream grid and the PSH unit, since its initial SOC was practically maximum. Interestingly, the spring week shows that on average the renewable generation slightly exceeds the load by about 4 kW. However, the average price of the exchanges with the grid is slightly positive, proving the balance of the implemented HES dispatch strategy that takes advantage of the existing surplus not only to profit, but also to leveling both SOCs during the 168 h. In turn, the summer week, leveraged by high wind and photovoltaic generation levels, illustrates not only significant net sales, but also enables the upstream pumping to refill partially the dam and the $SOC_2\,(168)$ is 18% higher than the initial one, while the battery bank is used punctually and remains with a relatively unchanged SOC. For the fall week, we can see an average renewable generation deficit, therefore the HES is called upon to alleviate this deficit, and the PSH and grid are responsible for serving the remaining load demand.
Additionally, the resulting energy HES dispatch for these 4 weeks is illustrated by the bar plot in Figure 7.18, where the share of energy generation (168 h) of each HES technology and its respective load level is represented. As we can see, the employed datasets clearly reveal the benefits of the used hybrid system sizing, with an installed wind power capacity far superior to the remaining generation technologies.
Furthermore, the statistics from each test week, i.e., the mean, standard deviation, minimum and maximum value and its mean load value are presented in Table 7.11, side-by-side with the respective forecasting performance, measured by the common MAPE (%) and RMSE metrics.

Table 7.11: Test weeks statistics and auxiliary STPF forecasting accuracy.

Test Week	Mean	St. Dev	Min	Max	Mean Load	MAPE(%)	RMSE (€)
12–18 February (Winter)	2.45×10^{-2}	9.75×10^{-3}	2.23×10^{-3}	4.65×10^{-2}	108.0	9.12	2.53×10^{-3}
18–24 May (Spring)	2.66×10^{-2}	9.31×10^{-3}	2.25×10^{-3}	4.23×10^{-2}	92.9	6.67	1.73×10^{-3}
10–16 August (Summer)	2.69×10^{-2}	6.66×10^{-3}	1.40×10^{-2}	4.00×10^{-2}	87.0	5.49	1.89×10^{-3}
11–17 November (Fall)	4.06×10^{-2}	9.75×10^{-3}	1.63×10^{-2}	7.58×10^{-2}	110.1	6.51	3.67×10^{-3}

A more detailed analysis of the results is provided by examining the weekly outcomes for one of the test weeks. This includes an exploration of the cost evolution, generation dispatch, and a comparison between predicted and actual prices. The additional discussion and visualizations are presented in section E.4.

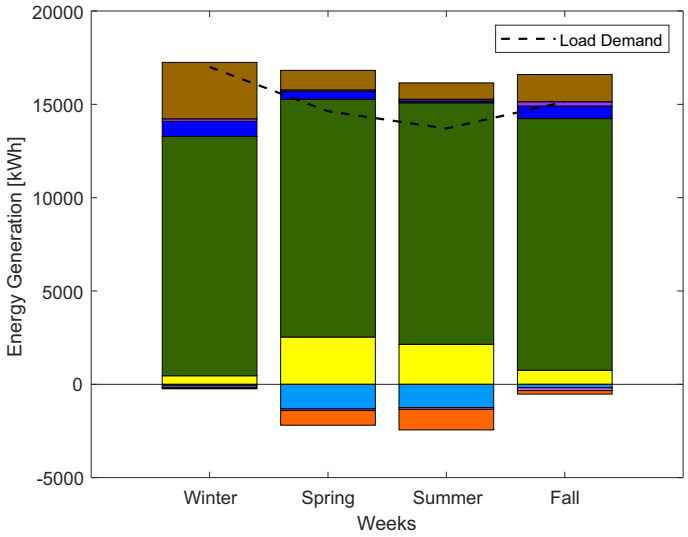


Figure 7.18: Energy generation share and load levels for each week. Where yellow, green and blue represent the PV, wind and hydro generation, respectively, light blue represents pumping periods, brown buying periods and orange selling periods. In turn, purple represents discharge periods, and lilac periods of charging the battery bank.

7.4 Generation Forecasting

Generation forecasting has become an increasingly vital task in the modern context of electricity markets and the growing penetration of renewable energy sources [617]. For energy producers — especially those operating non-dispatchable generation units such as photovoltaic (PV) and wind power—accurate forecasting is deemed essential. These producers must estimate their expected generation output in advance to participate effectively in market bidding processes, where they submit offers to sell electricity. Having reliable forecasts allows them to define confidence intervals around their expected generation, helping to mitigate risks associated with under- or over-delivery and to comply with market regulations.

This forecasting task is inherently challenging due to the intermittent and weather-dependent nature of renewable energy sources. As a result, generation forecasting has become a field largely dominated by PV and wind forecasting methodologies, which aim to predict short-term and medium-term output based on a variety of environmental and operational factors. However, with the exploration and emergence of other renewable energy sources over the past decade, there has been a notable proliferation of forecasting approaches extending beyond traditional PV and wind cases, encompassing other promising RESs — including ocean-related technologies such as wave and tidal energy.

Therefore, in the following sub-sections, and considering these recent developments, we will briefly explore the potential of Ocean Wave Energy, followed by a proposed methodology applied to real-world forecasting data.

7.4.1 Ocean Wave Forecast: A state-of-the-art

In the ever increasing quest to develop alternative and non-conventional renewable energy technologies to effectively tackle the increased demand for energy, escalating energy prices, and the environmental harms associated with fossil-based electricity production, researchers have made major efforts, over the last few decades, to develop and to facilitate the use of renewable energy sources (RES) [618]. These concerted efforts aim to enrich the increasingly 'particolored' energy portfolio. With this in mind, ocean renewable energy (ORE) stands out as one of the fittest contenders, with densities several times greater than wind and solar energies [619] coupled with virtually endless flows guaranteeing continuous energy availability (relatively high availability and utilization factor), i.e., with the possibility to operate "24/7" [620, 621]. These inherent characteristics of wave energy unveil an enormous potential, and several estimates suggest generated electricity figures ranging between 8 and 80 PWh per year [622, 623].

On the other hand, it is important to recognize that development costs, lack of convergence, potential impact on biodiversity, transmission requirements, and above all the (operational) harsh environment has delayed the commercialization and wide adoption of these emerging RES in comparison with more mature technologies [624]. Despite these challenges, the inherent qualities of ocean wave energy, together with the sizable RD&D investments on wave energy converters, particularly in the EU [625], allow us to anticipate a successful future integration of ORE resources into the grid. Hence, contributing to the ambitious targets in terms of RESs generation share, which the EU estimates to be at least 27% by 2030 and much higher by 2050 [626].

In this foreseeable scenario and following the lead of more mature technologies, predicting the energy production from RES, and in particular wave conditions/wave power, is critical to maintaining a stable electrical grid due to its intermittent and non-dispatchable nature of these generation technologies [626,627], especially as the amount of energy generated from these sources increases [628]. Additionally, these predictions are also key for project planning, decision-making and risk management of market players [629] and developers. Hence, if wave energy technologies are expected to play a significant role in a future cleaner energy mix, then it is crucial to have accurate estimates of the potential energy production in order to lay the foundations for suitable wave energy converters [630, 631].

However, the benefits of developing accurate forecasting models go far-beyond the latter, and the list includes aquaculture industry, optimize shipping routes, nautical sports, military operations, etc [627]. Consequently, ocean wave forecasting essentially focus on the challenging task of predicting characteristic wave heights and periods using readily available weather conditions [632].

The future role of wave energy technologies coupled with all the different applications of wave energy forecasting are the reason behind an increasing number of works in this field. This follows a broader trend in the renewable forecasting literature, which is attributed to the intense worldwide renewable integration efforts [617]. It also constitutes the main motivation of the presented work.

A literature survey reveals three conventional categories of methods to perform this task, namely: i) Numerical approaches primarily centered on wave energy-balance relationships [633]. For example, authors in [634, 635] used the popular wave numerical model WAVEWATCH III (WW3) to perform the forecasting task of ocean wave energy related variables, other popular choices are Simulating Wave Nearshore (SWAN) and Community Climate System Model (CCSM4) models [632, 634]. Even though wave numerical models offer a good accuracy, they tend to be computationally expensive, therefore, opening path for machine learning (ML) approaches [627, 632]. ii) Statistical methods such as regression, Box–Jenkins method applying Auto Regressive Moving Average (ARMA) [636] and Auto Regressive Moving Average (ARIMA) models, exponential smoothing, among others are a common choice to solve a time-series forecasting task. However, a common downsize is the difficulty to capture non-linear data patterns [637]. iii) Soft-computing approaches are amongst the trendiest options due to their ability of efficiently mapping large datasets and deal with non-linearities in the predicted and auxiliary time-series. The list includes trendy ML and Deep Learning (DL) based methods, fuzzy logic, expert systems, decision trees, genetic algorithms, and hybrid approaches [630, 638].

A particular highlight can be given to ML approaches, which overcome several shortcomings of traditional data analysis techniques. Thus providing accurate and efficient methods for predicting ocean related variables [639]. In this field, artificial neural networks (ANNs) [163, 640] and support vector machines (SVMs), multi-layer perceptron (MLP) [627, 632], Nonlinear autoregressive network with exogenous inputs (NARX) [633], recurrent neural networks (RNNs) [641], Random Forests (RF) [642], Extreme Learning Machine (ELM) [643], are among the most common choices used in forecasting.

An example of a hybrid approach is given in [644], where a Genetic Fuzzy System built on a three- staged algorithm is used to estimate significant wave height and energy flux. In another instance, a hybrid Grouping Genetic Algorithm - Extreme Learning Machine approach is proposed in [645]. Alternatively, a wavelet and neural network hybrid model was developed in combination with empirical orthogonal function analysis to better the performance of traditional ANNs [646]. In another instance a comparison between single and multi-layered RNN, LSTM, and seq-to-seq networks is conducted in [647] and in [648] ocean wave height is also predicted using a LSTM network, proving its superior accuracy in comparison with a MLP network, SVM and RF.

7.4.2 Wave power and sea state parameters

The main considerations about sea state parameters and wave power required to understand the predictable variable in the short-term wave power forecasting, as well as the auxiliary associated variables, are presented in this sub-section. Afterwards, the wave data employed in this work is illustrated together with the main statistics.

7.4.2.1 Wave power estimation

In the last few decades, climate models and particularly wave models were refined to produce a greater understanding of the different physical interactions responsible for wave phenomena, providing useful knowledge for public and private entities to better monitor and explore ocean resources, e.g., when we explore the potential of wave energy converters in certain sea state conditions. To achieve this goal, meteorological, sea and atmospheric agencies across the globe, such as National Oceanic and Atmospheric Agency (NOAA), have greatly expanded scalar buoy and nautical radar measurements of ocean waves (wind generated) to better study the sea state spectral parameters estimated from both sensors [649].

In this work, we are solely interested in the set of sea state parameters from where it is possible to derive a wave power expression. So, we define the mean transport rate of the wave energy through a vertical plane of unit width, parallel to a wave crest, as the wave energy flux or more broadly, wave power, J [kW m^{-1}] (Equation 7.13). And in deep water conditions, i.e., where the water depth is larger than half the wavelength, the same can be computed as follows [650]:

$$J = \frac{\left(\frac{\rho g^2}{64\pi} H_{m0}^2 T_e\right)}{1000} \approx 0.491 \left(H_{m0}^2 T_e\right) \quad (7.13)$$

where H_{m0} is the significant wave height; T_e the wave energy period; ρ the saltwater density; and g the acceleration due to gravity. When the significant wave height is given in metres (m) and the wave period in seconds, the result is the wave power in kilowatts (kW) per meter of the wave front length.

7.4.2.2 Wave Data

In this work, three different hourly datasets will be used to evaluate the forecasting performance of CNNs. These comprise a set of quality-controlled buoy measurements gathered by the US National Data Buoy Centre (NDBC), spanning between January 2007 and April 2010. The respective buoy IDs are: 41048 in the West Bermuda; 42056 Yucatan Basin between the Caribbean Sea and the Gulf of Mexico: 44014 Virginia Beach, Virginia. The focus of this work will be to predict the wave power or wave energy flux, J, however, the variables of interest to assemble the input data also include H_{m0}, T_e and windspeed, v.

With this regard, it is important to analyse the time-series statistics, and therefore Table 7.12 is presented, showing the mean, upper and lower occurrences, standard deviation, and the shape related statistics, kurtosis, and skewness.

These stats show a diversified scenario between datasets, which will help validate the forecasting performance over a wide range of sea state conditions. However, a commonality between the datasets is the greater standard deviations with highly right-skewed data and with a leptokurtic distribution. Moreover, the use of these specific datasets allows us to

Table 7.12: Statistics of the different wave power datasets $[\mathrm{kW\,m^{-1}}]$.

Buoy ID / Stats	Mean	Min	Max	St. Dev	Skew	Kurt
41048	18.19	0.72	461.62	24.93	3.93	28.44
42056	6.27	0.25	184.45	8.08	4.88	50.32
44014	12.42	0.39	261.90	19.73	4.33	29.42

benchmark the obtained results against several conventional approaches also considered in [163, 651].

A visual characterization of the energy density in terms of significant wave height and energy period is shown in Figure 7.19 for all the datasets.

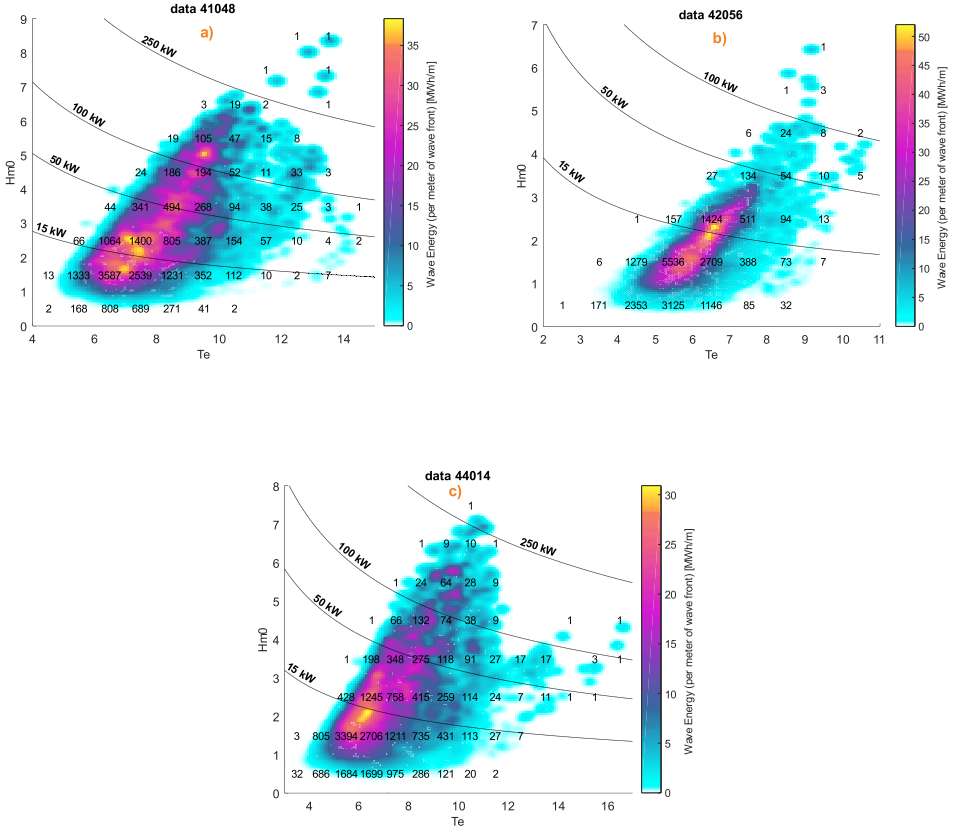


Figure 7.19: Kernel joint probability density estimation of the collected observations. The colour scale represents the wave energy per metre of wave front (in MWh/m). The numbers within the graphs indicate the occurrence of sea states (in number of hours per year) and the isolines refer to the wave power J (in kW).
(a) Buoy ID: 41048, (b) Buoy ID: 42056, (c) Buoy ID: 44014.

This allows us to clearly see the variability of sea state conditions, the absolute frequency of different sea state conditions (pair H_{m0} and T_e), as well as the associated (available) wave density and power. Thus, validating the representativeness of the considered datasets.

7.4.3 Methodology: theory and implementation

As the name suggests, this section introduces the important CNN design and implementation details, as well as the different stages of the proposed forecasting approach. Starting with the fundamentals of the Convolutional Neural Network topology, followed by the required adaptions to use it in the context of one-dimensional space data for a prediction-based problem, detailing the assemble stage of the input training and target samples.

7.4.3.1 Convolutional Neural Networks

The development of DL approaches has expanded past conventional classification problems, and the vast landscape of ML and DL techniques are being applied across various research fields. As we saw in the sub-section 7.4.1, the same is valid for wave power forecasting, where these approaches are gaining wide traction, much due to their ability to model complex non-linear relationships between labeled training sets.
With this respect, CNNs or ConvNets are one of the most promising and popular DL approaches [652, 653]. Originally conceived for large-scale image classification, where they have been tremendously successful [654], CNNs are a type of bio-inspired feed-forward type of ANN architecture, that eliminates the need for manual feature extraction [653], and is particularly tailored for analysing and finding patterns in two-dimensional data. However, the application scope of CNNs extends beyond image data, and it has been successfully applied in signal processing and time-series analysis (1D data), audio analysis (2D data) and video classification (3D data), etc. CNNs consist of a sequence of layers and its core building blocks are the Convolutional (ConvLayer) and Pooling (subsampling) layers which have different roles during the training process, as illustrated in Figure 7.20 [652, 655].

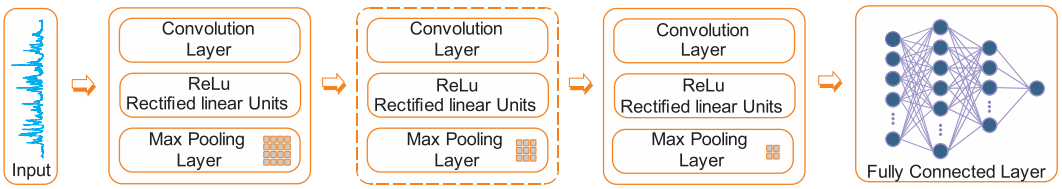


Figure 7.20: Convolutional Neural Network (CNN or ConvNet) architecture.

Convolution layers are composed by neurons with weights and biases which are then continuously updated (each training epoch), to minimize a loss function. However, for a given hidden layer the value of the weights and biases is the same for all neurons, which is a major contrast with the standard fully connected networks. Another major difference is that instead of a fully connected architecture, connections are local in space (across the width and height of the input volume) but always full along the entire depth of the input volume, i.e., there is an asymmetry in how spatial and depth dimensions are handled. The spatial extent of the local connection is called receptive field (inspired in the biological visual cortex) or filter size.

So, a ConvLayer contains an entire set of filters, each one producing an individual activation map which is stacked along the depth dimension resulting in the output volume, i.e., each filter is convolved across the width and height of the input volume. These (learnable) filters are sparse and shared across the entire input and are adjusted during the training process, in order to be activated when certain features are detected. Thus, we can define in Equation 7.14 the output result/feature map $y_k^{(m)}$ from the convolution operation that occurs in an arbitrary ConvLayer k, between an input from an arbitrary layer $k-1$, $x_{k-1}^{(c)}$, and the m^{th} filter of such channel c in layer k, $W_k^{(c,m)}$, followed by the added bias term $b_k^{(c)}$. Being f_j the chosen activation function [656].

$$y_k^{(m)} = f_j\left(\sum_{c=1}^{C} W_k^{(c,m)} * x_{k-1}^{(c)} + b_k^{(c)}\right) \quad (7.14)$$

Additionally, two common operations performed that influence the outcome of the convolution between the input and the filter/kernel are Padding and Striding. The first, padding p, defines the number of "pixels" added to the input (extremities), and it avoids a sizeable input shrinking and reduces the discrepancy in the treatment of central vs corner "pixels". The second, stride s, describes what is the step taken across spatial dimensions while consecutively convoluting filters with the inputs. An increased stride will lead to a reduced output dimension, which is the opposite result of padding. So, the output of each convolved input volume is used as the input to the next hidden layer, and for an arbitrary input size $[n_h \times n_w \times n_c]$ and filter size $[f_1 \times f_2]$, the generalized output dimensions of a ConvLayer with n_f filters are:

$$\left[\left(\frac{n_h + 2p - f_1}{s} + 1\right) \times \left(\frac{n_w + 2p - f_2}{s} + 1\right) \times n_f\right] \quad (7.15)$$

where subscripts $h,\ w$ and c represent height, width, and channels (depth), respectively.

In turn, a Pooling layer, is responsible for a nonlinear downsampling of the input volume, thereby, reducing the number of learnable parameters and hence speed up the computation cost. To execute the downsampling, a max or average Pooling layers are commonly used, and in the same fashion as ConvLayers, Pooling layer's hyperparameters also include kernel size, stride, and padding, thus the output dimensions are obtained similarly to Equation 7.15, with exception of depth that retains the size of the input volume that is forward to the pooling layer. Other common hidden layers in a ConvNet are Normalization layers, i.e., which normalize the output of the previous layers to improve the batch learning process; Flatten layers that, as the name suggests, flatten/collapse the input dimensions $[n_h \times n_w \times n_c]$ to a one-dimensional output, allowing the transition from convolution layers to a traditional Dense/fully connected layer. As for the output layer, it can be either categorical or continuous.

7.4.3.2 Proposed Methodology

In order to implement a CNN capable of predicting wave power for different forecasting horizons, we start the process by analysing the partial auto-correlation magnitudes (PACF), as well as the Pearson correlation, for the forecasted variable J, and the exogenous inputs H_{m0}, T_e and w_s. Besides, the hourly index of the corresponding target output J is also fed into the CNN to provide some account of the observed daily seasonalities. Since even though a CNN does not require a manual feature selection, due to its deep learning capabilities, it is still important to only feed input samples and parallel input variables with a significant correlation. In this work, a threshold level of 0.05 was considered. This avoids a more time-consuming training task and reduces the potential length (height) of the input samples, which can impact the kernel size. An illustration of this pre-processing and data selection stage is shown in Figure 7.21, where after performing a correlation analysis (auto and partial correlation) of the different time-series, the correspondent training and testing samples are assembled with its pairing target data.

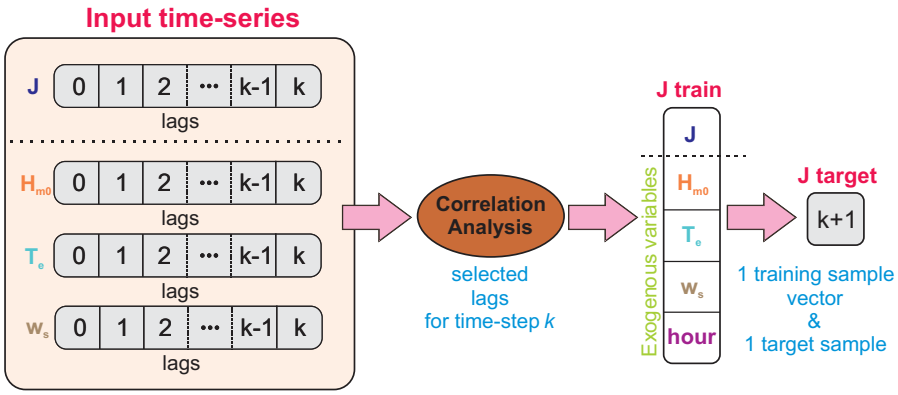


Figure 7.21: Wave pre-processing and input data selection.

Furthermore, it is important to notice that although CNNs are traditionally employed with 2D data, as we have seen in sub-section 7.4.3.1, they can also be used to model univariate time-series. The difference resides in the kernel size, which has the same width as the time-series, while the height may differ. In this way, the filter moves only across one dimension. In the event of a multivariate time-series, as is the case in this work, the several inputs are fed in parallel, analogously to the different colour channels in the 2D data, meaning they are treated fully along the depth dimension, but locally in terms of spatial dimensions. To sum up, a 1D CNN completes a 1D convolution operation and is followed by a dimensionality reduction, combining the feature extraction, transformation, data fusion and classification steps into one framework [657].

So, after having assembled the input data with exogenous variables, in the form of input samples with the corresponding target output, we moved on to building the forecasting CNN architecture. As can be seen in Figure 7.22 it has two ConvLayers, which is a preferred option over a single ConvLayer when dealing with long input sequences. The filters of these hidden layers will be activated, to signal certain input features. Then, the

customary non-linear activation function Rectified Linear Unit (ReLu) is called to map the convolution operations. This function has the particularity of only keeping the positive values (activated features), since negative values are set to zero. After those, two max Pooling layers are used to condense the output (feature maps) of ConvLayers, maintaining only the most salient values. These hidden layers are followed by a Flatten layer to compress the output volume to a single dimension vector. In turn, this will be the input to a Dense (fully connected) layer, whose task will be to interpret the features extracted by the convolution part of the architecture and to produce a single valued output. The Dense layer has also a Dropout rate of 20%, in order to improve its generalization capabilities.

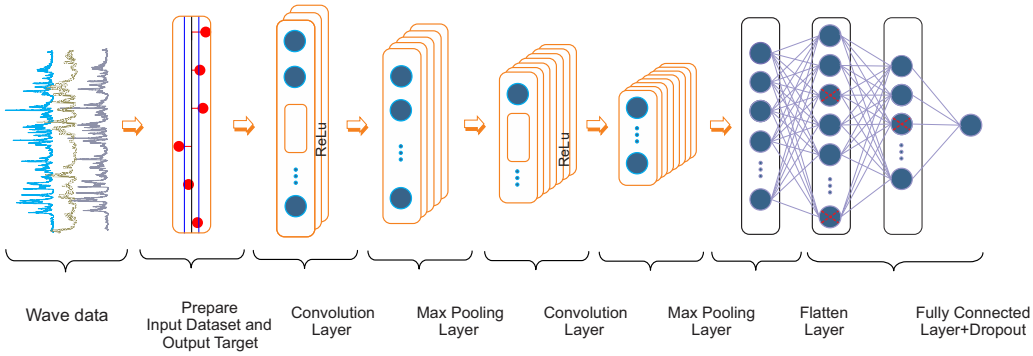


Figure 7.22: Proposed Methodology (Overview): CNN architecture fed with several 1D wave input data.

To implement the aforementioned CNN, the authors have resorted to Python 3.8.7, using the TensorFlow and Keras deep learning libraries and simulations were performed on Windows 10® with an Intel Xeon E5-1620 v4 CPU.

Additionally, regarding the input data size, we have considered the previous 30 days before the test week, to assemble the training input samples. This value enables a compromise between diversity in the training samples and reasonably recent data, i.e., a batch of roughly 720 samples, with a validation ratio of 15%. With these training and validation samples, the ConvNet will learn to map the input sequences, and each layer will be capable of learning to identify different features. To achieve this goal, the weights and biases are learned by minimizing the chosen loss function, which in this work was the MSE (as seen in sub-section 5.3.2).

With respect to the task of training the CNN, the Adaptive Moment Estimation (Adam) was chosen. This stochastic optimizer computes the individual adaptive learning rates for different parameters from estimates of first and second moments of the gradients, for each learning epoch [438].

7.4.4 Short-term Ocean Wave Forecasting

The proposed methodology is used to perform the short-term wave power forecast, J, for the three datasets, considering lead times of 1, 2, 3, 6, and 12 h.

For validation reasons, four weeks of testing are considered, one week per calendar season

of the year (northern hemisphere), which ensures that the results are not tied to a specific seasonality or trend. The testing dates for each dataset can be seen in Table 7.13.

Table 7.13: Forecasting: Testing Dates.

Buoy ID / Season	Winter Week	Spring Week	Summer Week	Fall Week
41048	07/02/08–13/02/08	01/05/08–07/05/08	12/08/08–16/08/08	22/10/08–28/10/08
42056	07/02/08–13/02/08	15/05/08–21/05/08	19/08/08–25/08/08	12/11/08–18/11/08
44014	31/01/08–06/02/08	08/05/08–14/05/08	22/07/08–28/07/08	29/10/08–05/11/08

Additionally, for each testing date, the CNN model was evaluated (trained and tested) for 15 times, i.e., 15 runs, this allows us to have a more solidified conclusion about the model performance. To evaluate the forecasting performance we employed common error metrics, namely the correlation coefficient (r), mean absolute percentual log-difference error (MAPElog), and root mean squared error (RMSE), as seen below in Equations 7.16-7.18.

$$r = \frac{\sum_{i=1}^{N_s} [(p_i - \overline{p}) \bullet (o_i - \overline{o})]}{\sqrt{\sum_{i=1}^{N_s} (p_i - \overline{p})^2 \bullet \sum_{i=1}^{N_s} (o_i - \overline{o})^2}} \quad (7.16)$$

$$\text{MAPE}\log\,[\%] = 100 \bullet \frac{\sum_{i=1}^{N_s} |\log p_i - \log o_i|}{N_s} \quad (7.17)$$

$$\text{RMSE}[kW/m] = \sqrt{\frac{\sum_{i=1}^{N_s} |p_i - o_i|}{N_s}} \quad (7.18)$$

where N_s is the number of testing samples; $\overline{p}$ and p_i are the average (for all N_s samples) predicted wave power and the predicted wave power at time-step i, respectively; and $\overline{o}$ and o_i are the average real wave power and the real wave power at time-step i, respectively. After completing the simulations, the mean forecasting results (r, MAPElog and RMSE for the 3 datasets, for each forecasting horizon, is compiled, see Tables 7.14-7.16.

Table 7.14: Forecasting errors: buoy 41048 (mean results).

For hor/ season	Winter Week			Spring Week			Summer Week			Fall Week		
	r	MAPE log	RMSE	r	MAPElog	RMSE	r	MAPElog	RMSE	r	MAPElog	RMSE
1h	0.96	14.40	6.22	0.95	14.34	4.51	0.94	15.48	3.04	0.97	11.12	3.19
2h	0.95	17.57	8.11	0.94	16.11	4.19	0.92	17.84	3.43	0.95	13.98	3.91
3h	0.93	21.49	9.36	0.92	18.82	5.18	0.89	20.29	4.01	0.92	17.16	4.92
6h	0.92	28.19	8.92	0.72	28.30	6.38	0.80	33.26	5.68	0.82	31.68	8.38
12h	0.85	38.91	12.32	0.29	52.81	12.66	0.69	39.05	6.88	0.54	51.84	19.75

So, by analyzing the tables we can see that naturally the forecasting accuracy decreases as the forecasting horizon increases, with error metrics, in terms of RMSE, increasing in comparison with the previous lead times, in the order of 15%–20% (between 1 and 2 h), 18.0%–19% (between 2 and 3h), 25%–50% (between 3 and 6 h), and 35%–57.4% (between 6 and 12). Moreover, there is an obvious connection between the prediction

Table 7.15: Forecasting errors: buoy 42056 (mean results).

For hor/ season	Winter Week			Spring Week			Summer Week			Fall Week		
	r	MAPElog	RMSE	r	MAPElog	RMSE	r	MAPElog	RMSE	r	MAPElog	RMSE
1h	0.97	10.10	0.78	0.95	10.05	0.27	0.97	12.31	0.27	0.94	12.18	2.52
2h	0.95	12.88	0.95	0.93	12.65	0.34	0.95	15.60	0.34	0.93	15.43	2.85
3h	0.93	16.01	1.13	0.90	15.45	0.42	0.95	14.66	0.39	0.90	18.74	3.37
6h	0.87	22.33	1.47	0.75	24.67	0.69	0.87	27.52	0.67	0.84	26.94	4.61
12h	0.67	31.60	2.21	0.40	45.08	1.70	0.84	24.15	0.78	0.69	41.76	7.02

Table 7.16: Forecasting errors: buoy 44014 (mean results).

For hor/ season	Winter Week			Spring Week			Summer Week			Fall Week		
	r	MAPElog	RMSE	r	MAPElog	RMSE	r	MAPElog	RMSE	r	MAPElog	RMSE
1h	0.97	13.80	5.84	0.97	19.21	14.96	0.91	15.79	0.43	0.93	21.46	3.73
2h	0.94	18.63	6.62	0.93	26.09	18.19	0.84	21.72	0.58	0.88	29.03	4.98
3h	0.93	23.58	6.68	0.93	35.14	20.51	0.73	26.79	0.75	0.84	39.05	5.94
6h	0.84	39.52	8.91	0.86	52.23	28.95	0.53	33.27	0.96	0.72	57.20	8.63
12h	0.68	57.95	10.64	0.80	68.79	36.15	0.41	43.29	1.20	0.51	84.89	12.89

accuracy and the season of the year, where the CNN testing is being conducted since, as we saw in Table 7.12, there is a great standard deviation in the data, with significant higher wave power during the winter and fall seasons.

Commencing with buoy data 41048, Table 7.14 reveals a relatively uniform accuracy across all seasons, with the exception of 12h horizon, with error metrics of 0.97–0.29 in terms of r, 11.12%–52.81% in terms of MAPE log and 3.19-19.75 kW/m in terms of RMSE. These results are in line with the regression and neural net models overall means in [651], the same dataset, with the exception being the 12h horizon, where the accuracy falls by 50% in terms of RMSE. Therefore, demonstrating the need to use combined models in this dataset, i.e., it would heavily benefit from the physical model "guidance". Now by analyzing the buoy 42056 (Table 7.15), error metrics vary in the range 0.97–0.40 in terms of r, 10.05%–45.08% in terms of MAPE log and 0.27-7.02 kW/m in terms of RMSE. In this case, we see an improvement in all forecasting horizons, when we compare them with the ones obtained by the physical models, regression, and neural network. As for buoy 44014 errors (Table 7.16), the error metrics average is in the range 0.95-0.60, 17.6%-63% and 6.24-15.22 kW/m, for r, MAPE log and RMSE, respectively.

In turn, the distribution (box-and-whisker plots) of the error metrics during the runs was also analysed, considering all seasons, and forecasting horizons, and Figure 7.23 is presented to demonstrate the behaviour of the CNN during the runs for buoy data 41048.

It is revealed that the employed ConvNet model presents a generally stable forecasting accuracy (a smaller interquartile range), certainly for the first 3 horizons; then from 6h onwards the error spread increases, which translate an added difficulty in extrapolating relevant information from the available input dataset. Additionally, there were no significant outliers.

A graphical representation of the short-term wave forecasting is given in Figure 7.24, where we can see the real and predicted wave power (left axis), as well as the absolute error (right axis), for all the considered forecasting horizons, regarding buoy data 41048 (spring testing week).

The effectiveness of the proposed method was evaluated in 3 different case studies, from

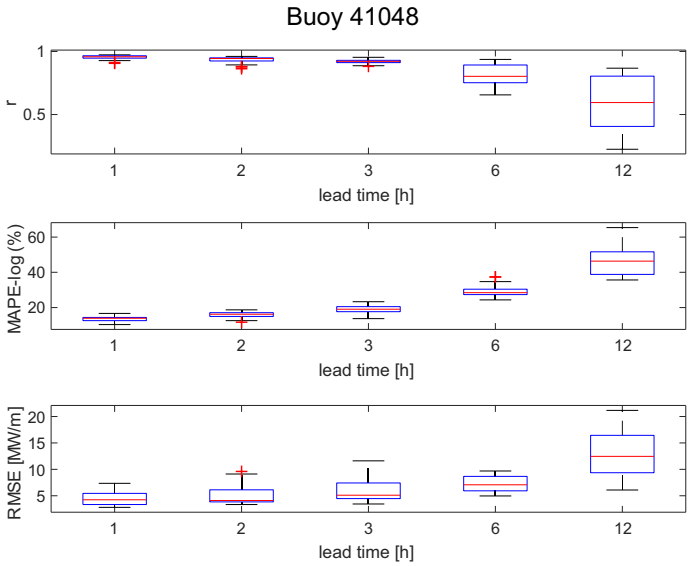
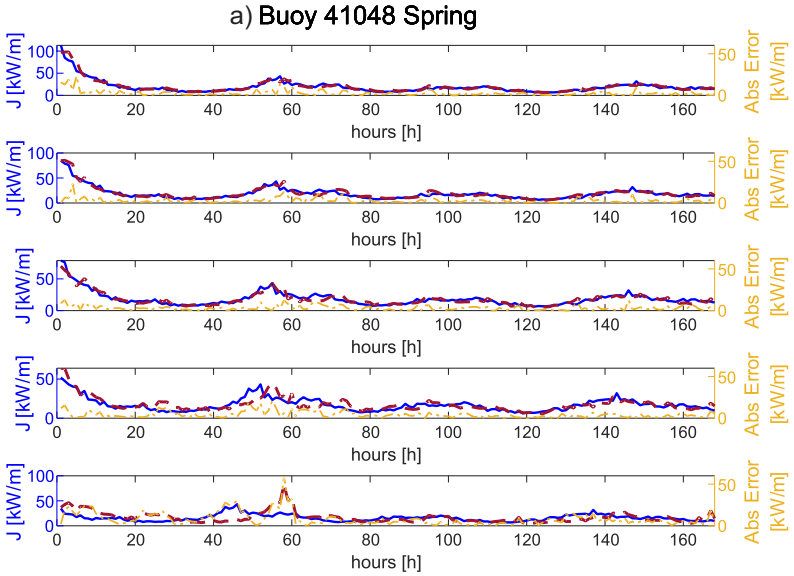


Figure 7.23: Box plots of the error metrics distributions, considering all forecasting horizons (buoy 41048).

NDBC buoy measurements on different sites. The proposed approach is versatile, proving the viability of CNNs, avoiding the need for time-consuming data pre-processing mechanisms. Numerical results reveal good performances, especially for horizons between 2h and 6h, where the CNN model can sometimes be quite fitting with a relatively small input training set (30 days). Meaning, average accuracies across all seasons ranged from 11% to 30%. As for the furthest forecasting horizon, the CNN model was sometimes unable to perform as well, which is normal in non-combined models, i.e., without information from a physical (numerical models).



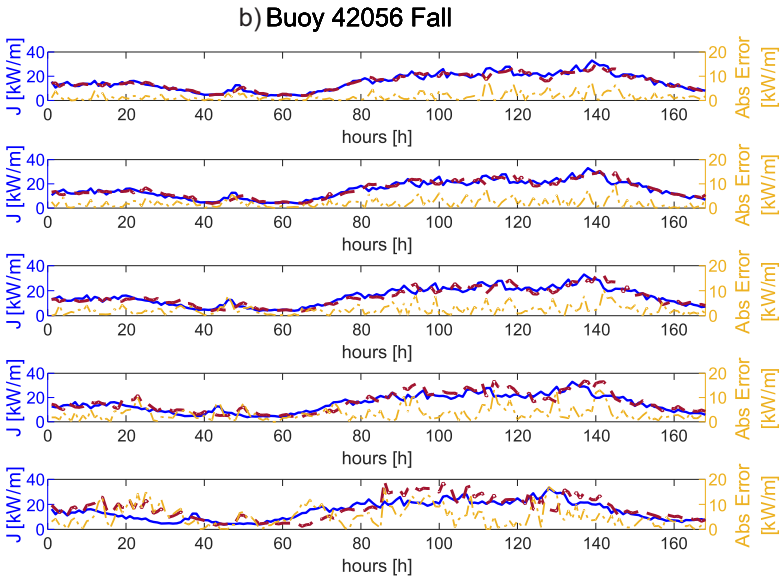


Figure 7.24: Test Results illustration for all horizons: Actual and forecasted wave power flux, J, respectively, blue solid line and maroon dashed line (left axis) and absolute error, canary yellow dot-dashed line (scaled right axis) ample graph with blue (dotted), green (solid) and red (dashed) lines
(a) Buoy 41048 Spring Week, (b) Buoy 42056 Fall Week.

Chapter 8

Conclusions

Reaching the concluding section of this extensive research marks the culmination of a long, winding odyssey of profound learning and rigorous discussion. This journey, a noble quest for what remains the holy grail of any academic endeavor, has taken us well beyond what we originally conceived — both in terms of scope but also in skills development — when devising this thesis's work plan and hypothesizing its title with my advisor.

With that said, and before delving into the more substantive findings of this thesis, it is crucial to begin with a reflective acknowledgment of this intellectual and collaborative expedition. This journey, though initiated with a clear plan, like any vast endeavor, naturally evolved through unforeseen research avenues. It demanded adaptability and a redirection of focus at several points, particularly evident in our broad exploration of energy policy and, in particular, the issue of system flexibility. Moreover, these new avenues led us to investigate the still elusive and emerging wave energy potential within the field of forecasting. This was a deliberate pivot from limiting our initial analysis to the more conventional paths of electricity market evolution and its connection with large-scale VRE integration, or more conventional EPS optimization and prediction-related avenues, such as the already exhaustive realm of load forecasting studies or the ED problem.

Furthermore, the very structure and writing style of this thesis, evident from its earliest stages, evolved in a highly instinctive manner. This approach reflects not only the author's own evolving understanding but also the synthesis of diverse contributions and facts, ranging from traditional journal papers, conference proceedings, and book chapters to the less orthodox, expert opinions, blog posts, and code repositories. While this more comprehensive style permeates its chapters, it simultaneously strives to be inherently self-explanatory. Put differently, and through an analogy familiar to any electrical engineer, this thesis is organized in a way that, much like electrical current, favors the path of least resistance.

Now, moving from these introductory and overarching considerations of the author's experience in developing this thesis, we can now turn our attention to the more specific and substantive conclusions drawn from this work contributions, in the section below.

8.1 Why the starting point matters

Though its pervasive presence often goes unacknowledged, electricity has become an integral commodity in our societal development, with its ease of access and use strongly correlated with economic progress. This development, throughout the years, has necessitated continuous expansion and process adjustments.

The consolidation of electricity markets was, without a doubt, a pivotal driver, fundamen-

tally changing the paradigm and interface connecting generation and consumption. This is a subtle yet critical point: from this stage onwards, energy policy, legislative frameworks, resource availability studies, planning and grid expansion, pricing mechanisms, demand elasticity, and even consumer perception have become significantly more complex, encompassing numerous ”moving parts”. These parts not only compete but are also governed by the promise of opportunity and welfare within a supposed level playing field. In this intricate environment, decision-making entities —ideally supported by robust academic consensus — must ensure a necessary balance of profitability, innovation, and social and environmental responsibility.

One of the most pernicious challenges facing the rich history of EPSs — a challenge that has certainly been heightened over the last decade — is the consensus regarding an urgent need to implement a transition. This is particularly evident in the evolution of the generation portfolio and the growing dependence on electricity as the primary end-use energy source. These shifts are crucial to addressing the environmental harms associated with more traditional energy sources. This constitutes the very premise of this thesis: the large-scale integration of variable renewable energies in electric power markets.

With that in mind, it was essential to begin this thesis with an introductory chapter that addressed some of the major energy security propositions concerning the operation and infrastructure of EPSs, as well as the role of different electricity markets. For instance, particular emphasis was placed on a topic central to the variability discussion: as large volumes of variable and intermittent RESs are brought online, the capacity to significantly queue and store electric energy remains limited — unlike most other commodities. This is especially relevant, as it restricts the options available for balancing supply and demand — which, in any way or form, must not compromise what is not merely a perceived convenience, but a guaranteed feature: instant availability for the end-user.

Having reviewed the non-negotiable principles that underpin the current functioning of EPSs — principles often discussed superficially and without operational grounding by energy policymakers — the next step was to examine, with precision, some of the structural lock-ins that hinder the transition toward a predominantly renewable portfolio. This analysis serves to clearly identify the research gap and, consequently, to establish the motivation for proposing contributions that directly address these challenges.

This intricate EPSs environment—characterized by an increasing number of interconnected and heterogeneous components—translates into a higher degree of uncertainty. This uncertainty is not only reflected in electricity market price volatility but also imbues a broader set of operational challenges, particularly in the ever-expanding domains of machine learning and optimization.

As such, it becomes essential to investigate more focused, concrete models and challenges faced by different stakeholders. This includes the critical role of forecasting diverse time series and applying optimization techniques to core EPSs problems such as the energy dispatch or OPF, among others. These efforts are not only about leveraging today’s most powerful data-handling tools but also about gaining a deeper, more grounded understanding

of how the path toward large-scale VRE integration is far more labyrinthine than straightforward at the micro (task/problem) level (contrasting with the mainstream narrative). Or, to return to our analogy: sometimes, current meets semiconductors — or even insulators — not just metals, and so something (often) must give (i.e., a compromise) — just as voltage drops when current flows through these less conductive materials.

8.2 Summary of Key Findings

The rapid integration of large volumes of VRE capacity has become a hallmark of many developed EPSs. The Iberian power system has excelled in this regard, corroborating the ambitious EU energy policy aimed at fostering a cleaner generation portfolio by leveraging the abundant endogenous resources of both countries. In parallel, with this fast-track adoption of renewables, regulators have been adapting the regulatory and fiscal frameworks to target carbon-intensive energy sources, including most of the existing fleet of dispatchable thermal capacity.
However, one lesson that societies should have learned in the last couple of years, first with the COVID-19 pandemic, the beginning of an inflationary environment in 2021, which was then aggravated by the war in 2022, is that these assumptions can change quickly. Hence, before rushing energy policy decisions to achieve a very high level of renewables in the mix, it is necessary to understand not only the set of operational and market challenges posed by the current penetration levels but also to carefully assess the crucial role of dispatchable generation in the overall system, that is, better planning of how and when to phase-out thermal plants, and what solutions can perform a similar role in a future grid.

- The first conclusion of the study is that the continuation of a narrow strategy, where wind and PV systems are added in large amounts, coupled with a few investments in biomass and cogeneration, not only requires sizable over-dimensioning, given the relatively low-capacity factors of the first two RESs; more importantly, it cannot by itself replace some of the beneficial traits of fired-generation to ensure the security of supply, price competitiveness, and power system reliability.

- Second, with that in mind, it is crucial for Iberian policy makers (”workhorse” case study), regulators, market participants and other involved parties, to understand that a one size fits all is not adequate. Several options should be considered, to find the “right” portfolio of technologies that reinforce system flexibility in a scenario dominated by RESs and in an environment where electrification is gaining momentum. Advisedly, a system permeated by the heterogeneity of primary sources is more robust in handling the ordinary seasonal variations of these sources. However, it also protects the system and the end user from sudden and/or unwarranted shocks that may affect some of them (as witnessed in the last two years).

- The third major outcome of this research work is a comprehensive overview of the main action vectors to achieve the goal: storage systems, interconnections, differ-

ent renewable technologies, demand response mechanisms, and modern and more efficient thermal generation technologies. Building on the extensive literature on these action vectors, and based on the specificities of the Iberian power system, this work presented and discussed the most tailored technologies and associated mechanisms; discussing the main functioning principle, strong suits of each, and generic LCOEs, which resulted in a thorough examination of the most promising solutions, complemented by a brief mention of illustrative real-life projects involving all the vectors of action and/or policy targets in both countries.

- A leading issue with the prevailing approach to large-scale VRE integration — at least as reflected in public discourse and perceived policy goals — is that it is often framed as a ”sprint” transition, pursued regardless of the associated ”costs.” However, from the perspective of such an intricate and interconnected system, with its inherent entropies and complexities, planning and system flexibility are two vital dimensions that must not be overlooked if the integrity of any EPS is to be preserved.

- It is also important to rethink the interaction of producers and consumers with the electricity market, including how they react to profound regulation changes, to better protect the final consumers from unsolicited effects. Besides, it is important for legislators to acknowledge the uniqueness of each national power system, and to promote an integration that preserves fair ruling, stability, and predictability to achieve a more competitive interconnected and secure electricity market. As a result, this work establishes important considerations that assist in the quest of lessening the mentioned disconnect.

- Given the augmented difficulty in regression and optimization type problems in EPSs sub-fields, it is a must to look and explore in detail the vast ”substratum” of ML models and optimization related resources.

- Metaheuristics have proven particularly well-suited for adjusting OPF decision variables — namely, generator outputs, transformer tap settings, and shunt capacitances — especially in scenarios requiring scalability and a more straightforward approach to handling sub-optimal solutions.

- In the forecasting context, traditional assumptions or pre-analysis metrics, including bi-variate correlations, should only be taken as an indicative figure, and a sensitive analysis with the chosen forecasting engine is always recommended. Moreover, different input combinations have excelled in different testing months, underlining how can they be complementary, and so, why is it necessary to tailor the input to the chosen case study data.

- The strategy of fostering input diversity among base learners — achieved through the use of distinct Box-Jenkins models — proved to be a promising design choice. These models were selected and evaluated from a pool of candidates based on data-driven

criteria and information-theoretic measures. This approach effectively addressed the common challenge of model identification (i.e., hyperparameter selection) in statistical modeling, thereby reinforcing the design rationale behind the proposed stacking ensemble methodology.

- In the foreseeable scenario and following the lead of more mature technologies, predicting the energy production from other RESs, in particular wave conditions/wave power, is critical to maintaining a stable electrical grid due to the intermittent and non-dispatchable nature of these generation technologies — especially as the amount of energy generated from these sources increases. Additionally, these predictions are also key for project planning, decision-making and risk management of market players considering investments in these technologies. For all these reasons, tailored analysis, input selection and forecasters should be investigated.

- By integrating energy resources into an optimal configuration in HES studies — a topic of growing interest with the emergence of micro-grids and energy communities — the inherent variability of solar and wind resources can be partially mitigated with storage technologies. This integration enhances overall system reliability and cost-effectiveness. In this context, and as specifically demonstrated in sub-section 7.3.2, short-term price forecasting can serve as a distinctive informational variable, in the cost formation equations, which then better supports the dispatch task.

- Finally, the clarity gained from our findings casts a definitive light on a critical conclusion: this transition must be pursued without compromising grid robustness, security of supply, or price affordability. Failing to do so risks not only losing public support but also undermining the core objective itself—as some isolated yet cautionary precedents have already demonstrated.

8.3 Future Works

Given the complex challenge of large-scale integration of variable renewable energies, it is easy to envision a range of follow-up studies. Just as the surge in electricity prices and subsequent regulatory changes served as a pretext for broader debates on energy policy, future research could similarly use regional imbalances in energy imports and exports as a starting point. Such studies could explore how the flow of electricity from low-price to high-price regions influences perceptions of energy dependence, and how end-users (final consumers) may not perceive the resulting savings — shedding light on the discrepancies between public energy-policy discourse and the underlying realities.
Still, in the Iberian context, a valuable line of inquiry would be to simulate scenarios of extremely high renewable energy penetration. This could include projections of capacity factors, transmission infrastructure requirements, and the potential impact of curtailments — especially if the peninsula's "energy island" status remains unchanged. Given that Spain and Portugal share a common electricity market, this research could offer critical insights

into the region's energy resilience and the barriers to further integration. From this more technical insights, one could derive and connect with additional OPF considerations and scenarios for multi-objective optimization problems.

A comparative study could also be conducted using the extensive review of flexibility options. For instance, applying these options to larger national loads, highly interconnected but with less favorable renewable conditions — such as Germany — could reveal how different flexibility strategies lead to distinct energy policy recommendations. This would be particularly relevant for more industrialized and electrified economies.

At a more granular level, there are also promising research avenues in forecasting and data modeling. One notable differentiator in some of the time-series studies reviewed in this work was the approach to input data selection and its significant impact on forecasting accuracy. This aspect could be further explored not only across different case studies, but also by testing a broader range of forecasting tools/models with varying hyperparameters in an online setting. Such work could compare emerging differences and investigate deeper connections with the generation mix, load behavior, regulatory frameworks, and renewable pricing incentive mechanisms, among other factors. For instance, it could conclude how ”artificial” the price formation mechanisms may be.

Bibliography

[1] M. R. Abram and M. Sforna, "Exploring complexity in electric power systems," in *Methods, Models, Simulations And Approaches Towards A General Theory Of Change*. WORLD SCIENTIFIC, mar 2012, pp. 229–244. [Online]. Available: http://www.worldscientific.com/doi/abs/10.1142/9789814383332_0016 1

[2] K. F. Zhang and X. Z. Dai, "Structural analysis of large-scale power systems," *Mathematical Problems in Engineering*, vol. 2012, 2012. 1

[3] A. Poullikkas, S. Papadouris, G. Kourtis, I. Hadjipaschalis, A. Poullikkas, S. Papadouris, G. Kourtis, and I. Hadjipaschalis, "Storage Solutions for Power Quality Problems in Cyprus Electricity Distribution Network," *AIMS Energy*, vol. 2, no. 1, pp. 1–17, 2014. [Online]. Available: http://www.aimspress.com/article/doi/10.3934/energy.2014.1.1http://www.aimspress.com/rticle/doi/10.3934/energy.2014.1.1 1

[4] C. Breyer, S. Khalili, D. Bogdanov, M. Ram, A. S. Oyewo, A. Aghahosseini, A. Gulagi, A. A. Solomon, D. Keiner, G. Lopez, P. A. Østergaard, H. Lund, B. V. Mathiesen, M. Z. Jacobson, M. Victoria, S. Teske, T. Pregger, V. Fthenakis, M. Raugei, H. Holttinen, U. Bardi, A. Hoekstra, and B. K. Sovacool, "On the history and future of 100% renewable energy systems research," *IEEE Access*, vol. 10, pp. 78 176–78 218, 2022. 2, 23, 68

[5] I. Z. F. bin Hussien, A. A. Rahim, and N. Abdullah, "Chapter 7 - Electric Power Transmission," in *Alternative Energy in Power Electronics*, M. H. Rashid, Ed. Boston: Butterworth-Heinemann, 2011, pp. 317–347. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B978012416714800007X 3

[6] M. Aybar-Mejía, J. Villanueva, D. Mariano-Hernández, F. Santos, A. Molina-García, M. . Villanueva, J. . Mariano-Hernández, D. . Santos, and F. . Molina-García, "A Review of Low-Voltage Renewable Microgrids: Generation Forecasting and Demand-Side Management Strategies," *Electronics*, vol. 10, no. 17, p. 2093, aug 2021. [Online]. Available: https://www.mdpi.com/2079-9292/10/17/2093/htm 3

[7] "General," in *UHV Transmission Technology*, The China Electric Power Research Institute, Ed. Academic Press, 2018, ch. 1, pp. 3–20. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/B978012805193100001X 3

[8] K. Prakash, A. Lallu, F. R. Islam, and K. A. Mamun, "Review of Power System Distribution Network Architecture," *Proceedings - Asia-Pacific World Congress on Computer Science and Engineering 2016 and Asia-Pacific World Congress on Engineering 2016, APWC on CSE/APWCE 2016*, pp. 124–130, jun 2017. 4

[9] L. Mehigan, J. P. Deane, B. P. Gallachóir, and V. Bertsch, “A review of the role of distributed generation (DG) in future electricity systems,” *Energy*, vol. 163, pp. 822–836, nov 2018. 4

[10] NERC, “An introduction to inverter-based resources on the bulk power system,” techreport, Jun. 2023. [Online]. Available: https://www.nerc.com/pa/Documents/2023_NERC_Guide_Inverter-Based-Resources.pdf 4

[11] E. Ridha, L. Nolting, and A. Praktiknjo, “Complexity profiles: A large-scale review of energy system models in terms of complexity,” *Energy Strategy Reviews*, vol. 30, p. 100515, jul 2020. 4

[12] S. Abeysinghe, M. Abeysekera, J. Wu, and M. Sooriyabandara, “Electrical properties of medium voltage electricity distribution networks,” *CSEE Journal of Power and Energy Systems*, vol. 7, no. 3, pp. 497–509, may 2021. 5

[13] S. P. Burger, J. D. Jenkins, S. C. Huntington, and I. J. Perez-Arriaga, “Why distributed?: A critical review of the tradeoffs between centralized and decentralized resources,” *IEEE Power and Energy Magazine*, vol. 17, no. 2, pp. 16–24, mar 2019. 5

[14] F. Martín-Martínez, A. Sánchez-Miralles, M. Rivier, and C. F. Calvillo, “Centralized vs distributed generation. A model to assess the relevance of some thermal and electric factors. Application to the Spanish case study,” *Energy*, vol. 134, pp. 850–863, sep 2017. 5

[15] K. Sadovskaia, D. Bogdanov, S. Honkapuro, and C. Breyer, “Power transmission and distribution losses – A model based on available empirical data and future trends for all countries globally,” *International Journal of Electrical Power & Energy Systems*, vol. 107, pp. 98–109, may 2019. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0142061518335075 5

[16] D. Marujo, G. L. Zanatta, and H. A. Floréz, “Optimal management of electrical power systems for losses reduction in the presence of active distribution networks,” *Electrical Engineering*, vol. 103, no. 3, pp. 1725–1736, jun 2021. [Online]. Available: https://link.springer.com/article/10.1007/s00202-020-01182-5 5

[17] F. d. S. Savian, J. C. M. Siluk, T. B. Garlet, F. M. do Nascimento, J. R. Pinheiro, and Z. Vale, “Non-technical losses: A systematic contemporary article review,” *Renewable and Sustainable Energy Reviews*, vol. 147, p. 111205, sep 2021. 6

[18] E. Benedict, T. Collins, D. Gotham, S. Hoffman, D. Karipides, S. Pekarek, and R. Ramabhadran, “LOSSES IN ELECTRIC POWER SYSTEMS,” West Lafayette, Indiana, Tech. Rep., 1992. [Online]. Available: https://docs.lib.purdue.edu/ecetr/266/ 6

[19] Council of European Energy Regulators, “2nd CEER Report on Power Losses,” Brussels, Belgium, Tech. Rep., mar 2020. [Online]. Available: https://www.ceer.eu/documents/104400/-/-/fd4178b4-ed00-6d06-5f4b-8b87d630b060 6

[20] K. P. Singh, A. Sharan, A. Agrawal, and R. S. Ambekar, “TRANSMISSION LINE LOSSES WITH REFERENCE TO TEMPERATURE,” *EPRA International Journal of Research & Development (IJRD)*, vol. 6, no. 7, pp. 1–1, jul 2021. [Online]. Available: https://eprajournals.com/IJSR/article/5532 7

[21] Y. Venu, T. Nireekshana, and B. Phanisaikrishna, “Mitigation of Ferranti Effect Using Thyristor Controlled Reactor,” *Lecture Notes in Electrical Engineering*, vol. 700, pp. 2533–2545, 2021. [Online]. Available: https://link.springer.com/chapter/10.1007/978-981-15-8221-9_236 7

[22] J. Blanes, D. Borge-Diez, A. Gonzalez-Martinez, and M. De Simon-Martin, “Scale model of a Very High Voltage alternating current Power Transmission Line integrated into a Smart Grid,” *IYCE 2015 - Proceedings: 2015 5th International Youth Conference on Energy*, aug 2015. 7

[23] A. S. Bretas, N. G. Bretas, J. B. London, and B. E. Carvalho, “Power flow in electrical systems,” *Cyber-Physical Power Systems State Estimation*, pp. 31–60, jan 2021. 7

[24] E. H. Allen, R. B. Stuart, and T. E. Wiedman, “No light in August: Power system restoration following the 2003 North American blackout,” *IEEE Power and Energy Magazine*, vol. 12, no. 1, pp. 24–33, jan 2014. 8

[25] P. Fernández-Bustamante, O. Barambones, I. Calvo, C. Napole, and M. Derbeli, “Provision of Frequency Response from Wind Farms: A Review,” *Energies*, vol. 14, no. 20, p. 6689, oct 2021. [Online]. Available: https://www.mdpi.com/1996-1073/14/20/6689/htm 8, 10

[26] ENTSO-E, “Inertia and Rate of Change of Frequency (RoCoF),” Brussels, Tech. Rep., 2020. [Online]. Available: https://eepublicdownloads.entsoe.eu/clean-documents/SOCdocuments/InertiaandRoCoF_v17_clean.pdf 8

[27] B. J. Kirby, J. Dyer, C. Martinez, R. A. Shoureshi, R. Guttromson, J. Dagle, and Others, *Frequency control concerns in the North American electric power system*. United States. Department of Energy, 2003. 9

[28] B. Potel, V. Debusschere, F. Cadoux, and L. de Alvaro Garcia, “Under-frequency Load Shedding Schemes Characteristics and Performance Criteria,” in *PowerTech 2017*, Manchester, United Kingdom, jun 2017. [Online]. Available: https://hal.science/hal-01549263 9, 11

[29] L. Badesa, F. Teng, and G. Strbac, “Pricing inertia and Frequency Response with diverse dynamics in a Mixed-Integer Second-Order Cone Programming formulation,” *Applied Energy*, vol. 260, p. 114334, feb 2020. 10

[30] Z. A. Obaid, L. M. Cipcigan, L. Abrahim, and M. T. Muhssin, "Frequency control of future power systems: reviewing and evaluating challenges and new control methods," *Journal of Modern Power Systems and Clean Energy*, vol. 7, no. 1, pp. 9–25, jan 2019. [Online]. Available: https://link.springer.com/article/10.1007/s40565-018-0441-1 10

[31] Y. G. Rebours, D. S. Kirschen, M. Trotignon, and S. Rossignol, "A survey of frequency and voltage control ancillary services - Part I: Technical features," *IEEE Transactions on Power Systems*, vol. 22, no. 1, pp. 350–357, feb 2007. 10

[32] P. Tumino, "Frequency control in a power system," oct 2020. [Online]. Available: https://eepower.com/technical-articles/frequency-control-in-a-power-system/ 10

[33] C. D. Parker, "Applications - Stationary | Energy Storage Systems: Batteries," *Encyclopedia of Electrochemical Power Sources*, pp. 53–64, 2009. 10

[34] Y. Rebours and D. Kirschen, "What is spinning reserve," *The University of Manchester*, vol. 174, no. 175, pp. 1–11, sep 2005. [Online]. Available: https://labs.ece.uw.edu/real/Library/Reports/What_is_spinning_reserve.pdf 10

[35] T. Sousa, J. A. Jardini, M. Masuda, and R. A. De Lima, "Spinning reserve opportunity cost in hydroelectric plants," *2005 IEEE Power Engineering Society General Meeting*, vol. 2, pp. 2064–2069, 2005. 10

[36] E. Potter, "Examining The California And Texas Power Outages," MSc Thesis, Johns Hopkins University, Baltimore, Maryland, dec 2021. 11

[37] J. Bialek, "What does the GB power outage on 9 August 2019 tell us about the current state of decarbonised power systems?" *Energy Policy*, vol. 146, p. 111821, nov 2020. 11

[38] National Grid ESO, "Past Frequency Events - A System Operability Framework," Tech. Rep., nov 2019. [Online]. Available: https://www.nationalgrideso.com/document/156761/download 11

[39] B. D. Raj, S. Kumar, S. Padhi, A. Sarkar, A. Mondal, and K. Ramamritham, "Brownout Based Blackout Avoidance Strategies in Smart Grids," *IEEE Transactions on Sustainable Computing*, vol. 6, no. 4, pp. 586–598, 2021. 11

[40] X. Luo, J. Wang, J. D. Wojcik, J. Wang, D. Li, M. Draganescu, Y. Li, and S. Miao, "Review of Voltage and Frequency Grid Code Specifications for Electrical Energy Storage Applications," *Energies*, vol. 11, no. 5, p. 1070, apr 2018. [Online]. Available: https://www.mdpi.com/1996-1073/11/5/1070/htm 12

[41] M. H. Othman, H. Mokhlis, M. Mubin, S. Talpur, N. F. A. Aziz, M. Dradi, and H. Mohamad, "Progress in control and coordination of energy storage

system-based VSG: a review," *IET Renewable Power Generation*, vol. 14, no. 2, pp. 177–187, feb 2020. [Online]. Available: https://onlinelibrary.wiley.com/doi/full/10.1049/iet-rpg.2019.0274 12

[42] S. Perras, "Electricity transmission line planning: Success factors for transmission system operators to reduce public opposition," Ph.D. dissertation, TU Dresden, 2015. 12

[43] L. M. Abadie and J. M. Chamorro, "Evaluation of a cross-border electricity interconnection: The case of Spain-France," *Energy*, vol. 233, p. 121177, oct 2021. 12

[44] D. Keles, J. Dehler-Holland, M. Densing, E. Panos, and F. Hack, "Cross-border effects in interconnected electricity markets - an analysis of the Swiss electricity prices," *Energy Economics*, vol. 90, p. 104802, aug 2020. 12

[45] G. Dimopoulos, C. Heussaff, and G. Zachmann, "The massive value of European Union cross-border electricity transmission," Brussels, 2023. [Online]. Available: https://www.bruegel.org/analysis/massive-value-european-union-cross-border-electricity-transmission 12

[46] E. M. Carlini, R. Schroeder, J. M. Birkebæk, and F. Massaro, "EU transition in power sector: How RES affects the design and operations of transmission power systems," *Electric Power Systems Research*, vol. 169, pp. 74–91, apr 2019. 13

[47] J. Lin and F. H. Magnago, *Electricity markets: Theories and applications*. John Wiley & Sons, 2017. 13, 20, 21

[48] A. G. Maksimov and D. V. S. and, "Forecasting of the electricity price on the day-ahead electricity market in russia," *Cogent Physics*, vol. 4, no. 1, p. 1316953, 2017. [Online]. Available: https://doi.org/10.1080/23311940.2017.1316953 13

[49] L. Meeus, *The Evolution of Electricity Markets in Europe*. Cheltenham: Edward Elgar Publishing, 11 2020. 15

[50] M. Torres, "Power market auctions 101: Explaining the eu electricity market (n-side)," Blog post, May 2023. [Online]. Available: https://energy.n-side.com/blog/power-market-auctions-101-explaining-the-eu-electricity-market 15, 17

[51] T. Mäkelä, "Optimising a renewable virtual power plant in day-ahead, intraday and balancing markets," mathesis, Aalto University, Dec. 2024. [Online]. Available: https://urn.fi/URN:NBN:fi:aalto-202501282109 16, 17, 18

[52] D. Gan, D. Feng, J. Xie, and D. Feng, *Electricity markets and power system economics*. CRC Press Boca Raton, FL, 2014. 16

[53] A. Bublitz, P. Ringler, M. Genoese, D. Keles, and W. Fichtner, "Agent-based simulation of interrelated wholesale electricity markets," in *2014 25th International Workshop on Database and Expert Systems Applications*, 2014, pp. 103–108. 17

[54] C. Graf, T. Kuppelwieser, and D. Wozabal, “Frequent auctions for intraday electricity markets,” *The Energy Journal*, vol. 45, no. 1, pp. 231–256, 2024. [Online]. Available: https://doi.org/10.5547/01956574.45.1.cgra 18

[55] F. Bevin-McCrimmon, I. Diaz-Rainey, M. McCarten, and G. Sise, “Liquidity and risk premia in electricity futures,” *Energy Economics*, vol. 75, pp. 503–517, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0140988318303700 20

[56] M. Beus, I. Pavić, I. Štritof, T. Capuder, and H. Pandžić, “Electricity market design in croatia within the european electricity market—recommendations for further development,” *Energies*, vol. 11, no. 2, 2018. [Online]. Available: https://www.mdpi.com/1996-1073/11/2/346 20

[57] A. GmbH, “Overview electricity markets in europe,” Market Report, 2023. [Online]. Available: https://www.amprion.net/Market/Market-Report/Overview-Electricity-Markets-in-Europe/ 20

[58] Á. Manso-Burgos, D. Ribó-Pérez, S. Mateo-Barcos, P. Carnero, and T. Gómez-Navarro, “Market value and agents benefits of enhanced short-term solar pv power generation forecasting,” *Machines*, vol. 10, no. 9, 2022. [Online]. Available: https://www.mdpi.com/2075-1702/10/9/730 20

[59] B. G. Magalhães, P. M. Bento, J. A. Pombo, M. R. Calado, and S. J. Mariano, “Spot price forecasting for best trading strategy decision support in the iberian electricity market,” *Expert Systems with Applications*, vol. 224, p. 120059, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417423005614 21

[60] N. Amjady and F. Keynia, “A new prediction strategy for price spike forecasting of day-ahead electricity markets,” *Applied Soft Computing*, vol. 11, no. 6, pp. 4246–4256, 2011. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1568494611001207 21

[61] D. Reiche and M. Bechberger, “Policy differences in the promotion of renewable energies in the eu member states,” *Energy Policy*, vol. 32, no. 7, pp. 843–849, 2004. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0301421502003439 22

[62] Y. Krozer, “Financing of the global shift to renewable energy and energy efficiency,” *Green Finance*, vol. 1, no. 3, pp. 264–278, 2019. [Online]. Available: https://www.aimspress.com/article/doi/10.3934/GF.2019.3.264 22

[63] A. Eitan and M. P. Hekkert, “Locked in transition? towards a conceptualization of path-dependence lock-ins in the renewable energy landscape,” *Energy*

Research & Social Science, vol. 106, p. 103316, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2214629623003766 23, 24

[64] M. K. Seibert and W. E. Rees, "Through the eye of a needle: An eco-heterodox perspective on the renewable energy transition," *Energies*, vol. 14, no. 15, 2021. [Online]. Available: https://www.mdpi.com/1996-1073/14/15/4508 23

[65] I. for Energy Research, "Renewable energy received record subsidies in 2024 (ier)," Commentary, Jan. 2025. [Online]. Available: https://www.instituteforenergyresearch.org/renewable/renewable-energy-received-record-subsidies-in-2024/ 23

[66] B. A. Carreras, P. Colet, J. M. Reynolds-Barredo, and D. Gomila, "Assessing blackout risk with high penetration of variable renewable energies," *IEEE Access*, vol. 9, pp. 132 663–132 674, 2021. 23

[67] D. Finon, "Intermittent renewables and the threat of blackouts, how to deal with it? (encyclopedie-energie)," Oct. 2022. [Online]. Available: https://www.encyclopedie-energie.org/en/intermittent-renewables-threat-of-blackouts-how-to-deal-with-it/ 23

[68] R. H. Schulte and F. C. Fletcher, "100conundrum," *The Electricity Journal*, vol. 32, no. 2, pp. 31–36, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1040619019300120 23

[69] K. De Vos, "Negative wholesale electricity prices in the german, french and belgian day-ahead, intra-day and real-time markets," *The Electricity Journal*, vol. 28, no. 4, pp. 36–50, 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1040619015000652 23

[70] P. Rövekamp, M. Schöpf, F. Wagon, M. Weibelzahl, and G. Fridgen, "Renewable electricity business models in a post feed-in tariff era," *Energy*, vol. 216, p. 119228, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0360544220323355 23

[71] H. B. Asmelash, "Energy subsidies and wto dispute settlement: Why only renewable energy subsidies are challenged," *Journal of International Economic Law*, vol. 18, no. 2, pp. 261–285, 05 2015. [Online]. Available: https://doi.org/10.1093/jiel/jgv024 23

[72] R. Rajan, F. M. Fernandez, and Y. Yang, "Primary frequency control techniques for large-scale pv-integrated power systems: A review," *Renewable and Sustainable Energy Reviews*, vol. 144, p. 110998, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1364032121002884 25

[73] Y. Xu, V. Ramanathan, and D. G. Victor, “Global warming will happen faster than we think,” *Nature*, vol. 564, no. 7734, pp. 30–32, dec 2018. [Online]. Available: https://www.nature.com/articles/d41586-018-07586-5 31

[74] D. Gielen, F. Boshell, D. Saygin, M. D. Bazilian, N. Wagner, and R. Gorini, “The role of renewable energy in the global energy transformation,” *Energy Strategy Reviews*, vol. 24, pp. 38–50, apr 2019. 31

[75] D. 2009/28/EC, “Directive 2009/28/ec of the european parliament and of the council of 23 april 2009 on the promotion of the use of energy from renewable sources,” 2009. [Online]. Available: https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2009:140:0016:0062:en:PDF 32

[76] E. Commission and D.-G. for Energy, *Energy – Roadmap 2050*. Publications Office, 2012. 32

[77] H. Fekete, T. Kuramochi, M. Roelfsema, M. den Elzen, N. Forsell, N. Höhne, L. Luna, F. Hans, S. Sterl, J. Olivier, H. van Soest, S. Frank, and M. Gusti, “A review of successful climate change mitigation policies in major emitting economies and the potential of global replication,” *Renewable and Sustainable Energy Reviews*, vol. 137, p. 110602, mar 2021. 32, 33

[78] J. Finkelstein, D. Frankel, and J. Noffsinger, “How to decarbonize global power systems,” may 2020. [Online]. Available: https://www.mckinsey.com/industries/electric-power-and-natural-gas/our-insights/how-to-decarbonize-global-power-systems 32

[79] P. Bento, H. Nunes, J. Pombo, M. d. R. Calado, and S. Mariano, “Daily Operation Optimization of a Hybrid Energy System Considering a Short-Term Electricity Price Forecast Scheme,” *Energies*, vol. 12, no. 5, p. 924, mar 2019. [Online]. Available: https://www.mdpi.com/1996-1073/12/5/924 32

[80] Y. Yao, J. H. Xu, and D. Q. Sun, “Untangling global levelised cost of electricity based on multi-factor learning curve for renewable energy: Wind, solar, geothermal, hydropower and bioenergy,” *Journal of Cleaner Production*, vol. 285, p. 124827, feb 2021. 32, 39

[81] E. Papadis and G. Tsatsaronis, “Challenges in the decarbonization of the energy sector,” *Energy*, vol. 205, p. 118025, aug 2020. 32

[82] A. Aghahosseini, D. Bogdanov, L. S. Barbosa, and C. Breyer, “Analysing the feasibility of powering the Americas with renewable energy and inter-regional grid interconnections by 2030,” *Renewable and Sustainable Energy Reviews*, vol. 105, pp. 187–205, may 2019. 32

[83] International Energy Agency, *The Power of Transformation: Wind, Sun and the Economics of Flexible Power Systems*. Paris: IEA, 2014. [Online]. Available: https://www.oecd-ilibrary.org/content/publication/9789264208032-en 32

[84] J. van Zalk and P. Behrens, "The spatial extent of renewable and non-renewable power generation: A review and meta-analysis of power densities and their application in the U.S." *Energy Policy*, vol. 123, pp. 83–91, dec 2018. 32

[85] Eurostat, "International trade in products related to green energy," oct 2021. [Online]. Available: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=International_trade_in_products_related_to_green_energy&oldid=545908 32

[86] P. Behuria, "The politics of late late development in renewable energy sectors: Dependency and contradictory tensions in India's National Solar Mission," *World Development*, vol. 126, p. 104726, feb 2020. 32

[87] M. Doepfert and R. Castro, "Techno-economic optimization of a 100% renewable energy system in 2050 for countries with high shares of hydropower: The case of Portugal," *Renewable Energy*, vol. 165, pp. 491–503, mar 2021. 32

[88] L. Fernandes and P. Ferreira, "Renewable energy scenarios in the Portuguese electricity system," *Energy*, vol. 69, pp. 51–57, may 2014. 32

[89] R. Gómez-Calvet, J. M. Martínez-Duart, and S. Serrano-Calle, "Current state and optimal development of the renewable electricity generation mix in Spain," *Renewable Energy*, vol. 135, pp. 1108–1120, may 2019. 32

[90] D. García-Gusano and D. Iribarren, "Prospective energy security scenarios in Spain: The future role of renewable power generation technologies and climate change implications," *Renewable Energy*, vol. 126, pp. 202–209, oct 2018. 33

[91] Red Eléctrica de España, "REData - Estructura generación emisiones asociadas | Red Eléctrica de España," 2022. [Online]. Available: https://www.ree.es/es/datos/generacion/estructura-generacion 33, 52, 66

[92] Direção-Geral de Energia e Geologia, "Produção anual e potência instalada," 2021. [Online]. Available: https://www.dgeg.gov.pt/pt/estatistica/energia/eletricidade/producao-anual-e-potencia-instalada/ 34

[93] C. V. Miguel, A. Mendes, and L. M. Madeira, "An overview of the portuguese energy sector and perspectives for power-to-gas implementation," *Energies*, vol. 11, p. 3259, 11 2018. [Online]. Available: https://www.mdpi.com/1996-1073/11/12/3259/htm 34

[94] Red Eléctrica de España, "El sistema eléctrico español 2020," Red Eléctrica de España, Madrid, Tech. Rep., 2021. [Online].

Available: https://www.ree.es/es/datos/publicaciones/informe-anual-sistema/informe-del-sistema-electrico-espanol-2020 37

[95] J. G. Navarro, "Solo cuatro centrales de carbón se mantendrán en funcionamiento en España," 2020. [Online]. Available: https://www.abc.es/economia/abci-solo-cuatro-centrales-carbon-mantendran-funcionamiento-espana-202007010202_noticia.html 37

[96] M. Planelles, "España desconecta siete centrales térmicas y arranca el proceso para enterrar el carbón | Sociedad | EL PAÍS," 2020. [Online]. Available: https://elpais.com/sociedad/2020-06-28/espana-desconecta-siete-termicas-y-arranca-el-proceso-para-enterrar-el-carbon.html 37

[97] María Castañeda, "EDP invierte 15 M€ en Aboño: otro paso hacia la descarbonización," nov 2020. [Online]. Available: https://www.merca2.es/2020/11/16/edp-inversion-abono-descarbonizacion/ 37

[98] Endesa, "Dismantling of the Litoral thermal power plant | Endesa," 2022. [Online]. Available: https://www.endesa.com/en/press/press-room/news/energy-transition/endesa-assigned-lezama-and-caldererias-litoral-thermal-power-plant-dismantling 38

[99] Global Energy Monitor, "As Pontes power station - Global Energy Monitor," 2024. [Online]. Available: https://www.gem.wiki/As_Pontes_power_station 38

[100] Global Energy Observatory, "Los Barrios Coal Power Plant Spain - GEO," 2012. [Online]. Available: http://globalenergyobservatory.org/geoid/43679 38

[101] EDP España, "Aboño | EDP España," 2022. [Online]. Available: https://espana.edp.com/en/abono 38

[102] ——, "Thermic Soto de Ribera | EDP España," 2022. [Online]. Available: https://espana.edp.com/en/thermic-soto-de-ribera 38

[103] Mike Fulwood, "Surging 2021 European Gas Prices – Why and How?" Oxford Institute for Energy Studies, Oxford, Tech. Rep., jan 2022. [Online]. Available: https://www.oxfordenergy.org/publications/surging-2021-european-gas-prices-why-and-how/ 38

[104] Market Observatory for Energy, "Quarterly report on European electricity markets: Volume 14 (Issue 4)," European Commission, Brussels, Tech. Rep., 2022. [Online]. Available: https://energy.ec.europa.eu/data-and-analysis/market-analysis_en 38

[105] P. K. Ozili, "Global Economic Consequence of Russian Invasion of Ukraine," *SSRN Electronic Journal*, p. 27, 2022. [Online]. Available: https://papers.ssrn.com/abstract=4064770 38

[106] Market Observatory for Energy, "Quarterly report On European electricity markets With focus on developments in annual wholesale prices," European Comission, Brussels, Tech. Rep., 2021. [Online]. Available: https://energy.ec.europa.eu/data-and-analysis/market-analysis_en 38

[107] Enrique Perez, "Coal soars in Europe in the midst of the energy crisis: from closing plants to a last-minute solution to face the winter," oct 2021. [Online]. Available: https://www.xataka.com/energia/carbon-se-dispara-europa-plena-crisis-energetica-cerrar-centrales-a-solucion-ultima-hora-para-afrontar-invierno 39

[108] Josu Alonso, "Spain rules out extending the life of thermal power plants and focuses on renewables | | News SER Chain," may 2022. [Online]. Available: https://cadenaser.com/asturias/2022/05/24/espana-descarta-alargar-la-vida-de-las-centrales-termicas-y-se-centra-en-las-renovables-ser-gijon/ 39

[109] Energía y Sociedad, "1.3. El valor de la flexibilidad de los ciclos combinados de gas natural - Energía y Sociedad," 2022. [Online]. Available: https://www.energiaysociedad.es/manual-de-la-energia/1-3-el-valor-de-la-flexibilidad-de-los-ciclos-combinados-de-gas-natural/ 39

[110] ENGIE, "Gas natural - ENGIE España," 2021. [Online]. Available: https://www.engie.es/actividades/energia-termica/gas-natural/ 39

[111] L. Fernández, "• Capacity of nuclear power plants Spain | Statista," 2021. [Online]. Available: https://www.statista.com/statistics/1229676/power-of-nuclear-power-plants-in-spain-by-plant/ 39

[112] Foro de la Industria Nuclear Española, "Nuclear power in Spain - Foro Nuclear," 2021. [Online]. Available: https://www.foronuclear.org/en/nuclear-power/nuclear-power-in-spain/ 39

[113] Y. T. Shah, *Thermal energy : sources, recovery, and applications*, 1st ed. CRC Press, 2018. [Online]. Available: https://books.google.com/books/about/Thermal_Energy.html?id=fEQPEAAAQBAJ 40

[114] P. Bento, S. Mariano, M. Calado, and J. Pombo, "Impacts of the COVID-19 pandemic on electric energy load and pricing in the Iberian electricity market," *Energy Reports*, vol. 7, pp. 4833–4849, nov 2021. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S2352484721004236 40, 76, 92

[115] S. Kalinowski, A. Łuczak, O. Bednář, A. A. Andreačečrdlová, B. Kadeřábková, and P. P. Pavelřežábek, "Energy prices impact on inflationary spiral," *Energies*, vol. 15, p. 3443, 5 2022. [Online]. Available: https://www.mdpi.com/1996-1073/15/9/3443/htm 41

[116] International Energy Agency, "Portugal 2021-Energy Policy Review," IEA, Paris, Tech. Rep., jul 2021. [Online]. Available: https://www.iea.org/reports/portugal-2021 41

[117] A. N. Nunes, "Energy changes in Portugal," *Mediterranee*, no. 130, pp. 1–20, sep 2018. [Online]. Available: http://journals.openedition.org/mediterranee/10113 41

[118] S. R. Sinsel, R. L. Riemke, and V. H. Hoffmann, "Challenges and solution technologies for the integration of variable renewable energy sources—a review," *Renewable Energy*, vol. 145, pp. 2271–2285, jan 2020. 42

[119] G. Ren, J. Liu, J. Wan, Y. Guo, and D. Yu, "Overview of wind power intermittency: Impacts, measurements, and mitigation solutions," *Applied Energy*, vol. 204, pp. 47–65, oct 2017. 42, 45, 47, 48

[120] D. Rowe, S. Sayeef, and G. Platt, "Intermittency: It's the Short-Term That Matters," *Future of Utilities - Utilities of the Future: How Technological Innovations in Distributed Energy Resources Will Reshape the Electric Power Sector*, pp. 129–150, jan 2016. 42

[121] L. M. Abadie, N. Goicoechea, F. González, and M. J. Presno, "Old Wind Farm Life Extension vs. Full Repowering: A Review of Economic Issues and a Stochastic Application for Spain," *Energies*, vol. 14, no. 12, p. 3678, jun 2021. [Online]. Available: https://www.mdpi.com/1996-1073/14/12/3678/htm 42

[122] A. Colmenar-Santos, S. Campíñez-Romero, C. Pérez-Molina, and F. Mur-Pérez, "Repowering: An actual possibility for wind energy in Spain in a new scenario without feed-in-tariffs," *Renewable and Sustainable Energy Reviews*, vol. 41, pp. 319–337, jan 2015. 42

[123] P. Bento, S. Mariano, P. Carvalho, M. do Rosário Calado, and J. Pombo, "Soaring electricity prices in the day-ahead iberian market: policy insights, regulatory challenges and lack of system flexibility," *International Journal of Energy Sector Management*, vol. ahead-of-p, 1 2023. [Online]. Available: https://doi.org/10.1108/IJESM-07-2022-0007 43

[124] F. Castro-Rodríguez and D. Miles-Touya, "Impact of spanish renewable support scheme reforms on the revenues of photovoltaic power plants," *Utilities Policy*, vol. 80, p. 101476, 2 2023. 43

[125] F. Amorim, J. Vasconcelos, I. C. Abreu, P. P. Silva, and V. Martins, "How much room for a competitive electricity generation market in Portugal?" *Renewable and Sustainable Energy Reviews*, vol. 18, pp. 103–118, feb 2013. 43

[126] Global Offshore Wind Alliance and International Renewable Energy Agency, "30 years of policies for wind energy - lessons from 12 wind energy markets," 2013.

[Online]. Available: https://policycommons.net/artifacts/1855903/30-years-of-policies-for-wind-energy/ 43

[127] J. J. Andreas, C. Burns, and J. Touza, "Portugal under austerity: from financial to renewable crisis?" *Environmental Research Communications*, vol. 1, p. 091005, 9 2019. [Online]. Available: https://iopscience.iop.org/article/10.1088/2515-7620/ab3cb0 43

[128] U. Energy Information Administration, "Levelized Costs of New Generation Resources in the Annual Energy Outlook 2022," Energy Information Administration, U.S., Washington, D.C., Tech. Rep., mar 2022. [Online]. Available: https://www.eia.gov/outlooks/aeo/pdf/electricity_generation.pdf 43, 44

[129] J. Aldersey-Williams and T. Rubert, "Levelised cost of energy – a theoretical justification and critical assessment," *Energy Policy*, vol. 124, pp. 169–179, 1 2019. 44

[130] G. R. Timilsina, "Demystifying the costs of electricity generation technologies," *Policy Research Working Paper Series 9303*, vol. The World Bank, 2020. 44

[131] C. Kost, S. Shammugam, V. Fluri, D. Peper, A. D. Memar, and T. Schlegel, "Levelized cost of electricity-renewable energy technologies," pp. 1–45, 2021. [Online]. Available: https://www.ise.fraunhofer.de/en/publications/studies/cost-of-electricity.html 44

[132] H. Zhao, Q. Wu, S. Hu, H. Xu, and C. N. Rasmussen, "Review of energy storage system for wind power integration support," *Applied Energy*, vol. 137, pp. 545–553, jan 2015. 45

[133] A. Castillo and D. F. Gayme, "Grid-scale energy storage applications in renewable energy integration: A survey," *Energy Conversion and Management*, vol. 87, pp. 885–894, nov 2014. 45

[134] A. A. Kebede, T. Kalogiannis, J. Van Mierlo, and M. Berecibar, "A comprehensive review of stationary energy storage devices for large scale renewable energy sources grid integration," *Renewable and Sustainable Energy Reviews*, vol. 159, p. 112213, may 2022. 45

[135] H. Zsiborács, N. H. Baranyai, A. Vincze, L. Zentkó, Z. Birkner, K. Máté, and G. Pintér, "Intermittent Renewable Energy Sources: The Role of Energy Storage in the European Power System of 2040," *Electronics*, vol. 8, no. 7, p. 729, jun 2019. [Online]. Available: https://www.mdpi.com/2079-9292/8/7/729/htm 45

[136] M. M. Rahman, A. O. Oni, E. Gemechu, and A. Kumar, "Assessment of energy storage technologies: A review," *Energy Conversion and Management*, vol. 223, p. 113295, nov 2020. 45

[137] P. Bento, J. Pombo, M. Calado, and S. Mariano, "A bat optimized neural network and wavelet transform approach for short-term price forecasting," *Applied Energy*, vol. 210, pp. 88–97, jan 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306261917314782 45, 233

[138] F. M. Camilo, R. Castro, M. E. Almeida, and V. F. Pires, "Economic assessment of residential PV systems with self-consumption and storage in Portugal," *Solar Energy*, vol. 150, pp. 353–362, jul 2017. 45

[139] V. Pandey, A. Sircar, K. Yadav, and N. Bist, "Pumped hydro storage for intermittent renewable energy: Present status and future potential in india," *MRS Energy & Sustainability*, vol. 10, no. 2, pp. 189–206, 2023. [Online]. Available: https://doi.org/10.1557/s43581-023-00064-0 45

[140] Renews.biz, "Iberdrola opens 1.1GW Portuguese pumped hydro - reNews - Renewable Energy News," jul 2022. [Online]. Available: https://renews.biz/79230/iberdrola-opens-11gw-portuguese-pumped-hydro/ 46

[141] Energias de Portugal, "Construção do parque solar flutuante da EDP no Alqueva vai arrancar | edp.com," may 2021. [Online]. Available: https://www.edp.com/pt-pt/noticias/construcao-do-parque-solar-flutuante-da-edp-no-alqueva-vai-arrancar 47

[142] International Renewable Energy Agency, "Electricity storage and renewables: Costs and markets to 2030," IRENA, Masdar City, Tech. Rep., oct 2017. [Online]. Available: https://www.irena.org/publications/2017/oct/electricity-storage-and-renewables-costs-and-markets 47, 56

[143] M. Mérida-Rodríguez, S. Reyes-Corredera, S. Pardo-García, and B. Zayas-Fernández, "Solar Photovoltaic Power in Spain," in *Renewable Energies and European Landscapes*, M. Frolova, M.-J. Prados, and A. Nadaï, Eds. Dordrecht: Springer Netherlands, 2015, pp. 63–80. [Online]. Available: https://doi.org/10.1007/978-94-017-9843-3_4 47, 66

[144] R. Zhang, J. Fang, and Y. Tang, "Inertia emulation through supercapacitor energy storage systems," in *2019 10th International Conference on Power Electronics and ECCE Asia (ICPE 2019 - ECCE Asia)*, 2019, pp. 1365–1370. 47

[145] P. Breeze, "Hydrogen Energy Storage," in *Power System Energy Storage Technologies*, 1st ed. Academic Press, may 2018, ch. 8, pp. 69–77. 48

[146] B. Uzum, A. Onen, H. M. Hasanien, and S. M. Muyeen, "Rooftop Solar PV Penetration Impacts on Distribution Network and Further Growth Factors—A Comprehensive Review," *Electronics*, vol. 10, no. 1, pp. 1–31, 2021. 48

[147] M. Brown and R. Sedano, *Electricity transmission: a primer*. Washington, D.C: National Council on Electric Policy, 2004. [Online]. Available: https://books.google.com/books/about/Electricity_Transmission.html?hl=pt-PT&id=yQwKAQAAMAAJ 48

[148] C. Fernandes, P. Frías, and L. Olmos, "Expanding interconnection capacity to integrate intermittent generation in the Iberian Peninsula," *IET Renewable Power Generation*, vol. 7, no. 1, pp. 45–54, jan 2013. [Online]. Available: https://onlinelibrary.wiley.com/doi/full/10.1049/iet-rpg.2012.0151 48

[149] E. Goncalves, "Is france an obstacle to the iberian peninsula's goal of becoming an energy supplier?" 6 2022. [Online]. Available: https://www.euronews.com/my-europe/2022/06/07/is-france-an-obstacle-to-the-iberian-peninsula-s-goal-of-becoming-an-energy-supplier 49

[150] S. Debnath, P. R. Marthi, Q. Xia, J. Pan, M. Saeedifard, V. N. Vipin, S. Chakraborty, and M. Arifujjaman, "Renewable Integration in Hybrid AC/DC Systems Using a Multi-Port Autonomous Reconfigurable Solar Power Plant (MARS)," *IEEE Transactions on Power Systems*, vol. 36, no. 1, pp. 603–612, jan 2021. 49

[151] H. Shen, Q. Dai, Q. Wu, J. Wu, Q. Zhou, J. Wang, W. Yang, R. Pestana, and R. Pastor, "The state-of-the-arts of the study on grid interconnection between Iberian Peninsula and West Maghreb region," *Global Energy Interconnection*, vol. 1, no. 1, pp. 20–28, jan 2018. 49

[152] Red Eléctrica de España, "Strengthening interconnections | Red Eléctrica," 2021. [Online]. Available: https://www.ree.es/en/red21/strengthening-interconnections 49

[153] Carlos Filipe de Matos Cardoso, "Interligação Eléctrica entre Portugal e Espanha," Ph.D. dissertation, Instituto Superior Técnico, Universidade Técnica de Lisboa, Lisbon, 2011. [Online]. Available: https://fenix.tecnico.ulisboa.pt/cursos/meec/dissertacao/2353642375441 49

[154] J. Price and M. Zeyringer, "highRES-Europe: The high spatial and temporal Resolution Electricity System model for Europe," *SoftwareX*, vol. 17, p. 101003, jan 2022. 50

[155] E. Ogungbemi, T. W. Awotwe, F. Nisar, T. Wilberforce, A. Alanzi, O. Ijaodola, F. N. Khatib, Z. El -Hassan, and A. G. Olabi, "The Future of Renewable Energy, Barriers and Solution Fuel cell power module View project The Future of Renewable Energy, Barriers and Solution," in *Proceedings of SEEP2018*, Paisley, may 2018, pp. 215–222. [Online]. Available: https://www.researchgate.net/publication/329179383 50

[156] A. Kovač, M. Paranos, and D. Marciuš, "Hydrogen in energy transition: A review," *International Journal of Hydrogen Energy*, vol. 46, no. 16, pp. 10 016–10 035, mar 2021. 51

[157] International Energy Agency, "The Future of Hydrogen – Seizing today's opportunities," IEA, Paris, Tech. Rep., jun 2019. [Online]. Available: https://www.iea.org/reports/the-future-of-hydrogen 51

[158] D. Wang, M. Muratori, J. Eichman, M. Wei, S. Saxena, and C. Zhang, "Quantifying the flexibility of hydrogen production systems to support large-scale renewable energy integration," *Journal of Power Sources*, vol. 399, pp. 383–391, sep 2018. 51

[159] M. Yue, H. Lambert, E. Pahon, R. Roche, S. Jemei, and D. Hissel, "Hydrogen energy systems: A critical review of technologies, applications, trends and challenges," *Renewable and Sustainable Energy Reviews*, vol. 146, p. 111180, aug 2021. 51

[160] G. Kakoulaki, I. Kougias, N. Taylor, F. Dolci, J. Moya, and A. Jäger-Waldau, "Green hydrogen in Europe – A regional assessment: Substituting existing production with electrolysis powered by renewables," *Energy Conversion and Management*, vol. 228, p. 113649, jan 2021. 51

[161] T. Wilberforce, Z. El Hassan, A. Durrant, J. Thompson, B. Soudan, and A. G. Olabi, "Overview of ocean power technology," *Energy*, vol. 175, pp. 165–181, may 2019. 51

[162] C. Constant, M. Kotarbinski, J. Stefek, R. Green, E. DeGeorge, and I. Baring-Gould, "Accelerating ocean-based renewable energy educational opportunities to achieve a clean energy future," *Progress in Energy*, vol. 3, no. 4, p. 042002, aug 2021. [Online]. Available: https://iopscience.iop.org/article/10.1088/2516-1083/ac1509 51, 53

[163] P. Bento, J. Pombo, R. Mendes, M. Calado, and S. Mariano, "Ocean wave energy forecasting using optimised deep learning neural networks," *Ocean Engineering*, vol. 219, p. 108372, jan 2020. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0029801820312798 51, 252, 254

[164] M. Z. A. . Khan, H. A. . Khan, M. Zafar, A. Khan, H. A. Khan, and M. Aziz, "Harvesting Energy from Ocean: Technologies and Perspectives," *Energies*, vol. 15, no. 9, p. 3456, may 2022. [Online]. Available: https://www.mdpi.com/1996-1073/15/9/3456/htm 52

[165] A. Rahman, O. Farrok, and M. M. Haque, "Environmental impact of renewable energy source based electrical power plants: Solar, wind, hydroelectric, biomass, geothermal, tidal, ocean, and osmotic," *Renewable and Sustainable Energy Reviews*, vol. 161, p. 112279, jun 2022. 52

[166] M. Banja, R. Sikkema, M. Jégard, V. Motola, and J. F. Dallemand, “Biomass for energy in the EU – The support framework,” *Energy Policy*, vol. 131, pp. 215–228, aug 2019. 52

[167] A. Wąs, P. Sulewski, N. Gerasymchuk, L. Stepasyuk, V. Krupin, Z. Titenko, and K. Pogodzińska, “The Potential of Ukrainian Agriculture’s Biomass to Generate Renewable Energy in the Context of Climate and Political Challenges—The Case of the Kyiv Region,” *Energies*, vol. 15, no. 18, p. 6547, sep 2022. [Online]. Available: https://www.mdpi.com/1996-1073/15/18/6547/htm 53

[168] Y. Niu and A. Korneev, “Present situation of biomass energy utilization-a comparison between China and the United States,” in *2022 2nd International Conference on Intelligent Technologies (CONIT)*. IEEE, jun 2022, pp. 1–5. [Online]. Available: https://ieeexplore.ieee.org/document/9848238/ 53

[169] F. B. Dovichi Filho, Y. Castillo Santiago, E. E. Silva Lora, J. C. Escobar Palacio, and O. A. Almazan del Olmo, “Evaluation of the maturity level of biomass electricity generation technologies using the technology readiness level criteria,” *Journal of Cleaner Production*, vol. 295, p. 126426, may 2021. 53

[170] A. Pérez-Navarro, D. Alfonso, C. Álvarez, F. Ibáñez, C. Sánchez, and I. Segura, “Hybrid biomass-wind power plant for reliable energy generation,” *Renewable Energy*, vol. 35, no. 7, pp. 1436–1443, jul 2010. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0960148109005679 53

[171] F. Rosillo-Calle, “A review of biomass energy-shortcomings and concerns,” *J Chem Technol Biotechnol*, vol. 91, pp. 1933–1945, 2016. [Online]. Available: www.soci.org 53

[172] I. Butnar, O. Broad, B. Solano Rodriguez, and P. E. Dodds, “The role of bioenergy for global deep decarbonization: CO2 removal or low-carbon energy?” *GCB Bioenergy*, vol. 12, no. 3, pp. 198–212, mar 2020. [Online]. Available: https://onlinelibrary.wiley.com/doi/full/10.1111/gcbb.12666 53

[173] Y. T. Shah, *Hybrid Power Generation, Storage, and Grids*, 1st ed. CRC Press, 2021. [Online]. Available: https://www.routledge.com/Hybrid-Power-Generation-Storage-and-Grids/Shah/p/book/9780367678401 53

[174] H. Shahbeig, A. Shafizadeh, M. A. Rosen, and B. F. Sels, “Exergy sustainability analysis of biomass gasification: a critical review,” *Biofuel Research Journal*, vol. 33, pp. 1592–1607, 2022. [Online]. Available: www.biofueljournal.com 53

[175] J. Li, G. Wang, Z. Li, S. Yang, W. T. Chong, and X. Xiang, “A review on development of offshore wind energy conversion system,” *International Journal of Energy Research*, vol. 44, no. 12, pp. 9283–9297, oct 2020. [Online]. Available: https://onlinelibrary.wiley.com/doi/full/10.1002/er.5751 53

[176] M. Shields, P. Beiter, J. Nunemaker, A. Cooperman, and P. Duffy, "Impacts of turbine and plant upsizing on the levelized cost of energy for offshore wind," *Applied Energy*, vol. 298, p. 117189, sep 2021. 53

[177] A. Bensalah, G. Barakat, and Y. Amara, "Electrical generators for large wind turbine: Trends and challenges," *Energies*, vol. 15, no. 18, 2022. [Online]. Available: https://www.mdpi.com/1996-1073/15/18/6700 53

[178] H. Polinder, J. A. Ferreira, B. B. Jensen, A. B. Abrahamsen, K. Atallah, and R. A. McMahon, "Trends in wind turbine generator systems," *IEEE Journal of Emerging and Selected Topics in Power Electronics*, vol. 1, no. 3, pp. 174–185, 2013. 53

[179] J.-M. Roldan-Fernandez, J. Serrano-Gonzalez, F. Gonzalez-Longatt, M. Burgos-Payan, J.-M. . Roldan-Fernandez, J. . Serrano-Gonzalez, and G. Soares, "Impact of Spanish Offshore Wind Generation in the Iberian Electricity Market: Potential Savings and Policy Implications," *Energies*, vol. 14, no. 15, p. 4481, jul 2021. [Online]. Available: https://www.mdpi.com/1996-1073/14/15/4481/htm 53, 54

[180] A. Moore, J. Price, and M. Zeyringer, "The role of floating offshore wind in a renewable focused electricity system for Great Britain in 2050," *Energy Strategy Reviews*, vol. 22, pp. 270–278, nov 2018. 53

[181] S. Carreno-Madinabeitia, G. Ibarra-Berastegi, J. Sáenz, and A. Ulazia, "Long-term changes in offshore wind power density and wind turbine capacity factor in the Iberian Peninsula (1900–2010)," *Energy*, vol. 226, p. 120364, jul 2021. 53

[182] Y. L. He, Y. Qiu, K. Wang, F. Yuan, W. Q. Wang, M. J. Li, and J. Q. Guo, "Perspective of concentrating solar power," *Energy*, vol. 198, p. 117373, may 2020. 54

[183] J. Lilliestam and R. Pitz-Paal, "Concentrating solar power for less than USD 0.07 per kWh: finally the breakthrough?" *Renewable Energy Focus*, vol. 26, pp. 17–21, sep 2018. 54

[184] A. Gamarra, S. Banacloche, Y. Lechon, and P. del Río, "Assessing the sustainability impacts of concentrated solar power deployment in Europe in the context of global value chains," *Renewable and Sustainable Energy Reviews*, vol. 171, p. 113004, jan 2023. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1364032122008851 54

[185] G. San Miguel and B. Corona, "Economic viability of concentrated solar power under different regulatory frameworks in Spain," *Renewable and Sustainable Energy Reviews*, vol. 91, pp. 205–218, aug 2018. 54

[186] Office of Electricity, "Demand Response | Department of Energy," 2023. [Online]. Available: https://www.energy.gov/oe/demand-response 54

[187] A. Stawska, N. Romero, M. de Weerdt, and R. Verzijlbergh, "Demand response: For congestion management or for grid balancing?" *Energy Policy*, vol. 148, p. 111920, jan 2021. 54

[188] A. R. Jordehi, "Optimisation of demand response in electric power systems, a review," *Renewable and Sustainable Energy Reviews*, vol. 103, pp. 308–319, apr 2019. 54, 55

[189] Y. Chen, P. Xu, J. Gu, F. Schmidt, and W. Li, "Measures to improve energy demand flexibility in buildings for demand response (DR): A review," *Energy and Buildings*, vol. 177, pp. 125–139, oct 2018. 55, 56

[190] International Energy Agency, "Demand Response," IEA, Paris, Tech. Rep., 2021. [Online]. Available: https://www.iea.org/reports/demand-response 55

[191] M. Bilton, M. Aunedi, M. Woolf, and G. Strbac, "Smart appliances for residential demand response Low Carbon London Learning Lab," Imperial College London, London, Tech. Rep., 2014. 55

[192] F. S. of Regulation, "The Spanish experience with dynamic tariffs - Florence School of Regulation," nov 2020. [Online]. Available: https://fsr.eui.eu/the-spanish-experience-with-dynamic-tariffs/#_ftn1 56

[193] A. Conchado, P. Linares, O. Lago, and A. Santamaría, "An estimation of the economic and environmental benefits of a demand-response electricity program for Spain," *Sustainable Production and Consumption*, vol. 8, pp. 108–119, oct 2016. 56

[194] S. Annala, G. Mendes, S. Honkapuro, L. Matos, and L. P. Klein, "Comparison of opportunities and challenges in demand response pilots in Finland and Portugal," *International Conference on the European Energy Market, EEM*, vol. 2018-June, sep 2018. 56

[195] F. Bandeiras, M. Gomes, P. Coelho, and J. Fernandes, "Towards net zero energy in industrial and commercial buildings in Portugal," *Renewable and Sustainable Energy Reviews*, vol. 119, p. 109580, mar 2020. 56

[196] M. L. Kubik, P. J. Coker, and J. F. Barlow, "Increasing thermal plant flexibility in a high renewables power system," *Applied Energy*, vol. 154, pp. 102–111, sep 2015. 57

[197] A. Fernández-Guillamón, E. Gómez-Lázaro, and Á. Molina-García, "Extensive frequency response and inertia analysis under high renewable energy source integration scenarios: application to the European interconnected power system," *IET Renewable Power Generation*, vol. 14, no. 15, pp. 2885–2896, nov 2020. [Online]. Available: https://onlinelibrary.wiley.com/doi/full/10.1049/iet-rpg.2020.0045 57

[198] R. Lozano and A. Reid, “Socially responsible or reprehensible? Investors, electricity utility companies, and transformative change in Europe,” *Energy Research & Social Science*, vol. 37, pp. 37–43, mar 2018. 57

[199] J. Cochran, D. Lew, and N. Kumar, “Flexible coal: Evolution from baseload to peaking plant (brochure),” 12 2013. [Online]. Available: http://www.osti.gov/servlets/purl/1110465/ 57

[200] J. Blondeau and J. Mertens, “Impact of intermittent renewable energy production on specific CO2 and NOx emissions from large scale gas-fired combined cycles,” *Journal of Cleaner Production*, vol. 221, pp. 261–270, jun 2019. 57

[201] R. Cowell, “The role of place in energy transitions: Siting gas-fired power stations and the reproduction of high-carbon energy systems,” *Geoforum*, vol. 112, pp. 73–84, jun 2020. 58

[202] U.S. Energy Information Administration, “Most planned U.S. natural gas-fired plants are near Appalachia and in Florida and Texas,” nov 2021. [Online]. Available: https://www.eia.gov/todayinenergy/detail.php?id=50436 58

[203] M. M. M. Naik, V. S. S. Murty, and B. D. Prasad, “Combined Cycle Power Plant Performance Enhancement Based on Inlet Air Cooling Techniques: A Technical Review,” *International Journal of Engineering Research and*, vol. 6, no. 12, dec 2017. 58

[204] Office of Energy Efficiency & Renewable Energy, “Combined heat and power technology fact sheet series: Gas turbines,” pp. 1–4, 7 2016. [Online]. Available: https://betterbuildingssolutioncenter.energy.gov/sites/default/files/attachments/CHP_Gas_Turbines.pdf 58

[205] U.S. Energy Information Administration, “U.S. Energy Information Administration - EIA - Independent Statistics and Analysis,” feb 2019. [Online]. Available: https://www.eia.gov/todayinenergy/detail.php?id=38312 58

[206] D. Talah and H. Bentarzi, “A general overview of combined cycle gas turbine plants,” *Algerian Journal of Signals and Systems*, vol. 7, pp. 135–155, 1 2022. [Online]. Available: https://www.ajss.dz/index.php/ajss/article/view/175 58

[207] U.S. Energy Information Administration, “Most combined-cycle power plants employ two combustion turbines with one steam turbine,” apr 2022. [Online]. Available: https://www.eia.gov/todayinenergy/detail.php?id=52158 58

[208] L. Hörnlein, “The value of gas-fired power plants in markets with high shares of renewable energy,” *Energy Economics*, vol. 81, pp. 1078–1098, jun 2019. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0140988319301276 59

[209] International Energy Agency, “Coal 2021,” IEA, Paris, Tech. Rep., dec 2021. [Online]. Available: https://www.iea.org/reports/coal-2021 59

[210] U. K. Pata, S. Erdogan, and B. Ozcan, “Evaluating the role of the share and intensity of renewable energy for sustainable development in Germany,” *Journal of Cleaner Production*, vol. 421, p. 138482, oct 2023. 59

[211] R. Rasheed, H. Javed, A. Rizwan, F. Sharif, A. Yasar, A. B. Tabinda, S. R. Ahmad, Y. Wang, and Y. Su, “Life cycle assessment of a cleaner supercritical coal-fired power plant,” *Journal of Cleaner Production*, vol. 279, p. 123869, jan 2021. 59

[212] H. Nomoto, “Advanced ultra-supercritical pressure (A-USC) steam turbines and their combination with carbon capture and storage systems (CCS),” *Advances in Steam Turbines for Modern Power Plants*, pp. 501–519, jan 2017. 59

[213] B. Tramošljika, P. Blecich, I. Bonefačić, and V. Glažar, “Advanced Ultra-Supercritical Coal-Fired Power Plant with Post-Combustion Carbon Capture: Analysis of Electricity Penalty and CO2 Emission Reduction,” *Sustainability*, vol. 13, no. 2, p. 801, jan 2021. [Online]. Available: https://www.mdpi.com/2071-1050/13/2/801/htm 60

[214] M. Finkenrath, J. Smith, and D. Volk, “Analysis of the Globally Installed Coal-Fired Power Plant Fleet,” *IEA Energy Papers*, no. 07, 2012. [Online]. Available: www.iea.org/about/copyright.asp 60

[215] Y. Yang, C. Li, N. Wang, and Z. Yang, “Progress and prospects of innovative coal-fired power plants within the energy internet,” *Global Energy Interconnection*, vol. 2, no. 2, pp. 160–179, apr 2019. 60

[216] M. Jaszczur, M. Dudek, M. A. Rosen, and Z. Kolenda, “An analysis of integration of a power plant with a lignite superheated steam drying unit,” *Journal of Cleaner Production*, vol. 243, p. 118635, jan 2020. 60

[217] G. Tibba, “Modeling the Inelastic Behavior of Heat Exchangers Accounting for Fluid-Structure Interactions,” PhD Thesis, Universität Magdeburg, 2013. [Online]. Available: https://www.researchgate.net/publication/283906648_Modeling_the_Inelastic_Behavior_of_Heat_Exchangers_Accounting_for_Fluid-Structure_Interactions 60

[218] S. H. Kim, T. A. Taiwo, and B. W. Dixon, “The Carbon Value of Nuclear Power Plant Lifetime Extensions in the United States,” *Nuclear Technology*, vol. 208, no. 5, pp. 775–793, 2022. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/00295450.2021.1951554 61

[219] F. Pomponi and J. Hart, “The greenhouse gas emissions of nuclear energy – Life cycle assessment of a European pressurised reactor,” *Applied Energy*, vol. 290, p. 116743, may 2021. 61

[220] World Nuclear Association, “Nuclear Power in the European Union - World Nuclear Association,” mar 2022. [Online]. Available: https://world-nuclear.org/information-library/country-profiles/others/european-union.aspx 61

[221] International Atomic Energy Agency, “IAEA Increases Projections for Nuclear Power Use in 2050 | IAEA,” 2021. [Online]. Available: https://www.iaea.org/newscenter/pressreleases/iaea-increases-projections-for-nuclear-power-use-in-2050 61

[222] GlobalData Energy, “Spain on track to complete nuclear power phase-out by 2035 - Power Technology,” oct 2021. [Online]. Available: https://www.power-technology.com/comment/spain-nuclear-power-phase-out/ 61

[223] S. Brinton and J. Freed, “When Nuclear Ends: How Nuclear Retirements Might Undermine Clean Power Plan Progress – Third Way,” ThirdWay Publishing, Washington, DC, Tech. Rep., aug 2015. [Online]. Available: https://www.thirdway.org/report/when-nuclear-ends-how-nuclear-retirements-might-undermine-clean-power-plan-progress#:\protect\TU\textdollar\sim\protect\TU\textdollar:text=IfAmerica%27snuclearplantsbegin,wepreservedthenuclearfleet. 62

[224] P. Fernández-Arias, D. Vergara, and J. A. Orosa, “A Global Review of PWR Nuclear Power Plants,” *Applied Sciences*, vol. 10, no. 13, p. 4434, jun 2020. [Online]. Available: https://www.mdpi.com/2076-3417/10/13/4434/htm 62

[225] S. Chavan, R. Rudrapati, and S. Manickam, “A comprehensive review on current advances of thermal energy storage and its applications,” *Alexandria Engineering Journal*, vol. 61, no. 7, pp. 5455–5463, jul 2022. 63

[226] V. Spasic, “Thermal energy storage TWEST secures production of CO2 free electricity in thermal power plants |RENEWABLES,” sep 2021. [Online]. Available: https://balkangreenenergynews.com/thermal-energy-storage-twest-enables-co2-free-electricity-production-at-thermal-power-plants/ 64

[227] P. Newswire, “E2S Power Agreement with India Power Paves the Way for Decarbonizing Fossil Fuel Power Plants,” dec 2022. [Online]. Available: https://www.prnewswire.com/news-releases/e2s-power-agreement-with-india-power-paves-the-way-for-decarbonizing-fossil-fuel-power-plants-301705458.html 64

[228] R. Kumar, R. Jilte, K. C. Nikam, and M. H. Ahmadi, “Status of carbon capture and storage in India’s coal fired power plants: A critical review,” *Environmental Technology & Innovation*, vol. 13, pp. 94–103, feb 2019. 64

[229] R. S. Elias, M. I. Wahab, and L. Fang, “Retrofitting carbon capture and storage to natural gas-fired power plants: A real-options approach,” *Journal of Cleaner Production*, vol. 192, pp. 722–734, aug 2018. 64

[230] E. Barbera, A. Mio, A. Massi Pavan, A. Bertucco, and M. Fermeglia, "Fuelling power plants by natural gas: An analysis of energy efficiency, economical aspects and environmental footprint based on detailed process simulation of the whole carbon capture and storage system," *Energy Conversion and Management*, vol. 252, p. 115072, jan 2022. 64

[231] R. L. Siegelman, P. J. Milner, E. J. Kim, S. C. Weston, and J. R. Long, "Challenges and opportunities for adsorption-based CO2 capture from natural gas combined cycle emissions," *Energy & Environmental Science*, vol. 12, no. 7, pp. 2161–2173, jul 2019. [Online]. Available: https://pubs.rsc.org/en/content/articlehtml/2019/ee/c9ee00505f 64

[232] A. Babin, C. Vaneeckhaute, and M. C. Iliuta, "Potential and challenges of bioenergy with carbon capture and storage as a carbon-negative energy source: A review," *Biomass and Bioenergy*, vol. 146, p. 105968, mar 2021. 65

[233] X. Sun, J. Alcalde, M. Bakhtbidar, J. Elío, V. Vilarrasa, J. Canal, J. Ballesteros, N. Heinemann, S. Haszeldine, A. Cavanagh, D. Vega-Maza, F. Rubiera, R. Martínez-Orio, G. Johnson, R. Carbonell, I. Marzan, A. Travé, and E. Gomez-Rivas, "Hubs and clusters approach to unlock the development of carbon capture and storage – case study in spain," *Applied Energy*, vol. 300, p. 117418, 10 2021. 65

[234] S. G. Carroll, "Spanish eu presidency draws criticism on carbon capture technology," 7 2023. [Online]. Available: https://www.euractiv.com/section/agriculture-food/news/spanish-eu-presidency-draws-criticism-on-carbon-capture-technology/ 65

[235] Direção-Geral de Energia e Geologia, "Carbon dioxide geological storage," 3 2023. [Online]. Available: https://www.dgeg.gov.pt/en/vertical-areas/geology/petroleum-co2-geological-storage/carbon-dioxide-geological-storage/ 65

[236] B. Schäfer, C. Beck, K. Aihara, D. Witthaut, and M. Timme, "Non-gaussian power grid frequency fluctuations characterized by lévy-stable laws and superstatistics," *Nature Energy*, vol. 3, no. 2, pp. 119–126, 2018. [Online]. Available: https://doi.org/10.1038/s41560-017-0058-z 67

[237] International Energy Agency, "Global EV Outlook 2023," IEA, Paris, Tech. Rep., 2023. [Online]. Available: https://www.iea.org/reports/global-ev-outlook-2023 69

[238] A. Thomas and J. D. Erickson, "Rethinking the geography of energy transitions: low carbon energy pathways through energyshed design," *Energy Research & Social Science*, vol. 74, p. 101941, 4 2021. 69

[239] M. Blonsky, A. Nagarajan, S. Ghosh, K. McKenna, S. Veda, and B. Kroposki, "Potential Impacts of Transportation and Building Electrification on the Grid: A Review of Electrification Projections and Their Effects on Grid Infrastructure, Operation, and Planning," *Current Sustainable/Renewable Energy Reports*, vol. 6, no. 4, pp. 169–176, dec 2019. [Online]. Available: https://link.springer.com/article/10.1007/s40518-019-00140-5 70

[240] A. Syla, A. Rinaldi, D. Parra, and M. K. Patel, "Optimal capacity planning for the electrification of personal transport: The interplay between flexible charging and energy system infrastructure," *Renewable and Sustainable Energy Reviews*, vol. 192, p. 114214, 3 2024. 70

[241] T. T. Mai, P. Jadun, J. S. Logan, C. A. McMillan, M. Muratori, D. C. Steinberg, L. J. Vimmerstedt, B. Haley, R. Jones, and B. Nelson, "Electrification futures study: scenarios of electric technology adoption and power consumption for the united states," 2018. 70

[242] T. Meintjes, R. Castro, and A. J. Pires, "Impact of vehicle charging on portugal's national electricity load profile in 2030," *Utilities Policy*, vol. 73, p. 101310, 12 2021. 70

[243] R. Lauvergne, Y. Perez, M. Françon, and A. Tejeda De La Cruz, "Integration of electric vehicles into transmission grids: A case study on generation adequacy in Europe in 2040," *Applied Energy*, vol. 326, p. 120030, nov 2022. 70

[244] International Renewable Energy Agency, "Innovation Outlook: Smart Charging for Electric Vehicles," IRENA, Abu Dhabi, Tech. Rep., 2019. 70

[245] M. Carrión and R. Zárate-Miñano, "Operation of renewable-dominated power systems with a significant penetration of plug-in electric vehicles," *Energy*, vol. 90, pp. 827–835, 10 2015. 70

[246] G. Blumberg, R. Broll, and C. Weber, "The impact of electric vehicles on the future European electricity system – A scenario analysis," *Energy Policy*, vol. 161, p. 112751, feb 2022. 70

[247] L. Wang, V. Nian, H. Li, and J. Yuan, "Impacts of electric vehicle deployment on the electricity sector in a highly urbanised environment," *Journal of Cleaner Production*, vol. 295, p. 126386, may 2021. 71

[248] P. M. Bento, S. J. Mariano, J. A. Pombo, and M. R. Calado, "Large-scale penetration of renewables in the iberian power system: Evolution, challenges and flexibility options," *Renewable and Sustainable Energy Reviews*, vol. 204, p. 114794, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1364032124005203 71

[249] L. A. Lambert, J. Tayah, C. Lee-Schmid, M. Abdalla, I. Abdallah, A. H. Ali, S. Esmail, and W. Ahmed, "The EU's natural gas Cold War and diversification challenges," *Energy Strategy Reviews*, vol. 43, p. 100934, sep 2022. 71

[250] C. Agathokleous and J. Ehnberg, "A quantitative study on the requirement for additional inertia in the european power system until 2050 and the potential role of wind power," *Energies*, vol. 13, no. 9, 2020. [Online]. Available: https://www.mdpi.com/1996-1073/13/9/2309 72

[251] L. Yang, Z. Hu, S. Xie, S. Kong, and W. Lin, "Adjustable virtual inertia control of supercapacitors in pv-based ac microgrid cluster," *Electric Power Systems Research*, vol. 173, pp. 71–85, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0378779619301440 72

[252] R. Jones, B. Haley, G. Kwok, J. Hargreaves, and J. Williams, "Electrification and the future of electricity markets: Transitioning to a low-carbon energy system," *IEEE Power and Energy Magazine*, vol. 16, no. 4, pp. 79–89, 2018. 72

[253] D. Wozabal, C. Graf, and D. Hirschmann, "The effect of intermittent renewables on the electricity price variance," *OR Spectrum*, vol. 38, no. 3, pp. 687–709, Jul 2016. [Online]. Available: https://doi.org/10.1007/s00291-015-0395-x 73

[254] E. L. Boasson, M. D. Leiren, and J. Wettestad, *Comparative Renewables Policy - Political, Organizational and European Fields*, 1st ed., T. Hoerber and J. Fairbrass, Eds. Routledge, 2020. [Online]. Available: https://doi.org/10.4324/9780429198144 75

[255] M. Xiao, T. Junne, J. Haas, and M. Klein, "Plummeting costs of renewables - are energy scenarios lagging?" *Energy Strategy Reviews*, vol. 35, p. 100636, 5 2021. 75

[256] G. M. Pereira, R. Castro, and P. Santos, "Carbon emissions and renewables' share in the future iberian power system," *Inventions*, vol. 7, p. 4, 12 2021. [Online]. Available: https://www.mdpi.com/2411-5134/7/1/4/htm 75

[257] R. Langarita, I. Cazcarro, J. Sánchez-Chóliz, and C. Sarasa, "The role of fiscal measures in promoting renewable electricity in spain," *Energy Conversion and Management*, vol. 244, p. 114480, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0196890421006567 75

[258] M. E. Biresselioglu, M. H. Demir, M. D. Kaplan, and B. Solak, "Individuals, collectives, and energy transition: Analysing the motivators and barriers of european decarbonisation," *Energy Research & Social Science*, vol. 66, p. 101493, 8 2020. 75

[259] P. Caputo, *The Role of Renewable Energy Sources in Green Planning of Cities and Communities*. Springer, Cham, 2020, pp. 229–251. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-030-41072-8_10 75

[260] S. Hampton, T. Fawcett, J. Rosenow, C. Michaelis, and R. Mayne, "Evaluation in an emergency: Assessing transformative energy policy amidst the climate crisis," *Joule*, vol. 5, pp. 285–289, 2 2021. 75

[261] L. Bird, D. Lew, M. Milligan, E. M. Carlini, A. Estanqueiro, D. Flynn, E. Gomez-Lazaro, H. Holttinen, N. Menemenlis, A. Orths, P. B. Eriksen, J. C. Smith, L. Soder, P. Sorensen, A. Altiparmakis, Y. Yasuda, and J. Miller, "Wind and solar energy curtailment: A review of international experience," *Renewable and Sustainable Energy Reviews*, vol. 65, pp. 577–586, 11 2016. 75

[262] W. H. Wang, V. Moreno-Casas, J. H. D. Soto, A. Peña-Ramos, L. Aldieri, and P. V. Schaeffer, "A free-market environmentalist transition toward renewable energy: The cases of germany, denmark, and the united kingdom," *Energies*, vol. 14, p. 4659, 7 2021. [Online]. Available: https://www.mdpi.com/1996-1073/14/15/4659/htm 76

[263] M. M. Mwampashi, C. S. Nikitopoulos, O. Konstandatos, and A. Rai, "Wind generation and the dynamics of electricity prices in australia," *Energy Economics*, vol. 103, p. 105547, 11 2021. 76, 90

[264] J. Meus, H. Pittomvils, S. Proost, and E. Delarue, "Distortions of national policies to renewable energy cooperation mechanisms," *The Energy Journal*, vol. 43, 10 2022. [Online]. Available: www.iaee.org/energyjournal/article/3854 76

[265] J. Meus, S. D. Vits, N. S'heeren, E. Delarue, and S. Proost, "Renewable electricity support in perfect markets: Economic incentives under diverse subsidy instruments," *Energy Economics*, vol. 94, p. 105066, 2 2021. 76

[266] N. V. Andrea Pescatori, Martin Stuermer, "Surging energy prices may not ease until next year – imf blog," Blog Post, Oct. 2021. [Online]. Available: https://blogs.imf.org/2021/10/21/surging-energy-prices-may-not-ease-until-next-year/ 76

[267] O. Analytica, "Beyond energy, euro-area inflation remains modest," *Expert Briefings*, 2021. 76

[268] International Energy Agency, "What is the impact of increasing commodity and energy prices on solar pv, wind and biofuels? – analysis - iea," 12 2021. [Online]. Available: https://www.iea.org/articles/what-is-the-impact-of-increasing-commodity-and-energy-prices-on-solar-pv-wind-and-biofuels 76

[269] A. Akrami, M. Doostizadeh, and F. Aminifar, "Power system flexibility: an overview of emergence to evolution," *Journal of Modern Power Systems and*

Clean Energy, vol. 7, pp. 987–1007, 9 2019. [Online]. Available: https://link.springer.com/article/10.1007/s40565-019-0527-4 77

[270] Z. Liu, Y. Zhang, Y. Wang, N. Wei, and C. Gu, "Development of the interconnected power grid in europe and suggestions for the energy internet in china," *Global Energy Interconnection*, vol. 3, pp. 111–119, 4 2020. 78

[271] A. Crowley-Vigneau, Y. Kalyuzhnova, and N. Ketenci, "What motivates the 'green' transition: Russian and european perspectives," *Resources Policy*, p. 103128, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0301420722005712 78

[272] P. O. of the European Union, "Commission regulation (eu) 2015/1222," 7 2015. [Online]. Available: https://eur-lex.europa.eu/legal-content/PT/ALL/?uri=CELEX:32015R1222 79

[273] J. Borges, "Mibel and the market splitting - 5th international conference on the european electricity market," in *2008 5th International Conference on the European Electricity Market, EEM*. IEEE, 2008. [Online]. Available: https://ieeexplore.ieee.org/document/4579105 81

[274] K. Jones, *Negative pricing in U.S. electric power production and distribution*, 1st ed. Emerald Group Publishing Limited, 2013, vol. 29, pp. 153–165. 81

[275] OMIE, "Anexo i-informe justificativo," 2019. [Online]. Available: https://www.omie.es/sites/default/files/inline-files/anexo_i_informe_justificativo.pdf 82

[276] I. Peña, I. Azevedo, and L. A. F. M. Ferreira, "Lessons from wind policy in portugal," *Energy Policy*, vol. 103, pp. 193–202, 2017. [Online]. Available: http://dx.doi.org/10.1016/j.enpol.2016.11.033 82, 237

[277] EDP, "Edp anticipates closure of coal plants in portugal and spain | edp.com," 7 2020. [Online]. Available: https://www.edp.com/en/news/2020/07/13/edp-anticipates-closure-coal-plants-portugal-and-spain 82

[278] M. Jabłońska, S. Viljainen, J. Partanen, and T. Kauranne, "The impact of emissions trading on electricity spot market price behavior," *International Journal of Energy Sector Management*, vol. 6, pp. 343–364, 9 2012. 83

[279] R. E. de España, "Bienvenido | esios electricidad · datos · transparencia," 2021. [Online]. Available: https://www.esios.ree.es/es?locale=es 83

[280] R. E. Nacional, "Ren data hub," 2021. [Online]. Available: https://datahub.ren.pt/ 83

[281] OMIE, "Resultados del mercado," 2021. [Online]. Available: https://www.omie.es/es 83

[282] S. Mosquera-López and A. Nursimulu, "Drivers of electricity price dynamics: Comparative analysis of spot and futures markets," *Energy Policy*, vol. 126, pp. 76–87, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0301421518307432 87

[283] Štefan Bojnec and A. Križaj, "Electricity markets during the liberalization: The case of a european union country," *Energies*, vol. 14, 2021. [Online]. Available: https://www.mdpi.com/1996-1073/14/14/4317 90

[284] G. E. Francés, "Market or geopolitics? the europeanization of eu's energy corridors," *International Journal of Energy Sector Management*, vol. 5, pp. 39–59, 4 2011. 90

[285] V. Costantini, F. Gracceva, A. Markandya, and G. Vicini, "Security of energy supply: Comparing scenarios from a european perspective," *Energy Policy*, vol. 35, pp. 210–226, 1 2007. 90

[286] A. Elbassoussy, "European energy security dilemma: major challenges and confrontation strategies," *Review of Economics and Political Science*, vol. 4, pp. 321–343, 10 2019. 90

[287] T. D. Couture, D. Jacobs, O. Zinaman, and J. Cochran, "The future of renewable energy policy: Adapting to the rise of low-cost renewables," *Renewable Energy Law and Policy Review*, vol. 10, pp. 17–36, 7 2021. [Online]. Available: https://www.elgaronline.com/view/journals/relp/10-2/relp.2021.02.02.xml 90

[288] D. P. Macedo, A. C. Marques, and O. Damette, "The impact of the integration of renewable energy sources in the electricity price formation: is the merit-order effect occurring in portugal?" *Utilities Policy*, vol. 66, p. 101080, 10 2020. 90

[289] C. F. Alves and P. D. Pinto, "Impact of special regime generation management on electricity prices: the portuguese case," *International Journal of Energy Sector Management*, vol. 16, pp. 511–528, 3 2022. 91

[290] P. P. da Silva and P. Horta, "The effect of variable renewable energy sources on electricity price volatility: the case of the iberian market," *International Journal of Sustainable Energy*, vol. 38, pp. 794–813, 9 2019. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/14786451.2019.1602126 91

[291] S. Zalzar and E. F. Bompard, "The impacts of an integrated european day-ahead and intraday electricity market on market performance: The iberian region case," *2019 IEEE Milan PowerTech, PowerTech 2019*, 6 2019. 91

[292] O. Analytica, "Environmental policy is shaping the eu's energy market," *Expert Briefings*, 2019. 91

[293] International Energy Agency, "What is behind soaring energy prices and what happens next?" 2021. [Online]. Available: https://www.iea.org/commentaries/what-is-behind-soaring-energy-prices-and-what-happens-next 91

[294] J. A. Clemente-Almendros, F. T. Boldeanu, and L. A. Seguí-Amórtegui, "Impact of covid-19 on listed european electricity companies: a comparative analysis of investment in renewable and traditional electricity," *Journal of Economic Studies*, vol. ahead-of-print, 2021. 92

[295] T. L. Afonso, A. C. Marques, and J. A. Fuinhas, "Accommodating renewable energy sources in a small electricity market: An analysis considering the interactions of sources within portugal and spain," *Heliyon*, vol. 5, p. e02354, 8 2019. 92

[296] O. Prokhorov and D. Dreisbach, "The impact of renewables on the incidents of negative prices in the energy spot markets," *Energy Policy*, vol. 167, p. 113073, 8 2022. 92

[297] ERSE, "Comunicado erse propõe medidas extraordinárias para minimizar o efeito dos preços grossistas no mercado de energia," 10 2021. 92

[298] J. Bueno, D. Romero-Jordán, and P. del Río, "Analysing the drivers of electricity demand in spain after the economic crisis," *Energies*, vol. 13, 2020. [Online]. Available: https://www.mdpi.com/1996-1073/13/20/5336 93

[299] T. G. Chirwa and N. M. Odhiambo, "Electricity consumption and economic growth : New evidence from advanced, emerging and developing economies," *International Journal of Energy Sector Management*, vol. 14, pp. 1–19, 1 2020. 93

[300] S. Serfaty, "Lagrange and the calculus of variations," *Lettera Matematica*, vol. 2, no. 1, pp. 39–46, 2014. [Online]. Available: https://doi.org/10.1007/s40329-014-0049-x 96

[301] J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd ed., ser. Springer Series in Operations Research and Financial Engineering. Springer, 2006. 96, 99, 100, 102

[302] D.-Z. Du, P. M. Pardalos, and W. Wu, *History of Optimization*. Boston, MA: Springer US, 2009, pp. 1538–1542. [Online]. Available: https://doi.org/10.1007/978-0-387-74759-0_268 96

[303] P. G. Harhammer, *Optimization in Operation of Electric and Energy Power Systems*. Boston, MA: Springer US, 2009, pp. 2824–2832. [Online]. Available: https://doi.org/10.1007/978-0-387-74759-0_485 97

[304] F. A. Lootsma, *Optimization and Decision Support Systems*. Boston, MA: Springer US, 2009, pp. 2803–2807. [Online]. Available: https://doi.org/10.1007/978-0-387-74759-0_481 97

[305] M. Janga Reddy and D. Nagesh Kumar, "Evolutionary algorithms, swarm intelligence methods, and their applications in water resources engineering: a state-of-the-art review," *H2Open Journal*, vol. 3, no. 1, pp. 135–188, 06 2020. [Online]. Available: https://doi.org/10.2166/h2oj.2020.128 97, 100

[306] S. P. Bradley, A. C. Hax, and T. L. Magnanti, *Applied Mathematical Programming*. Addison-Wesley Publishing Company, 1977. [Online]. Available: https://books.google.pt/books?id=MSWdWv3Gn5cC 97

[307] V. Piccialli and M. Sciandrone, "Nonlinear optimization and support vector machines," *Annals of Operations Research*, vol. 314, no. 1, pp. 15–47, Jul 2022. [Online]. Available: https://doi.org/10.1007/s10479-022-04655-x 99

[308] R. Parker and C. Coffrin, "Managing power balance and reserve feasibility in the ac unit commitment problem," 2024. [Online]. Available: https://arxiv.org/abs/2404.00200 99

[309] S. J. Wright, "Continuous optimization (nonlinear and linear programming)," *Foundations of Computer-Aided Process Design*, vol. 7, 1999. 100

[310] M. H. K. Roni, M. S. Rana, H. R. Pota, M. M. Hasan, and M. S. Hussain, "Recent trends in bio-inspired meta-heuristic optimization techniques in control applications for electrical systems: a review," *International Journal of Dynamics and Control*, vol. 10, no. 3, pp. 999–1011, Jun 2022. [Online]. Available: https://doi.org/10.1007/s40435-021-00892-3 100, 104

[311] X.-S. Yang, *Nature-Inspired Optimization Algorithms*, 2nd ed. Academic Press, 9 2020. 100, 105

[312] T. Ma, "Why do local methods solve nonconvex problems?" 2021. [Online]. Available: https://arxiv.org/abs/2103.13462 101

[313] O. A. Elwakeil and J. S. Arora, "Global optimization of structural systems using two new methods," *Structural optimization*, vol. 12, no. 1, pp. 1–10, 1996. [Online]. Available: https://doi.org/10.1007/BF01270438 104

[314] S. Selvarajan, "A comprehensive study on modern optimization techniques for engineering applications," *Artificial Intelligence Review*, vol. 57, no. 8, p. 194, 2024. [Online]. Available: https://doi.org/10.1007/s10462-024-10829-9 104

[315] S. Voß, *Metaheuristics*. Boston, MA: Springer US, 2009, pp. 2061–2075. [Online]. Available: https://doi.org/10.1007/978-0-387-74759-0_367 104

[316] W. Liu and J. Wang, "Recursive elimination current algorithms and a distributed computing scheme to accelerate wrapper feature selection," *Information Sciences*, vol. 589, pp. 636–654, 4 2022. 104, 114

[317] F. A. Ozbay and B. Alatas, "A novel approach for detection of fake news on social media using metaheuristic optimization algorithms," *Elektronika ir Elektrotechnika*, vol. 25, pp. 62–67, 8 2019. 105

[318] D. Molina, A. LaTorre, and F. Herrera, "An insight into bio-inspired and evolutionary algorithms for global optimization: Review, analysis, and lessons learnt over a decade of competitions," *Cognitive Computation*, vol. 10, pp. 517–544, 8 2018. 105

[319] H. Nunes, D. Duarte, J. Pombo, S. Mariano, and M. Calado, "Comparative study of mppt metaheuristics for pv systems under partial shading conditions," in *2022 International Conference on Smart Energy Systems and Technologies (SEST)*, 2022, pp. 1–6. 105

[320] A. Kanungo, M. Mittal, and L. D. and, "Critical analysis of optimization techniques for a mrpid thermal system controller," *IETE Journal of Research*, vol. 69, no. 1, pp. 149–164, 2023. [Online]. Available: https://doi.org/10.1080/03772063.2020.1808092 105

[321] W. J. Gutjahr and R. Montemanni, "Stochastic search in metaheuristics," *International Series in Operations Research and Management Science*, vol. 272, pp. 513–540, 2019. 105

[322] J. Kennedy and R. Eberhart, "Particle swarm optimization," *Neural Networks, 1995. Proceedings., IEEE International Conference on*, vol. 4, pp. 1942–1948 vol.4, 1995. 105

[323] X. S. Yang and S. Deb, "Cuckoo search: Recent advances and applications," *Neural Computing and Applications*, vol. 24, pp. 169–174, 2014. 105

[324] S. Das, S. S. Mullick, and P. N. Suganthan, "Recent advances in differential evolution – an updated survey," *Swarm and Evolutionary Computation*, vol. 27, pp. 1–30, 4 2016. 105

[325] S. Mirjalili and A. Lewis, "The whale optimization algorithm," *Advances in Engineering Software*, vol. 95, pp. 51–67, 2016. 105

[326] Y. Li, X. Lin, and J. Liu, "An improved gray wolf optimization algorithm to solve engineering problems," *Sustainability (Switzerland)*, vol. 13, pp. 1–23, 2021. 105

[327] A. P. Piotrowski, J. J. Napiorkowski, and A. E. Piotrowska, "Population size in particle swarm optimization," *Swarm and Evolutionary Computation*, vol. 58, p. 100718, 11 2020. 105

[328] B. Morales-Castañeda, D. Zaldívar, E. Cuevas, F. Fausto, and A. Rodríguez, "A better balance in metaheuristic algorithms: Does it exist?" *Swarm and Evolutionary Computation*, vol. 54, p. 100671, 5 2020. 105

[329] R. Datta and K. Deb, "Uniform adaptive scaling of equality and inequality constraints within hybrid evolutionary-cum-classical optimization," *Soft Computing - A Fusion of Foundations, Methodologies and Applications*, vol. 20, no. 6, pp. 2367–2382, jun 2016. [Online]. Available: https://dl.acm.org/doi/10.1007/s00500-015-1646-0 107, 108

[330] M. Clerc, "Confinements and Biases in Particle Swarm Optimization," vol. 0, p. 9, 2006. 107, 120

[331] N. v. Thoai, *Global Optimization Methods for Systems of Nonlinear Equations*. Boston, MA: Springer US, 2009, pp. 1364–1366. [Online]. Available: https://doi.org/10.1007/978-0-387-74759-0_241 107

[332] R. M. Freund, "Penalty and barrier methods for constrained optimization (lecture notes, massachusetts institute of technology)," 2004. [Online]. Available: https://ocw.aprende.org/courses/sloan-school-of-management/15-084j-nonlinear-programming-spring-2004/lecture-notes/lec10_penalty_mt.pdf 108

[333] D. G. Luenberger and Y. Ye, "Penalty and Barrier Methods," in *Linear and Nonlinear Programming*, D. G. Luenberger and Y. Ye, Eds. Cham: Springer International Publishing, 2016, pp. 397–428. [Online]. Available: https://doi.org/10.1007/978-3-319-18842-3_13 108, 110

[334] A. Zobaa, S. Abdel Aleem, and A. Abdelaziz, *Classical and Recent Aspects of Power System Optimization 1st Edition*, 06 2018. 110

[335] M. Helbig and A. Engelbrecht, "Dynamic vector-evaluated pso with guaranteed convergence in the sub-swarms," in *2015 IEEE Symposium Series on Computational Intelligence*, 2015, pp. 1286–1293. 111

[336] C. Sun, Y. Jin, J. Zeng, and Y. Yu, "A two-layer surrogate-assisted particle swarm optimization algorithm," *Soft Computing*, vol. 19, pp. 1461–1475, 6 2015. 113, 114, 115, 116, 124, 125, 126

[337] Z. Tang, L. Xu, and S. Luo, "Adaptive dynamic surrogate-assisted evolutionary computation for high-fidelity optimization in engineering," *Applied Soft Computing*, vol. 127, p. 109333, 9 2022. 114

[338] C. Srithapon, P. Fuangfoo, P. K. Ghosh, A. Siritaratiwat, and R. Chatthaworn, "Surrogate-assisted multi-objective probabilistic optimal power flow for distribution network with photovoltaic generation and electric vehicles," *IEEE Access*, vol. 9, pp. 34 395–34 414, 2021. 114

[339] H. Tong, C. Huang, L. L. Minku, and X. Yao, "Surrogate models in evolutionary single-objective optimization: A new taxonomy and experimental study," *Information Sciences*, vol. 562, pp. 414–437, 7 2021. 114

[340] F. Li, Y. Li, X. Cai, and L. Gao, "A surrogate-assisted hybrid swarm optimization algorithm for high-dimensional computationally expensive problems," *Swarm and Evolutionary Computation*, vol. 72, p. 101096, 7 2022. 114, 115

[341] Z. Zhou, Y. S. Ong, M. H. Nguyen, and D. Lim, "A study on polynomial regression and gaussian process global surrogate model in hierarchical surrogate-assisted evolutionary algorithm," vol. 3, 2005, pp. 2832–2839. 114

[342] D. Han, W. Du, X. Wang, and W. Du, "A surrogate-assisted evolutionary algorithm for expensive many-objective optimization in the refining process," *Swarm and Evolutionary Computation*, vol. 69, p. 100988, 3 2022. 114

[343] J. Tian, C. Sun, Y. Tan, and J. Zeng, "Granularity-based surrogate-assisted particle swarm optimization for high-dimensional expensive optimization," *Knowledge-Based Systems*, vol. 187, p. 104815, 1 2020. 114

[344] S. M. Clarke, J. H. Griebsch, and T. W. Simpson, "Analysis of support vector regression for approximation of complex engineering analyses," *Journal of Mechanical Design, Transactions of the ASME*, vol. 127, pp. 1077–1087, 11 2005. 114

[345] V. Volz, G. Rudolph, and B. Naujoks, "Investigating uncertainty propagation in surrogate-assisted evolutionary algorithms," vol. 8. Association for Computing Machinery, Inc, 7 2017, pp. 881–888. 114

[346] J. Liu, Y. Wang, G. Sun, and T. Pang, "Multisurrogate-assisted ant colony optimization for expensive optimization problems with continuous and categorical variables," *IEEE Transactions on Cybernetics*, 2021. 114

[347] X. Cai, H. Qiu, L. Gao, C. Jiang, and X. Shao, "An efficient surrogate-assisted particle swarm optimization algorithm for high-dimensional expensive problems," *Knowledge-Based Systems*, vol. 184, p. 104901, 11 2019. 114

[348] J. Eason and S. Cremaschi, "Adaptive sequential sampling for surrogate model generation with artificial neural networks," *Computers & Chemical Engineering*, vol. 68, pp. 220–232, 9 2014. 114

[349] T. Zhang, F. Li, X. Zhao, W. Qi, and T. Liu, "A convolutional neural network-based surrogate model for multi-objective optimization evolutionary algorithm based on decomposition," *Swarm and Evolutionary Computation*, vol. 72, p. 101081, 7 2022. 114

[350] W. Wang, H. L. Liu, and K. C. Tan, "A Surrogate-Assisted Differential Evolution Algorithm for High-Dimensional Expensive Optimization Problems," *IEEE Transactions on Cybernetics*, apr 2022. 114

[351] C. Chen, X. Wang, H. Dong, and P. Wang, "Surrogate-assisted hierarchical learning water cycle algorithm for high-dimensional expensive optimization," *Swarm and Evolutionary Computation*, vol. 75, p. 101169, dec 2022. 114

[352] M. Yu, J. Liang, K. Zhao, and Z. Wu, "An aRBF surrogate-assisted neighborhood field optimizer for expensive problems," *Swarm and Evolutionary Computation*, vol. 68, p. 100972, feb 2022. 114

[353] Y. Liu, J. Liu, and S. Tan, "Decision space partition based surrogate-assisted evolutionary algorithm for expensive optimization," *Expert Systems with Applications*, vol. 214, p. 119075, mar 2023. 114

[354] S. C. Chu, Z. G. Du, Y. J. Peng, and J. S. Pan, "Fuzzy hierarchical surrogate assists probabilistic particle swarm optimization for expensive high dimensional problem," *Knowledge-Based Systems*, vol. 220, p. 106939, 5 2021. 114

[355] J. S. Pan, L. G. Zhang, S. C. Chu, C. S. Shieh, and J. Watada, "Surrogate-assisted hybrid meta-heuristic algorithm with an add-point strategy for a wireless sensor network," *Entropy 2023, Vol. 25, Page 317*, vol. 25, p. 317, 2 2023. 114

[356] W. Chen, H. Dong, P. Wang, and X. Wang, "Surrogate-assisted global transfer optimization based on adaptive sampling strategy," *Advanced Engineering Informatics*, vol. 56, p. 101914, apr 2023. 114

[357] H. Dong, P. Wang, X. Yu, and B. Song, "Surrogate-assisted teaching-learning-based optimization for high-dimensional and computationally expensive problems," *Applied Soft Computing*, vol. 99, p. 106934, 2 2021. 114

[358] G. Chen, K. Zhang, X. Xue, L. Zhang, C. Yao, J. Wang, and J. Yao, "A radial basis function surrogate model assisted evolutionary algorithm for high-dimensional expensive optimization problems," *Applied Soft Computing*, vol. 116, p. 108353, 2 2022. 114

[359] H. Yu, Y. Tan, C. Sun, and J. Zeng, "A generation-based optimal restart strategy for surrogate-assisted social learning particle swarm optimization," *Knowledge-Based Systems*, vol. 163, pp. 14–25, 1 2019. 114, 126

[360] H. Yu, Y. Tan, J. Zeng, C. Sun, and Y. Jin, "Surrogate-assisted hierarchical particle swarm optimization," *Information Sciences*, vol. 454-455, pp. 59–72, 7 2018. 114, 126

[361] P. Hu, J. S. Pan, S. C. Chu, and C. Sun, "Multi-surrogate assisted binary particle swarm optimization algorithm and its application for feature selection," *Applied Soft Computing*, vol. 121, p. 108736, 5 2022. 114

[362] H. Dong, X. Li, Z. Yang, L. Gao, and Y. Lu, "A two-layer surrogate-assisted differential evolution with better and nearest option for optimizing the spring of hydraulic series elastic actuator," *Applied Soft Computing*, vol. 100, p. 107001, 3 2021. 114, 126

[363] X. Ji, Y. Zhang, D. Gong, and X. Sun, "Dual-surrogate-assisted cooperative particle swarm optimization for expensive multimodal problems," *IEEE Transactions on Evolutionary Computation*, vol. 25, pp. 794–808, 8 2021. 114

[364] F. Li, W. Shen, X. Cai, L. Gao, and G. G. Wang, "A fast surrogate-assisted particle swarm optimization algorithm for computationally expensive problems," *Applied Soft Computing*, vol. 92, p. 106303, 7 2020. 115

[365] F. Zhao, H. Zhang, L. Wang, R. Ma, T. Xu, N. Zhu, and Jonrinaldi, "A surrogate-assisted jaya algorithm based on optimal directional guidance and historical learning mechanism," *Engineering Applications of Artificial Intelligence*, vol. 111, p. 104775, 5 2022. 115

[366] J. S. Pan, N. Liu, S. C. Chu, and T. Lai, "An efficient surrogate-assisted hybrid optimization algorithm for expensive optimization problems," *Information Sciences*, vol. 561, pp. 304–325, 6 2021. 115

[367] I. Loshchilov, "Surrogate-Assisted Evolutionary Algorithms," Theses, Université Paris Sud - Paris XI ; Institut national de recherche en informatique et en automatique - INRIA, Jan. 2013. [Online]. Available: https://theses.hal.science/tel-00823882 115

[368] H. Dong and Z. Dong, "Surrogate-assisted grey wolf optimization for high-dimensional, computationally expensive black-box problems," *Swarm and Evolutionary Computation*, vol. 57, p. 100713, 9 2020. 115

[369] C. S. K. Dash, A. K. Behera, S. Dehuri, and S. B. Cho, "Radial basis function neural networks: A topical state-of-the-art survey," *Open Computer Science*, vol. 6, pp. 33–63, 2016. 116

[370] S. Li, H. Chen, M. Wang, A. A. Heidari, and S. Mirjalili, "Slime mould algorithm: A new method for stochastic optimization," *Future Generation Computer Systems*, vol. 111, pp. 300–323, 2020. 118

[371] R. G. Regis and C. A. Shoemaker, "A stochastic radial basis function method for the global optimization of expensive functions," *INFORMS Journal on Computing*, vol. 19, pp. 497–509, 7 2007. 121

[372] G. Fu, C. Sun, Y. Tan, G. Zhang, and Y. Jin, "A surrogate-assisted evolutionary algorithm with random feature selection for large-scale expensive problems," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12269 LNCS, pp. 125–139, 2020. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-030-58112-1_9 126

[373] H. Yu, L. Kang, Y. Tan, C. Sun, and J. Zeng, “Truncation-learning-driven surrogate assisted social learning particle swarm optimization for computationally expensive problem,” *Applied Soft Computing*, vol. 97, p. 106812, dec 2020. 126

[374] H. Wang, Y. Jin, and J. Doherty, “Committee-Based Active Learning for Surrogate-Assisted Particle Swarm Optimization of Expensive Problems,” *IEEE Transactions on Cybernetics*, vol. 47, no. 9, pp. 2664–2677, sep 2017. 126

[375] P. Bento, “Sasma benchmark code,” 2023. [Online]. Available: https://gitfront.io/r/pedrobento/hWud6E83nsJt/SASMAsims/ 127

[376] K. Kassoul, N. Zufferey, N. Cheikhrouhou, and S. B. Belhaouari, “Exponential particle swarm optimization for global optimization,” *IEEE Access*, vol. 10, pp. 78 320–78 344, 2022. 127

[377] J. Ma, H. Li, J. Ma, and H. Li, “Research on Rosenbrock Function Optimization Problem Based on Improved Differential Evolution Algorithm,” *Journal of Computer and Communications*, vol. 7, no. 11, pp. 107–120, nov 2019. [Online]. Available: http://www.scirp.org/journal/PaperInformation.aspx?PaperID=96749 128

[378] Z. Guo, H. Huang, C. Deng, X. Yue, and Z. Wu, “An Enhanced Differential Evolution with Elite Chaotic Local Search,” *Computational Intelligence and Neuroscience*, vol. 2015, p. 583759, 2015. [Online]. Available: /pmc/articles/PMC4561320//pmc/articles/PMC4561320/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC4561320/ 128

[379] P. N. Suganthan, N. Hansen, J. J. Liang, K. Deb, Y. P. Chen, A. Auger, and S. Tiwari, “Problem definitions and evaluation criteria for the CEC 2005 special session on real-parameter optimization,” Singapore, Tech. Rep. January, 2005. 128, 335

[380] R. Cordeschi, “Ai turns fifty: Revisiting its origins,” *Applied Artificial Intelligence*, vol. 21, no. 4-5, pp. 259–279, 2007. [Online]. Available: https://doi.org/10.1080/08839510701252304 135

[381] D. A. Bloch, “Machine Learning: Models And Algorithms,” *SSRN Electronic Journal*, 2023. [Online]. Available: https://www.ssrn.com/abstract=4493977 135, 138

[382] S. Raschka, J. Patterson, and C. Nolet, “Machine Learning in Python: Main Developments and Technology Trends in Data Science, Machine Learning, and Artificial Intelligence,” *Information*, vol. 11, no. 4, 2020. [Online]. Available: https://www.mdpi.com/2078-2489/11/4/193 135

[383] I. H. Sarker, “Deep learning: A comprehensive overview on techniques, taxonomy, applications and research directions,” *SN Computer Science*, vol. 2, no. 6, p. 420, Aug 2021. [Online]. Available: https://doi.org/10.1007/s42979-021-00815-1 136

[384] H. Shahinzadeh, A. Mahmoudi, A. Asilian, H. Sadrarhami, M. Hemmati, and Y. Saberi, "Deep learning: A overview of theory and architectures," in *2024 20th CSI International Symposium on Artificial Intelligence and Signal Processing (AISP)*, 2024, pp. 1–11. 136

[385] J. Holdsworth and M. Scapicchio, "What Is Deep Learning? | IBM — ibm.com," https://www.ibm.com/topics/deep-learning, [Accessed 20-09-2024]. 136

[386] L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M. A. Fadhel, M. Al-Amidie, and L. Farhan, "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," *Journal of Big Data*, vol. 8, no. 1, p. 53, mar 2021. [Online]. Available: https://doi.org/10.1186/s40537-021-00444-8 136

[387] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015. [Online]. Available: https://doi.org/10.1038/nature14539 136, 168

[388] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778. 136

[389] M. S. Rahman, F. Khomh, A. Hamidi, J. Cheng, G. Antoniol, and H. Washizaki, "Machine learning application development: practitioners' insights," *Software Quality Journal*, vol. 31, no. 4, pp. 1065–1119, 2023. [Online]. Available: https://doi.org/10.1007/s11219-023-09621-9 137

[390] I. H. Sarker, "Machine Learning: Algorithms, Real-World Applications and Research Directions," *SN Computer Science*, vol. 2, no. 3, p. 160, 2021. [Online]. Available: https://doi.org/10.1007/s42979-021-00592-x 137, 138, 139

[391] K. Bayoudh, "A survey of multimodal hybrid deep learning for computer vision: Architectures, applications, trends, and challenges," *Information Fusion*, vol. 105, p. 102217, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S156625352300533X 137

[392] A. Shahraki, T. Ohlenforst, and F. Kreyß, "When machine learning meets network management and orchestration in edge-based networking paradigms," *Journal of Network and Computer Applications*, vol. 212, p. 103558, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1084804522001990 137

[393] D. Quintero, B. He, B. C. Faria, A. Jara, C. Parsons, S. Tsukamoto, R. Wale *et al.*, *IBM PowerAI: Deep Learning Unleashed on IBM Power Systems Servers*. IBM Redbooks, 2019. 140

[394] X. Sui, Q. Wu, J. Liu, Q. Chen, and G. Gu, “A review of optical neural networks,” *IEEE Access*, vol. 8, pp. 70 773–70 783, 2020. 140

[395] Y. van de Burgt and P. Gkoupidenis, “Organic materials and devices for brain-inspired computing: From artificial implementation to biophysical realism,” *MRS Bulletin*, vol. 45, pp. 631–640, 2020. [Online]. Available: https://doi.org/10.1557/mrs.2020.194 142

[396] P. M. R. Bento, “Hybrid artificial intelligence algorithms for short-term load and price forecasting in competitive electric markets,” Master’s thesis, Universidade da Beira Interior (Portugal), 2017. 142, 144, 151, 154, 158

[397] K.-L. Du, C.-S. Leung, W. H. Mow, and M. N. S. Swamy, “Perceptron: Learning, generalization, model selection, fault tolerance, and role in the deep learning era,” *Mathematics*, vol. 10, no. 24, 2022. [Online]. Available: https://www.mdpi.com/2227-7390/10/24/4730 143

[398] R. McLean, “Swarm-based algorithms for neural network training,” mathesis, Department of Computer Science, Brock University, 2019. [Online]. Available: http://hdl.handle.net/10464/14563 145, 167

[399] Q. Wang, Y. Ma, K. Zhao, and Y. Tian, “A comprehensive survey of loss functions in machine learning,” *Annals of Data Science*, vol. 9, pp. 187–212, 2022. [Online]. Available: https://doi.org/10.1007/s40745-020-00253-5 148

[400] C. Aggarwal, *The Backpropagation Algorithm*. Cham: Springer International Publishing, 2023, pp. 29–71. [Online]. Available: https://doi.org/10.1007/978-3-031-29642-0_2 149

[401] M. Li, T. Zhang, Y. Chen, and A. J. Smola, “Efficient mini-batch training for stochastic optimization,” in *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD ’14. New York, NY, USA: Association for Computing Machinery, 2014, p. 661–670. [Online]. Available: https://doi.org/10.1145/2623330.2623612 150

[402] Vidushi and M. Agarwal, *Performance Analysis on the Basis of Learning Rate*. Singapore: Springer Singapore, 2020, pp. 41–46. [Online]. Available: https://doi.org/10.1007/978-981-15-2043-3_6 152

[403] H. Apaydin, H. Feizi, M. T. Sattari, M. S. Colak, S. Shamshirband, and K.-W. Chau, “Comparative analysis of recurrent neural network architectures for reservoir inflow forecasting,” *Water*, vol. 12, no. 5, 2020. [Online]. Available: https://www.mdpi.com/2073-4441/12/5/1500 153

[404] O. V. Johnson, C. XinYing, O. E. Johnson, K. W. Khaw, and M. H. Lee, “Learning rate schedules and optimizers, a game changer for deep neural networks,” in

Advances in Intelligent Computing Techniques and Applications, F. Saeed, F. Mohammed, and Y. Fazea, Eds. Cham: Springer Nature Switzerland, 2024, pp. 327–340. 152

[405] I. Sutskever, J. Martens, G. Dahl, and G. Hinton, "On the importance of initialization and momentum in deep learning," in *Proceedings of the 30th International Conference on International Conference on Machine Learning - Volume 28*, ser. ICML'13. Atlanta, GA, USA: JMLR.org, 2013, p. III–1139–III–1147. 155, 158

[406] X. Wang, "Method of steepest descent and its applications," *IEEE Microwave and Wireless Components Letters*, vol. 12, pp. 24–26, 2008. 156

[407] D. Kumar, S. Gupta, and P. Sehgal, "Improved training of predictive ann with gradient techniques," *Proc., Int. Multiconf. of Engineers and Computer Scientists, International Association of Engineers, Hong Kong*, 2014. 156

[408] W. W. Hager and H. Zhang, "A survey of nonlinear conjugate gradient methods," 2005. [Online]. Available: https://api.semanticscholar.org/CorpusID:55715146 156

[409] M. V. Narkhede, P. P. Bartakke, and M. S. Sutaone, "A review on weight initialization strategies for neural networks," *Artificial Intelligence Review*, vol. 55, no. 1, pp. 291–322, Jan 2022. [Online]. Available: https://doi.org/10.1007/s10462-021-10033-z 158

[410] F. Y. Barrera, "Random initialization of weights in a neural network," Mar. 2024. [Online]. Available: https://www.baeldung.com/cs/ml-neural-network-weights 158

[411] "Xavier initialization." [Online]. Available: https://paperswithcode.com/method/xavier-initialization 159

[412] L. L. Lu Lu, Y. S. Yeonjong Shin, Y. S. Yanhui Su, and G. E. K. George Em Karniadakis, "Dying relu and initialization: Theory and numerical examples," *Communications in Computational Physics*, vol. 28, no. 5, p. 1671–1706, Jan. 2020. [Online]. Available: http://dx.doi.org/10.4208/cicp.OA-2020-0165 159

[413] X. Ying, "An overview of overfitting and its solutions," *Journal of Physics: Conference Series*, vol. 1168, no. 2, p. 022022, feb 2019. [Online]. Available: https://dx.doi.org/10.1088/1742-6596/1168/2/022022 160, 165, 169

[414] B. Neal, S. Mittal, A. Baratin, V. Tantia, M. Scicluna, S. Lacoste-Julien, and I. Mitliagkas, "A modern take on the bias-variance tradeoff in neural networks," 2019. [Online]. Available: https://arxiv.org/abs/1810.08591 161

[415] Z. Yang, Y. Yu, C. You, J. Steinhardt, and Y. Ma, "Rethinking bias-variance tradeoff for generalization of neural networks," in *Proceedings of the 37th International Conference on Machine Learning*, ser. ICML'20. JMLR.org, 2020. 162

[416] S. Urolagin, P. K.V., and N. V. S. Reddy, “Generalization capability of artificial neural network incorporated with pruning method,” in *Advanced Computing, Networking and Security*, P. S. Thilagam, A. R. Pais, K. Chandrasekaran, and N. Balakrishnan, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 171–178. 162, 167, 169

[417] D. Johari, T. Rahman, and I. Musirin, “Ann model selection and performance evaluation for lightning prediction system,” in *Proceedings of the 2nd Seminar and Information Technology*, 2009. 163, 165

[418] M. Yaghini, M. M. Khoshraftar, and M. Fallahi, “A hybrid algorithm for artificial neural network training,” *Engineering Applications of Artificial Intelligence*, vol. 26, no. 1, pp. 293–301, 2013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0952197612000371 163

[419] R. Bayat, M. Pezeshki, E. Dohmatob, D. Lopez-Paz, and P. Vincent, “The pitfalls of memorization: When memorization hurts generalization,” 2024. [Online]. Available: https://arxiv.org/abs/2412.07684 163

[420] J. F. Triki, “Analysis of word embeddings: A clustering and topological approach,” mathesis, Department of Informatics, The University of Bergen, Jun. 2021. [Online]. Available: https://bora.uib.no/bora-xmlui/bitstream/handle/11250/2769947/Masters-Thesis-in-ML---Jonas-Folkvord-Triki.pdf?sequence=1&isAllowed=y 164, 165

[421] W. Wang, P. H. A. J. M. Van Gelder, and J. K. Vrijling, “Some issues about the generalization of neural networks for time series prediction,” in *Artificial Neural Networks: Formal Models and Their Applications – ICANN 2005*, W. Duch, J. Kacprzyk, E. Oja, and S. Zadrożny, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 559–564. 164, 167

[422] M. H. Beale, M. T. Hagan, and H. B. Demuth, *Neural Network Toolbox™ 6 User’s Guide*, The MathWorks, Inc., 2009. [Online]. Available: https://www.mit.bme.hu/system/files/oktatas/targyak/9109/MATLAB_-_Neural_Network_Toolbox_6_-_Users_Guide_2009.pdf 165, 168, 169

[423] I. K. Nti, O. N. yarko Boateng, and J. Aning, “Performance of machine learning algorithms with different k values in k-fold crossvalidation,” *International Journal of Information Technology and Computer Science*, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:245472837 166

[424] V. C. Ebhota and V. M. Srivastava, “Performance analysis of learning rate parameter on prediction of signal power loss for network optimization and better generalization,” *Wirel. Pers. Commun.*, vol. 118, no. 2, p. 1111–1128, May 2021. [Online]. Available: https://doi.org/10.1007/s11277-020-08061-z 167

[425] D. Hunter, H. Yu, M. Pukish, J. Kolbusz, and B. Wilamowski, "Selection of proper neural network sizes and architectures—a comparative study," *Industrial Informatics, IEEE Transactions on*, vol. 8, pp. 228–240, 05 2012. 167

[426] F. S. Panchal and M. Panchal, "Review on methods of selecting number of hidden nodes in artificial neural network," *International Journal of Computer Science and Mobile Computing*, vol. 3, no. 11, pp. 455–464, 2014. 167

[427] K. G. Sheela and S. N. Deepa, "Review on methods to fix number of hidden neurons in neural networks," *Mathematical Problems in Engineering*, vol. 2013, no. 1, p. 425740, 2013. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1155/2013/425740 167

[428] T. Jin, M. Carbin, D. M. Roy, J. Frankle, and G. K. Dziugaite, "Pruning's effect on generalization through the lens of training and regularization," in *Proceedings of the 36th International Conference on Neural Information Processing Systems*, ser. NIPS '22. Red Hook, NY, USA: Curran Associates Inc., 2022. 169

[429] M. Shen, P. Molchanov, H. Yin, and J. M. Alvarez, "When to prune? a policy towards early structural pruning," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 12 237–12 246. 169

[430] K. Pratama and K. Dae-Ki, "The effect of hyperparameter choice on relu and selu activation function," *International journal of advanced smart convergence*, vol. 6, no. 4, pp. 73–79, 12 2017. 169, 170

[431] S. Santurkar, D. Tsipras, A. Ilyas, and A. Madry, "How does batch normalization help optimization? (no, it is not about internal covariate shift)," in *Neural Information Processing Systems*, 2018. [Online]. Available: https://api.semanticscholar.org/CorpusID:53104146 171

[432] S. N. Sivanandam, S. Sumathi, and S. N. Deepa, *Introduction to neural networks using MATLAB 6.0*, 1st ed. New Delhi: McGraw Hill Education (India) Private Limited, 2006. 171, 172

[433] M. Bhuvaneshwari, E. G. M. Kanaga, J. Anitha, K. Raimond, and S. T. George, "A comprehensive review on deep learning techniques for a BCI-based communication system," *Demystifying Big Data, Machine Learning, and Deep Learning for Healthcare Analytics*, pp. 131–157, jan 1 2021. 173

[434] V. Arul, "Deep learning methods for data classification," *Artificial Intelligence in Data Mining*, pp. 87–108, jan 1 2021. 173

[435] C. Okinda, I. Nyalala, T. Korohou, C. Okinda, J. Wang, T. Achieng, P. Wamalwa, T. Mang, and M. Shen, "A review on computer vision systems in monitoring of poultry: A welfare perspective," *Artificial Intelligence in Agriculture*, vol. 4, pp. 184–208, jan 1 2020. 173

[436] M. Kazemzadeh, C. L. Hisey, K. Zargar-Shoshtari, W. Xu, and N. G. Broderick, "Deep convolutional neural networks as a unified solution for Raman spectroscopy-based classification in biomedical applications," *Optics Communications*, vol. 510, p. 127977, may 1 2022. 173

[437] S. Li, L. Goel, and P. Wang, "An ensemble approach for short-term load forecasting by extreme learning machine," *Applied Energy*, vol. 170, pp. 22–29, 2016. 174

[438] P. Bento, J. Pombo, S. Mariano, and M. d. R. Calado, "Short-Term Load Forecasting using optimized LSTM Networks via Improved Bat Algorithm," in *2018 International Conference on Intelligent Systems (IS)*. IEEE, sep 2018, pp. 351–357. [Online]. Available: https://ieeexplore.ieee.org/document/8710498/ 175, 218, 258

[439] W. Guo, L. Che, M. Shahidehpour, and X. Wan, "Machine-Learning based methods in short-term load forecasting," *The Electricity Journal*, vol. 34, no. 1, p. 106884, jan 1 2021. 175, 179, 181

[440] J. Chung, Ç. Gülçehre, K. Cho, and Y. Bengio, "Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling," Tech. Rep., 2014. 175

[441] H. Jaeger, "Echo state network," *Scholarpedia*, vol. 2, no. 9, p. 2330, 2007. 176

[442] P. V. Aceituno, G. Yan, and Y. Y. Liu, "Tailoring Echo State Networks for Optimal Learning," *iScience*, vol. 23, no. 9, p. 101440, sep 25 2020. 176

[443] W. Rawat and Z. Wang, "Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review," *Neural Computation*, vol. 29, no. 9, pp. 2352–2449, sep 1 2017. 176

[444] H. Liu, "Single-point wind forecasting methods based on reinforcement learning," *Wind Forecasting in Railway Engineering*, pp. 177–214, jan 1 2021. 178

[445] J. Ren, M. Green, and X. Huang, "From traditional to deep learning: Fault diagnosis for autonomous vehicles," *Learning Control*, pp. 205–219, jan 1 2021. 178

[446] Y. Kurtgoz and E. Deniz, *Comparison of ANN, Regression Analysis, and ANFIS Models in Estimation of Global Solar Radiation for Different Climatological Locations*. Academic Press, jan 1 2018, pp. 133–148. 178, 179

[447] L. Yin, Z. Sun, F. Gao, and H. Liu, "Deep Forest Regression for Short-Term Load Forecasting of Power Systems," *IEEE Access*, vol. 8, pp. 49 090–49 099, 2020. 180

[448] G. Dudek, "Neural networks for pattern-based short-term load forecasting: A comparative study," *Neurocomputing*, vol. 205, pp. 64–74, sep 12 2016. 180

[449] S. Song, M. Wang, and Y. Lin, "An improved algorithm for incremental extreme learning machine," *Systems Science & Control Engineering*, vol. 8, no. 1, pp. 308–317, jan 1 2020. 180

[450] Q. Chen, W. Zhang, K. Zhu, D. Zhou, H. Dai, and Q. Wu, "A novel trilinear deep residual network with self-adaptive Dropout method for short-term load forecasting," *Expert Systems with Applications*, vol. 182, p. 115272, nov 15 2021. 180

[451] Z. Deng, B. Wang, Y. Xu, T. Xu, C. Liu, and Z. Zhu, "Multi-scale convolutional neural network with time-cognition for multi-step short-Term load forecasting," *IEEE Access*, vol. 7, pp. 88 058–88 071, 2019. 181

[452] V. Veeramsetty, K. R. Reddy, M. Santhosh, A. Mohnot, and G. Singal, "Short-term electric power load forecasting using random forest and gated recurrent unit," *Electrical Engineering*, vol. 104, no. 1, pp. 307–329, feb 1 2022. 181

[453] H. Chelabi, M. T. Khadir, B. Chikhaoui, and A. J. Telmoudi, "Comparison of Deep Learning Architectures for Short-Term Electrical Load Forecasting Based on Multi-Modal Data," *Cybernetics and Systems*, vol. 53, no. 1, pp. 186–207, 2022. 181

[454] N. B. Vanting, Z. Ma, and B. N. Jørgensen, "A scoping review of deep neural networks for electric load forecasting," *Energy Informatics*, vol. 4, no. 2, pp. 1–13, sep 1 2021. 181

[455] C. Fan, C. Ding, L. Xiao, F. Cheng, and Z. Ai, "Deep belief ensemble network based on MOEA/D for short-term load forecasting," *Nonlinear Dynamics*, vol. 105, no. 3, pp. 2405–2430, aug 1 2021. 181

[456] C. Medina, C. R. M. Ana, and G. González, "Transmission grids to foster high penetration of large-scale variable renewable energy sources -a review of challenges, problems, and solutions," *International Journal of Renewable Energy Research*, vol. 12, pp. 146–169, 2022. 183

[457] D. Q. Hung, M. R. Shah, and N. Mithulananthan, *Technical challenges, security and risk in grid integration of renewable energy*. Springer International Publishing, 2016, vol. 57, pp. 99–118. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-319-30427-4_6 183

[458] K. L. Maheswari, B. Sathya, and A. M. A. Jeylani, "Mitigating measures to address challenges of renewable integration—forecasting, scheduling, dispatch, balancing, monitoring, and control," *Integration of Renewable Energy Sources with Smart Grid*, pp. 281–304, 8 2021. [Online]. Available: https://onlinelibrary.wiley.com/doi/full/10.1002/9781119751908.ch13 183

[459] V. Krasko and S. Rebennack, *Chapter 15: Global Optimization: Optimal Power Flow Problem*, 2017, pp. 187–205. [Online]. Available: https://epubs.siam.org/doi/abs/10.1137/1.9781611974683.ch15 183

[460] M. Ghasemi, S. Ghavidel, M. Gitizadeh, and E. Akbari, "An improved teaching–learning-based optimization algorithm using lévy mutation strategy for non-

smooth optimal power flow," *International Journal of Electrical Power & Energy Systems*, vol. 65, pp. 375–384, 2 2015. 184

[461] J. Simões, "Modelo de optimal power flow utilizando sequential linear programming," masterthesis, Universidade do Porto, 2007. 184

[462] F. Marzbani and A. Abdelfatah, "Economic Dispatch Optimization Strategies and Problem Formulation: A Comprehensive Review," *Energies*, vol. 17, no. 3, 2024. [Online]. Available: https://www.mdpi.com/1996-1073/17/3/550 184

[463] E. Obio, S. Ali, I. Oyebanjo, D. Abara, F. Suleiman, P. Ohiero, D. Oku, V. Ogar, E. Obio, S. Ali, I. Oyebanjo, D. Abara, F. Suleiman, P. Ohiero, D. Oku, and V. Ogar, "Comparison of Economic Dispatch, OPF and Security Constrained-OPF in Power System Studies R^N," *Journal of Power and Energy Engineering*, vol. 10, no. 8, pp. 54–74, aug 2022. [Online]. Available: http://www.scirp.org/journal/PaperInformation.aspx?PaperID=119491 184

[464] J. Lin and F. H. Magnago, *Optimal Power Flow*, 1st ed. Wiley-IEEE Press, 10 2017, pp. 147–171. 184

[465] B. Stott and O. Alsac, "Optimal Power Flow — Basic Requirements for Real-Life Problems and their Solutions," 2012. [Online]. Available: https://api.semanticscholar.org/CorpusID:8051665 184

[466] Y. Y. Hong and G. F. D. Apolinario, "Uncertainty in unit commitment in power systems: A review of models, methods, and applications," *Energies*, vol. 14, p. 6658, 10 2021. [Online]. Available: https://www.mdpi.com/1996-1073/14/20/6658/htm 184

[467] J. A. Taylor, *Convex Optimization of Power Systems*. Cambridge University Press, 2015. [Online]. Available: https://www.cambridge.org/core/product/4CCA9CC35F35AE7EB222B07F2AD7FA98 184

[468] M. B. Cain, R. P. O'Neill, and A. Castillo, "History of optimal power flow and formulations," Federal Energy Regulatory Commission, Tech. Rep., 2012. [Online]. Available: https://api.semanticscholar.org/CorpusID:17073922 184

[469] P. Tang, Z. Liu, L. Zhu, K. Hou, and W. Pei, "An Online Calculation Method for Optimal Power Flow Based on Optimal Operation Region," *2022 4th International Conference on Smart Power and Internet Energy Systems, SPIES 2022*, pp. 433–437, 2022. 184

[470] M. Chamanbaz, F. Dabbene, and C. Lagoa, "AC optimal power flow in the presence of renewable sources and uncertain loads," 2019. [Online]. Available: https://arxiv.org/abs/1702.02967 184, 197

[471] O. A. Afolabi, W. H. Ali, P. Cofie, J. Fuller, P. Obiomon, and E. S. Kolawole, "Analysis of the load flow problem in power system planning studies," *Energy and Power Engineering*, vol. 7, pp. 509–523, 9 2015. [Online]. Available: http://www.scirp.org/journal/PaperInformation.aspx?PaperID=60124 184, 185, 207

[472] M. Albadi, *Power Flow Analysis*. IntechOpen, 2019. [Online]. Available: https://doi.org/10.5772/intechopen.83374 185, 205, 209

[473] S. Li, W. Gong, L. Wang, and Q. Gu, "Multi-objective optimal power flow with stochastic wind and solar power," *Applied Soft Computing*, vol. 114, p. 108045, 1 2022. 185

[474] E. Mohagheghi, M. Alramlawi, A. Gabash, and P. Li, "A Survey of Real-Time Optimal Power Flow," *Energies*, vol. 11, no. 11, p. 3142, nov 2018. [Online]. Available: https://www.mdpi.com/1996-1073/11/11/3142/htm 185

[475] D. Dharamjit, "Load flow analysis on ieee 30 bus system," *International Journal of Scientific and Research Publications*, vol. 2, 2012. [Online]. Available: http://www.ijsrp.org/research-paper-1112.php?rp=P11375 185, 206

[476] T. Kulworawanichpong, "Simplified newton–raphson power-flow solution method," *International Journal of Electrical Power & Energy Systems*, vol. 32, pp. 551–558, 7 2010. 185

[477] N. Vojnović, J. Vidaković, and M. Vidaković, "Decoupled load flow for large-scale multiphase distribution networks," *Sustainable Energy, Grids and Networks*, vol. 34, p. 101049, 6 2023. 185

[478] R. K. Portelinha, C. C. Durce, O. L. Tortelli, and E. M. Lourenço, "Fast-decoupled power flow method for integrated analysis of transmission and distribution systems," *Electric Power Systems Research*, vol. 196, p. 107215, 7 2021. 185, 209

[479] W. P. Guamán, G. N. Pesántez, X. A. Proaño, E. M. Pérez, and W. V. Tigse, "Power Flow Solution Combining Newton-Raphson and Fast Decoupled Methods," in *International Conference on Innovation and Research*, M. Botto-Tobar, M. Zambrano Vizuete, and A. Díaz Cadena, Eds. Cham: Springer International Publishing, 2021, pp. 222–233. 185

[480] K. Hussain, W. Zhu, and M. N. Mohd Salleh, "Long-term memory harris' hawk optimization for high dimensional and optimal power flow problems," *IEEE Access*, vol. 7, pp. 147 596–147 616, 2019. [Online]. Available: https://ieeexplore.ieee.org/document/8864968 185

[481] A. S. Rana, M. Bajaj, S. Gairola, A. S. Rana, M. Bajaj, and S. Gairola, "Optimal Power Flow Solution in Smart Grid Environment Using SVC and

TCSC," *Advanced Communication and Control Methods for Future Smartgrids*, nov 2019. [Online]. Available: https://www.intechopen.com/chapters/66959undefined/chapters/66959 185

[482] B. G. Risi, F. Riganti-Fulginei, and A. Laudani, "Modern techniques for the optimal power flow problem: State of the art," *Energies*, vol. 15, p. 6387, 9 2022. [Online]. Available: https://www.mdpi.com/1996-1073/15/17/6387/htm 186, 204

[483] A. Ali, A. Hassan, M. U. Keerio, N. H. Mugheri, G. Abbas, M. Hatatah, E. Touti, and A. Yousef, "A novel solution to optimal power flow problems using composite differential evolution integrating effective constrained handling techniques," *Scientific Reports*, vol. 14, no. 1, pp. 1–27, mar 2024. [Online]. Available: https://www.nature.com/articles/s41598-024-56590-5 186

[484] S. A. Mohamed, N. Anwer, and M. M. Mahmoud, "Solving optimal power flow problem for IEEE-30 bus system using a developed particle swarm optimization method: towards fuel cost minimization," *International Journal of Modelling and Simulation*, apr 2023. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/02286203.2023.2201043 186

[485] P. P. Biswas, P. N. Suganthan, R. Mallipeddi, and G. A. Amaratunga, "Optimal power flow solutions using differential evolution algorithm integrated with effective constraint handling techniques," *Engineering Applications of Artificial Intelligence*, vol. 68, pp. 81–100, 2 2018. 186

[486] A. A. El-Fergany and H. M. Hasanien, "Tree-seed algorithm for solving optimal power flow problem in large-scale power systems incorporating validations and comparisons," *Applied Soft Computing*, vol. 64, pp. 307–316, 3 2018. 186

[487] H. Xiao, Z. Dong, L. Kong, W. Pei, and Z. Zhao, "Optimal power flow using a novel metamodel based global optimization method," *Energy Procedia*, vol. 145, pp. 301–306, 7 2018. 186

[488] C. Shilaja and T. Arunprasath, "Optimal power flow using moth swarm algorithm with gravitational search algorithm considering wind power," *Future Generation Computer Systems*, vol. 98, pp. 708–715, 9 2019. 187, 200, 201

[489] V. Roberge and M. Tarbouchi, "Hybrid method based on metaheuristics and interior point for optimal power flow," *5th International Conference on Intelligent Computing in Data Sciences, ICDS 2021*, 2021. 187

[490] H. Wu, Y. Chen, Z. Cai, A. A. Heidari, H. Chen, and G. Liang, "Gradient pyramid mechanism and nelder-mead simplex enhanced colony predation algorithm for optimal power flow problems," *Energy Reports*, vol. 11, pp. 2901–2920, 6 2024. 187

[491] S. Corsi, *Relationship Between Voltage and Active and Reactive Powers*. Springer London, 2015, pp. 3–11. [Online]. Available: https://doi.org/10.1007/978-1-4471-6636-8_1 189

[492] I. Khan, V. Bhattacharjee, and M. Nasir, "Effect of the approximation of voltage angle difference on the opf algorithms in the power network," *Energy Systems*, vol. 11, pp. 471–490, 4 2019. [Online]. Available: https://link.springer.com/article/10.1007/s12667-018-00323-x 189, 209

[493] M. Gaiceanu, V. Solcanu, T. Gaiceanu, and I. Ghenea, *Numerical Methods of Electric Power Flow in Interconnected Systems*. Springer, Cham, 2021, pp. 901–933. [Online]. Available: https://doi.org/10.1007/978-3-030-62191-9_33 189

[494] M. Vaiman and M. Vaiman, "Using phase angle for steady-state voltage stability assessment," 2017. [Online]. Available: https://www.researchgate.net/publication/319493283_Using_Phase_Angle_for_Steady-State_Voltage_Stability_Assessment 190

[495] B. A. Robbins, H. Zhu, and A. D. Dominguez-Garcia, "Optimal tap setting of voltage regulation transformers in unbalanced distribution systems," *IEEE Transactions on Power Systems*, vol. 31, pp. 256–267, 1 2016. 190

[496] D. Nedic, "Tap adjustment in ac load flow," University of Manchester Institute of Science and Technology (UMIST), Tech. Rep., 9 2002. 191

[497] J. H. Ye, P. Dehghanian, and T. Overbye, "Power flow consideration of impedance correction for phase shifting transformers," *2019 IEEE Texas Power and Energy Conference, TPEC 2019*, 3 2019. 191

[498] F. D. Leon, A. Farazmand, and P. Joseph, "Comparing the t and π equivalent circuits for the calculation of transformer inrush currents," *IEEE Transactions on Power Delivery*, vol. 27, pp. 2390–2398, 2012. 191

[499] C. A. Ferreira and R. B. Prada, "Improved model for tap-changing transformer," *IET Generation, Transmission & Distribution*, vol. 7, pp. 1289–1295, 11 2013. [Online]. Available: https://onlinelibrary.wiley.com/doi/full/10.1049/iet-gtd.2012.0466 191

[500] J. M. Cano, M. R. R. Mojumdar, J. G. Norniella, and G. A. Orcajo, "Phase shifting transformer model for direct approach power flow studies," *International Journal of Electrical Power & Energy Systems*, vol. 91, pp. 71–79, 10 2017. 192

[501] O. G. Ibe, "Concepts of reactive power control and voltage stability methods in power system network," *IOSR Journal of Computer Engineering*, vol. 11, pp. 15–25, 2013. 194

[502] Q. He, W. Yang, L. He, and Y. Mao, "Evaluation of reactive power and voltage support capability of power grid based on big data," *Proceedings - 2021 International Conference on Networking, Communications and Information Technology, NetCIT 2021*, pp. 261–266, 2021. 194

[503] D. Langner, G. Pilz, and M. Gurbiel, "Application benefits of line- or self-commutated static var compensator," *IEEE Power and Energy Society General Meeting*, vol. 2014-October, 10 2014. 194

[504] B. Kazemtabrizi and E. Acha, "A comparison study between mathematical models of static var compensators aimed at optimal power flow solutions," *Proceedings of the Mediterranean Electrotechnical Conference - MELECON*, pp. 1125–1128, 2012. 194

[505] A. A. Edris, R. Adapa, M. H. Baker, L. Bohmann, K. Clark, K. Habashi, L. Gyugyi, J. Lemay, A. S. Mehraban, A. K. Myers, J. Reeve, F. Sener, D. R. Torgerson, and R. R. Wood, "Proposed terms and definitions for flexible ac transmission system (facts)," *IEEE Transactions on Power Delivery*, vol. 12, pp. 1848–1853, 1997. 194

[506] P. P. Biswas, P. Arora, R. Mallipeddi, P. N. Suganthan, and B. K. Panigrahi, "Optimal placement and sizing of facts devices for optimal power flow in a wind power integrated electrical network," *Neural Computing and Applications*, vol. 33, pp. 6753–6774, 6 2021. [Online]. Available: https://link.springer.com/article/10.1007/s00521-020-05453-x 194

[507] P. M. Tuson, "Transmission transformer tap changer optimisation while minimising system losses using security unconstrained optimal power flow (opf) techniques," Master's thesis, University of the Witwatersrand, 2011. [Online]. Available: https://core.ac.uk/download/pdf/39670461.pdf 194

[508] H. Cui, "Bus admittance matrix revisited: Performance challenges on modern computers," *IEEE Open Access Journal of Power and Energy*, vol. 11, pp. 83–93, 2024. 197

[509] S. W. Park, R. Y. Zhang, J. Lavaei, and R. Baldick, "Monotonicity between phase angles and power flow and its implications for the uniqueness of solutions," T. X. Bui, Ed., vol. 2019-Janua. IEEE Computer Society, 1 2019, pp. 3607–3616. [Online]. Available: http://hdl.handle.net/10125/59797 197

[510] S. Frank, I. Steponavice, S. Rebennack, S. Frank, I. Steponavice, and S. Rebennack, "Optimal power flow: a bibliographic survey I," *Energy Systems*, vol. 3, no. 3, pp. 221–258, apr 2012. [Online]. Available: https://link.springer.com/article/10.1007/s12667-012-0056-y 197, 198

[511] E. Naderi, H. Narimani, M. Pourakbari-Kasmaei, F. V. Cerna, M. Marzband, and M. Lehtonen, "State-of-the-Art of Optimal Active and Reactive Power Flow: A

Comprehensive Review from Various Standpoints," *Processes*, vol. 9, no. 8, p. 1319, jul 2021. [Online]. Available: https://www.mdpi.com/2227-9717/9/8/1319/htm 197, 198

[512] F. Capitanescu, W. Rosehart, and L. Wehenkel, "Optimal power flow computations with constraints limiting the number of control actions," in *2009 IEEE Bucharest PowerTech*. Bucharest, Romania: IEEE, 2009, pp. 1–8. 198

[513] M. J. Khan and Y. Kumar, "Optimal power flow (OPF) Formulation and reactive power optimization of power systems using conventional optimization technique (interior point method)," *International Journal of Mathematics Research*, vol. 8, no. 3, pp. 251–263, 2016. [Online]. Available: http://www.irphouse.com/ijmr16/ijmrv8n3_11.pdf 198

[514] Z. Wang, A. Younesi, M. V. Liu, G. C. Guo, and C. L. Anderson, "AC Optimal Power Flow in Power Systems with Renewable Energy Integration: A Review of Formulations and Case Studies," *IEEE Access*, vol. 11, pp. 102 681–102 712, 2023. 198, 201

[515] M. Al-Kaabi, V. Dumbrava, and M. Eremia, "Single and multi-objective optimal power flow based on hunger games search with pareto concept optimization," *Energies*, vol. 15, p. 8328, 11 2022. [Online]. Available: https://www.mdpi.com/1996-1073/15/22/8328/htm 198

[516] M. Abuella and C. Hatziadoniu, "Selection of most effective control variables for solving optimal power flow using sensitivity analysis in particle swarm algorithm," 2016. [Online]. Available: https://arxiv.org/abs/1601.04150 199

[517] D. Asija, P. V. Astick, and P. Choudekar, "Minimizing fuel cost of generators using ga-opf," A. K. Somani, S. Srivastava, A. Mundra, and S. Rawat, Eds. Springer Singapore, 2018, pp. 331–339. [Online]. Available: https://link.springer.com/chapter/10.1007/978-981-10-5828-8_32 199

[518] L. V. Hoorebeeck, P. A. Absil, and A. Papavasiliou, "Solving non-convex economic dispatch with valve-point effects and losses with guaranteed accuracy," *International Journal of Electrical Power & Energy Systems*, vol. 134, p. 107143, 1 2022. 200

[519] N. Sinsuphan, U. Leeton, and T. Kulworawanichpong, "Optimal power flow solution using improved harmony search method," *Applied Soft Computing*, vol. 13, pp. 2364–2374, 5 2013. 200

[520] S. Fadil, A. Yazici, and B. Urazel, "A solution to security constrained non-convex economic dispatch problem by modified subgradient algorithm based on feasible values," *International Journal of Electrical Power & Energy Systems*, vol. 43, pp. 849–858, 12 2012. 202

[521] I. Mironov, “Effect of placement of pmus on state estimation in a power system,” Master’s thesis, UiT Norges arktiske universitet, 2016. 202

[522] F. Hasan, A. Kargarian, and J. Mohammadi, “Hybrid learning aided inactive constraints filtering algorithm to enhance ac opf solution time,” *IEEE Transactions on Industry Applications*, vol. 57, pp. 1325–1334, 3 2021. 204

[523] R. N. Alcalá-Galiano, “Analysis and implementation of optimal power flow (opf) tool for grid operation and planning under future potential flexibility scenarios-application to i-de network,” Master’s thesis, Universidad Pontificia Comillas, 2021. [Online]. Available: https://repositorio.comillas.edu/xmlui/handle/11531/55707?show=full 204

[524] M. Dehghan and A. Shirilord, “Three-step iterative methods for numerical solution of systems of nonlinear equations,” *Engineering with Computers*, vol. 38, pp. 1015–1028, 4 2022. [Online]. Available: https://link.springer.com/article/10.1007/s00366-020-01072-1 204

[525] M. Bennani and T. Braconnier, “Stopping criteria for eigensolvers,” *CERFACS, Toulouse, France, Tech. Rep. TR/PA/94/22*, 1994. 205

[526] Q. Li, C. A. Micchelli, L. Shen, al, S. W. Davis, J. M. Stone, Y.-F. Jiang, A. A. Yahyia, A. A. Abbas, and A. S. Jaber, “Acceleration factor optimization of gauss-seidel numerical analysis for power flow using fa,” *Journal of Physics: Conference Series*, vol. 2467, p. 012018, 5 2023. [Online]. Available: https://iopscience.iop.org/article/10.1088/1742-6596/2467/1/012018 205

[527] T. L. Tienand, F. T. Hoeand, and N. Janardan, “Some new findings on gauss–seidel technique for load flow analysis,” M. T. M. S. H. Amylia and Mustaffa, Eds. Springer Singapore, 2014, pp. 529–535. 206

[528] J. S. Arora, “Optimum Design Concepts,” in *Introduction to Optimum Design*, 2nd ed., J. S. Arora, Ed. San Diego: Academic Press, 2004, ch. 4, pp. 83–174. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780120641550500045 207

[529] W. Murray, T. T. De Rubira, and A. Wigington, “Improving the robustness of Newton-based power flow methods to cope with poor initial points,” in *2013 North American Power Symposium (NAPS)*. IEEE, sep 2013, pp. 1–6. [Online]. Available: http://ieeexplore.ieee.org/document/6666905/ 207

[530] T. J. McDougall and S. J. Wotherspoon, “A simple modification of Newton’s method to achieve convergence of order 1+2,” *Applied Mathematics Letters*, vol. 29, pp. 20–25, mar 2014. 207

[531] W. Li, *Appendix C: Power Flow Models*, 2nd ed. John Wiley & Sons, Ltd, 2014, pp. 497–501. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118849972.app3 209, 212

[532] M. A. Salam, *Power System Stability Analysis*. Singapore: Springer Singapore, 2020, ch. 9, pp. 411–460. [Online]. Available: https://doi.org/10.1007/978-981-15-3212-2_9 209

[533] R. K. Portelinha, C. C. Durce, O. L. Tortelli, E. M. Lourenço, and B. C. Pal, "Unified Transmission and Distribution Fast Decoupled Power Flow," *Journal of Control, Automation and Electrical Systems*, vol. 30, no. 6, pp. 1051–1058, 2019. [Online]. Available: https://doi.org/10.1007/s40313-019-00509-3 209

[534] V. Norkin, "The exact projective penalty method for constrained optimization," *Journal of Global Optimization*, pp. 1–18, 1 2024. [Online]. Available: https://link.springer.com/article/10.1007/s10898-023-01350-4 212

[535] A. A. Heidari, S. Mirjalili, H. Faris, I. Aljarah, M. Mafarja, and H. Chen, "Harris hawks optimization: Algorithm and applications," *Future Generation Computer Systems*, vol. 97, pp. 849–872, 8 2019. 213, 341

[536] T. L. Duong, N. A. Nguyen, and T. T. Nguyen, "Application of meta-heuristic algorithm for finding the best solution for the optimal power flow problem," *International Journal of Intelligent Engineering and Systems*, vol. 14, pp. 528–538, 12 2021. [Online]. Available: https://inass.org/wp-content/uploads/2021/10/2021123147.pdf 214

[537] R. D. Zimmerman, C. E. Murillo-Sánchez, and R. J. Thomas, "Matpower: Steady-state operations, planning, and analysis tools for power systems research and education," *IEEE Transactions on Power Systems*, vol. 26, pp. 12–19, 2 2011. 214

[538] M. Mishra, J. Nayak, B. Naik, and A. Abraham, "Deep learning in electrical utility industry: A comprehensive review of a decade of research," *Engineering Applications of Artificial Intelligence*, vol. 96, p. 104000, nov 2020. 217

[539] F. Y. Xu, X. Cun, M. Yan, H. Yuan, Y. Wang, and L. L. Lai, "Power Market Load Forecasting on Neural Network with Beneficial Correlated Regularization," *IEEE Transactions on Industrial Informatics*, vol. 3203, no. c, pp. 1–10, 2018. 217

[540] Y. Xie, Y. Ueda, and M. Sugiyama, "A Two-Stage Short-Term Load Forecasting Method Using Long Short-Term Memory and Multilayer Perceptron," *Energies*, vol. 14, no. 18, p. 5873, sep 2021. [Online]. Available: https://www.mdpi.com/1996-1073/14/18/5873/htm 217

[541] M. El-Hendawi and Z. Wang, "An ensemble method of full wavelet packet transform and neural network for short term electrical load forecasting," *Electric Power Systems Research*, vol. 182, p. 106265, may 2020. 218

[542] L. Ghelardoni, S. Member, A. Ghio, D. Anguita, and S. Member, "Energy Load Forecasting Using Empirical Mode Decomposition and Support Vector Regression," *IEEE Transactions on Smart Grid*, vol. 4, no. 1, pp. 549–556, 2013. 218

[543] H. Yuansheng, H. Shenhai, and S. Jiayin, "A Novel Hybrid Method for Short-Term Power Load Forecasting," *Journal of Electrical and Computer Engineering*, vol. 2016, pp. 1–11, 2016. 218

[544] E. a. Feinberg and D. Genethliou, "Applied Mathematics for Restructured Electric Power Systems," in *IEEE Transactions on Automatic Control*, 2005, vol. 50, no. 9, pp. 269–285. [Online]. Available: http://link.springer.com/chapter/10.1007/0-387-23471-3_12 218

[545] F. D. Rueda, J. D. Suárez, and A. d. R. Torres, "Short-Term Load Forecasting Using Encoder-Decoder WaveNet: Application to the French Grid," *Energies*, vol. 14, no. 9, p. 2524, apr 2021. [Online]. Available: https://www.mdpi.com/1996-1073/14/9/2524/htm 218

[546] A. Acakpovi, A. T. Ternor, N. Y. Asabere, P. Adjei, and A. S. Iddrisu, "Time Series Prediction of Electricity Demand Using Adaptive Neuro-Fuzzy Inference Systems," *Mathematical Problems in Engineering*, vol. 2020, 2020. 218

[547] P. Bento, J. Pombo, M. Calado, and S. Mariano, "Optimization of neural network with wavelet transform and improved data selection using bat algorithm for short-term load forecasting," *Neurocomputing*, vol. 358, pp. 53–71, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925231219306927 218, 219, 221, 227

[548] J. Moon, Y. Kim, M. Son, and E. Hwang, "Hybrid Short-Term Load Forecasting Scheme Using Random Forest and Multilayer Perceptron," *Energies*, vol. 11, no. 12, p. 3283, nov 2018. [Online]. Available: https://www.mdpi.com/1996-1073/11/12/3283/htm 218

[549] Y. K. Semero, J. Zhang, and D. Zheng, "EMD–PSO–ANFIS-based hybrid approach for short-term load forecasting in microgrids," *IET Generation, Transmission & Distribution*, vol. 14, no. 3, pp. 470–475, feb 2020. [Online]. Available: https://onlinelibrary.wiley.com/doi/full/10.1049/iet-gtd.2019.0869 218, 219

[550] E. Ceperic, V. Ceperic, and A. Baric, "A strategy for short-term load forecasting by support vector regression machines," *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 4356–4364, 2013. 218

[551] T.-L. Yuan, D.-S. Jiang, S.-Y. Huang, Y.-Y. Hsu, H.-C. Yeh, M.-N. L. Huang, and C.-N. Lu, "Recurrent Neural Network Based Short-Term Load Forecast with Spline Bases and Real-Time Adaptation," *Applied Sciences*, vol. 11, no. 13, p. 5930, jun

2021. [Online]. Available: https://www.mdpi.com/2076-3417/11/13/5930/htm 218

[552] T. Ciechulski and S. Osowski, "High Precision LSTM Model for Short-Time Load Forecasting in Power Systems," *Energies*, vol. 14, no. 11, p. 2983, may 2021. [Online]. Available: https://www.mdpi.com/1996-1073/14/11/2983/htm 218

[553] W. Wu, W. Liao, J. Miao, and G. Du, "Using Gated Recurrent Unit Network to Forecast Short-Term Load Considering Impact of Electricity Price," *Energy Procedia*, vol. 158, pp. 3369–3374, feb 2019. 218

[554] C.-U. Yeom and K.-C. Kwak, "Short-Term Electricity-Load Forecasting Using a TSK-Based Extreme Learning Machine with Knowledge Representation," *Energies*, vol. 10, no. 10, p. 1613, oct 2017. [Online]. Available: https://www.mdpi.com/1996-1073/10/10/1613/htm 218

[555] Y. Yu, T. Y. Ji, M. S. Li, and Q. H. Wu, "Short-term Load Forecasting Using Deep Belief Network with Empirical Mode Decomposition and Local Predictor," *IEEE Power and Energy Society General Meeting*, vol. 2018-Augus, dec 2018. 218

[556] M. Mansoor, F. Grimaccia, S. Leva, and M. Mussetta, "Comparison of echo state network and feed-forward neural networks in electrical load forecasting for demand response programs," *Mathematics and Computers in Simulation*, vol. 184, pp. 282–293, jun 2021. 218

[557] S. K. Acharya, Y.-M. Wi, and J. Lee, "Short-Term Load Forecasting for a Single Household Based on Convolution Neural Networks Using Data Augmentation," *Energies*, vol. 12, no. 18, p. 3560, sep 2019. [Online]. Available: https://www.mdpi.com/1996-1073/12/18/3560/htm 218

[558] B. N. Oreshkin, G. Dudek, P. Pełka, and E. Turkina, "N-BEATS neural network for mid-term electricity load forecasting," *Applied Energy*, vol. 293, p. 116918, jul 2021. 218

[559] J. Sowinski, "The Impact of the Selection of Exogenous Variables in the ANFIS Model on the Results of the Daily Load Forecast in the Power Company," *Energies*, vol. 14, no. 2, p. 345, jan 2021. [Online]. Available: https://www.mdpi.com/1996-1073/14/2/345/htm 219

[560] Y. Jin, H. Guo, J. Wang, and A. Song, "A Hybrid System Based on LSTM for Short-Term Power Load Forecasting," *Energies*, vol. 13, no. 23, p. 6241, nov 2020. [Online]. Available: https://www.mdpi.com/1996-1073/13/23/6241/htm 219

[561] M. Khashei and Z. Hajirahimi, "A comparative study of series arima/mlp hybrid models for stock price forecasting," *Communications in Statistics - Simulation and Computation*, vol. 48, no. 9, pp. 2625–2640, oct 2018. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/03610918.2018.1458138 219, 221

[562] M. S. Nazar, A. E. Fard, A. Heidari, M. Shafie-khah, and J. P. Catalão, "Hybrid model using three-stage algorithm for simultaneous load and price forecasting," *Electric Power Systems Research*, vol. 165, pp. 214–228, dec 2018. 219

[563] H. Nie, G. Liu, X. Liu, and Y. Wang, "Hybrid of ARIMA and SVMs for short-term load forecasting," in *Energy Procedia*, vol. 16, no. PART C, 2011, pp. 1455–1460. 219

[564] Ü. Ç. Büyükşahin and Ş. Ertekin, "Improving forecasting accuracy of time series data using a new ARIMA-ANN hybrid method and empirical mode decomposition," *Neurocomputing*, vol. 361, pp. 151–163, oct 2019. 219

[565] W. Guo, L. Che, M. Shahidehpour, and X. Wan, "Machine-Learning based methods in short-term load forecasting," *The Electricity Journal*, vol. 34, no. 1, p. 106884, jan 2021. 219

[566] Y. Shen, Y. Ma, S. Deng, C.-J. Huang, and P.-H. Kuo, "An Ensemble Model based on Deep Learning and Data Preprocessing for Short-Term Electrical Load Forecasting," *Sustainability*, vol. 13, no. 4, p. 1694, feb 2021. [Online]. Available: https://www.mdpi.com/2071-1050/13/4/1694/htm 219

[567] M. Massaoudi, S. S. Refaat, I. Chihi, M. Trabelsi, F. S. Oueslati, and H. Abu-Rub, "A novel stacked generalization ensemble-based hybrid LGBM-XGB-MLP model for Short-Term Load Forecasting," *Energy*, vol. 214, p. 118874, jan 2021. 219

[568] T. Salmi, J. Kiljander, and D. Pakkala, "Stacked Boosters Network Architecture for Short-Term Load Forecasting in Buildings," *Energies*, vol. 13, no. 9, p. 2370, may 2020. [Online]. Available: https://www.mdpi.com/1996-1073/13/9/2370/htm 219

[569] C. S. Lai, Z. Mo, T. Wang, H. Yuan, W. W. Ng, and L. L. Lai, "Load forecasting based on deep neural network and historical data augmentation," *IET Generation, Transmission & Distribution*, vol. 14, no. 24, pp. 5927–5934, dec 2020. [Online]. Available: https://onlinelibrary.wiley.com/doi/full/10.1049/iet-gtd.2020.0842 219

[570] R. R. Sharma, M. Kumar, S. Maheshwari, and K. P. Ray, "EVDHM-ARIMA-Based Time Series Forecasting Model and Its Application for COVID-19 Cases," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, 2021. 220

[571] Z. Zhao, C. Wang, M. Nokleby, and C. J. Miller, "Improving short-term electricity price forecasting using day-ahead LMP with ARIMA models," *IEEE Power and Energy Society General Meeting*, vol. 2018-Janua, pp. 1–5, jan 2018. 221

[572] W. Chen, H. Xu, Z. Chen, and M. Jiang, "A novel method for time series prediction based on error decomposition and nonlinear combination of forecasters," *Neurocomputing*, vol. 426, pp. 85–103, feb 2021. 221, 223

[573] R. A. Rajagukguk, R. A. A. Ramadhan, and H.-J. Lee, "A Review on Deep Learning Models for Forecasting Time Series Data of Solar Irradiance and Photovoltaic Power," *Energies*, vol. 13, no. 24, pp. 1–23, 2020. [Online]. Available: https://ideas.repec.org/a/gam/jeners/v13y2020i24p6623-d462571.html 221

[574] A. Shrestha and A. Mahmood, "Review of deep learning algorithms and architectures," *IEEE Access*, vol. 7, pp. 53 040–53 065, 2019. 221

[575] T. Shinozaki and S. Watanabe, "Structure discovery of deep neural network based on evolutionary algorithms," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, apr 2015, pp. 4979–4983. [Online]. Available: http://ieeexplore.ieee.org/document/7178918/ 222

[576] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, "Chapter 10 - Deep learning," in *Data Mining: Practical Machine Learning Tools and Techniques*, I. H. Witten, E. Frank, M. A. Hall, and C. J. B. T. D. M. F. E. Pal, Eds. Morgan Kaufmann, 2017, ch. Deep learn, pp. 417–466. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780128042915000106 222

[577] R. K. Jagait, M. N. Fekri, K. Grolinger, and S. Mir, "Load Forecasting Under Concept Drift: Online Ensemble Learning with Recurrent Neural Network and ARIMA," *IEEE Access*, pp. 1–1, 2021. [Online]. Available: https://ieeexplore.ieee.org/document/9476011/ 223

[578] S. Barak and S. S. Sadegh, "Forecasting energy consumption using ensemble ARIMA–ANFIS hybrid algorithm," *International Journal of Electrical Power & Energy Systems*, vol. 82, pp. 92–104, nov 2016. 223

[579] M. A. Ganaie, M. Hu, M. Tanveer*, and P. N. Suganthan*, "Ensemble deep learning: A review," apr 2021. [Online]. Available: https://arxiv.org/abs/2104.02395v1 223

[580] F. Prado, M. C. Minutolo, and W. Kristjanpoller, "Forecasting based on an ensemble Autoregressive Moving Average - Adaptive neuro - Fuzzy inference system – Neural network - Genetic Algorithm Framework," *Energy*, vol. 197, p. 117159, apr 2020. 223, 225

[581] R. J. Hyndman, A. V. Kostenko, R. Hyndman, and A. V. Kostenko, "Minimum Sample Size requirements for Seasonal Forecasting Models," *Foresight: The International Journal of Applied Forecasting*, no. 6, pp. 12–15, 2007. [Online]. Available: https://econpapers.repec.org/RePEc:for:ijafaa:y:2007:i:6:p:12-15 224

[582] J. Shetty and G. Shobha, "An ensemble of automatic algorithms for forecasting resource utilization in cloud," in *FTC 2016 - Proceedings of Future Technologies Conference*. San Francisco: Institute of Electrical and Electronics Engineers Inc., jan 2017, pp. 301–306. 224

[583] T. Van Giang, V. Debusschere, and S. Bacha, "One week hourly electricity load forecasting using Neuro-Fuzzy and Seasonal ARIMA models," *IFAC Proceedings Volumes*, vol. 45, no. 21, pp. 97–102, jan 2012. 225

[584] M. F. Møller, "A scaled conjugate gradient algorithm for fast supervised learning," *Neural Networks*, vol. 6, no. 4, pp. 525–533, 1993. 227

[585] M. R. Delavar, "Hybrid machine learning approaches for classification and detection of fractures in carbonate reservoir," *Journal of Petroleum Science and Engineering*, vol. 208, p. 109327, jan 2022. 227

[586] I. N. England, "ISO New England - Energy, Load, and Demand Reports," https://www.iso-ne.com/isoexpress/web/reports/load-and-demand, 2021, online; accessed 2021-07-26. 227

[587] P. Pinson, L. Mitridati, C. Ordoudis, and J. Ostergaard, "Towards fully renewable energy systems: Experience and trends in Denmark," *CSEE Journal of Power and Energy Systems*, vol. 3, no. 1, pp. 26–35, mar 20 2017. 232

[588] H. Geman and A. Roncoroni, "Understanding the fine structure of electricity prices," *Journal of Business*, vol. 79, no. 3, pp. 1225–1261, 5 2006. 232

[589] S. Zedda, G. B. Masala, S. Zedda, and G. B. Masala, "Price Spikes in the Italian Electricity Market: An Economic Analysis," *Modern Economy*, vol. 10, no. 9, pp. 2145–2156, sep 11 2019. 232

[590] J. Lago, G. Marcjasz, B. De Schutter, and R. Weron, "Forecasting day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark," *Applied Energy*, vol. 293, p. 116983, jul 1 2021. 233

[591] M. Lehna, F. Scheller, and H. Herwartz, "Forecasting day-ahead electricity prices: A comparison of time series and neural network models taking external regressors into account," *Energy Economics*, vol. 106, p. 105742, feb 1 2022. 233

[592] G. Memarzadeh and F. Keynia, "Short-term electricity load and price forecasting by a new optimal LSTM-NN based prediction algorithm," *Electric Power Systems Research*, vol. 192, p. 106995, mar 1 2021. 233

[593] F. Zhang, H. Fleyeh, and C. Bales, "A hybrid model based on bidirectional long short-term memory neural network and Catboost for short-term electricity spot price forecasting," *Journal of the Operational Research Society*, vol. 73, 2022. 233

[594] C. J. Huang, Y. Shen, Y. H. Chen, and H. C. Chen, "A novel hybrid deep neural network model for short-term electricity price forecasting," *International Journal of Energy Research*, vol. 45, no. 2, pp. 2511–2532, feb 1 2021. 234

[595] W. Yang, S. Sun, Y. Hao, and S. Wang, "A novel machine learning-based electricity price forecasting model based on optimal model selection strategy," *Energy*, vol. 238, p. 121989, jan 1 2022. 234

[596] J. Zhang, Z. Tan, and Y. Wei, "An adaptive hybrid model for short term electricity price forecasting," *Applied Energy*, vol. 258, p. 114087, jan 15 2020. 234, 238

[597] H. Zhang, Y. Yang, Y. Zhang, Z. He, W. Yuan, Y. Yang, W. Qiu, and L. Li, "A combined model based on SSA, neural networks, and LSSVM for short-term electric load and price forecasting," *Neural Computing and Applications*, vol. 33, no. 2, pp. 773–788, jan 1 2021. 234

[598] OMIE, "Day-ahead hourly price," 2022, [Online; accessed 2022-03-30]. [Online]. Available: https://www.omie.es/en/market-results/daily/daily-market/daily-hourly-price 235

[599] J. F. F. Januário, "Electricity price forecasting utilizing machine learning in MIBEL," mathesis, Instituto Universitário de Lisboa, Nov. 2019. [Online]. Available: https://repositorio.iscte-iul.pt/handle/10071/20235 236

[600] A. Wagner, E. Ramentol, F. Schirra, and H. Michaeli, "Short- and long-term forecasting of electricity prices using embedding of calendar information in neural networks," *Journal of Commodity Markets*, p. 100246, feb 20 2022. 236

[601] J. P. S. Catalão, S. J. P. S. Mariano, V. M. F. Mendes, and L. A. F. M. Ferreira, "Short-term electricity prices forecasting in a competitive market: A neural network approach," *Electric Power Systems Research*, vol. 77, no. 10, pp. 1297–1304, 2007, 1. 236

[602] A. Heydari, M. Majidi Nezhad, E. Pirshayan, D. Astiaso Garcia, F. Keynia, and L. De Santoli, "Short-term electricity price and load forecasting in isolated power grids based on composite neural network and gravitational search optimization algorithm," *Applied Energy*, vol. 277, p. 115503, nov 1 2020. 237

[603] P. O. Oviroh and T. C. Jen, "The energy cost analysis of hybrid systems and diesel generators in powering selected base transceiver station locations in Nigeria," *Energies*, vol. 11, no. 3, pp. 7–9, 2018. 243

[604] S. Upadhyay and M. P. Sharma, "A review on configurations, control and sizing methodologies of hybrid energy systems," *Renewable and Sustainable Energy Reviews*, vol. 38, pp. 47–63, 2014. 243, 362

[605] K. Shivarama Krishna and K. Sathish Kumar, "A review on hybrid renewable energy systems," *Renewable and Sustainable Energy Reviews*, vol. 52, pp. 907–916, 2015. 243, 244, 364

[606] S. Nojavan, M. Majidi, and K. Zare, "Performance improvement of a battery/PV/-fuel cell/grid hybrid energy system considering load uncertainty modeling using IGDT," *Energy Conversion and Management*, vol. 147, pp. 29–39, 2017. 243

[607] A. Arsalis and G. E. Georghiou, "A Decentralized , Hybrid Photovoltaic-Solid Oxide Fuel Cell System for Application to a Commercial Building," *Energies*, vol. 11, no. 12, p. 3512, 2018. 243, 244

[608] S. S. Reddy, "Optimal scheduling of thermal-wind-solar power system with storage," *Renewable Energy*, vol. 101, pp. 1357–1368, 2017. 243

[609] A. Malheiro, P. M. Castro, R. M. Lima, and A. Estanqueiro, "Integrated sizing and scheduling of wind/PV/diesel/battery isolated systems," *Renewable Energy*, vol. 83, pp. 646–657, 2015. 243, 362

[610] V. M. Sanchez, A. U. Chavez-Ramirez, S. M. Duron-Torres, J. Hernandez, L. G. Arriaga, and J. M. Ramirez, "Techno-economical optimization based on swarm intelligence algorithm for a stand-alone wind-photovoltaic-hydrogen power system at south-east region of Mexico," *International Journal of Hydrogen Energy*, vol. 39, no. 29, pp. 16 646–16 655, 2014. 243

[611] F. Fodhil, A. Hamidat, and O. Nadjemi, "Potential, optimization and sensitivity analysis of photovoltaic-diesel-battery hybrid energy system for rural electrification in Algeria," *Energy*, vol. 169, pp. 613–624, 2019. 243

[612] V. Khare, S. Nema, and P. Baredar, "Solar-wind hybrid renewable energy system: A review," *Renewable and Sustainable Energy Reviews*, vol. 58, pp. 23–33, 2016. 244

[613] W. Dong, Y. Li, and J. Xiang, "Optimal sizing of a stand-alone hybrid power system based on battery/hydrogen with an improved ant colony optimization," *Energies*, vol. 9, no. 10, 2016. 244

[614] H. Chitsaz, P. Zamani-Dehkordi, H. Zareipour, and P. P. Parikh, "Electricity price forecasting for operational scheduling of behind-the-meter storage systems," *IEEE Transactions on Smart Grid*, vol. 9, no. 6, pp. 6612–6622, 2018. 246

[615] Neon, Technical University of Berlin, ETH Zürich, and DIW Berlin, "Open Power System Data," https://open-power-system-data.org/, 2018, [Online; accessed 2018-11-15]. [Online]. Available: https://open-power-system-data.org/ 246

[616] Statista, "Prices of diesel fuel in Czechia 2004-2018," https://www.statista.com/statistics/603765/diesel-fuel-prices-czech-republic/, 2018, [Online; accessed 2018-11-19]. [Online]. Available: https://www.statista.com/statistics/603765/diesel-fuel-prices-czech-republic/ 247

[617] T. Hong, P. Pinson, Y. Wang, R. Weron, D. Yang, and H. Zareipour, "Energy Forecasting: A Review and Outlook," *IEEE Open Access Journal of Power and Energy*, vol. 7, pp. 376–388, oct 9 2020. 250, 251

[618] A. Jahanshahi, M. Kamali, M. Khalaj, and Z. Khodaparast, "Delphi-based prioritization of economic criteria for development of wave and tidal energy technologies," *Energy*, vol. 167, pp. 819–827, jan 15 2019. 251

[619] R. Mendes, M. Calado, and S. Mariano, "Wave energy potential in Portugal–Assessment based on probabilistic description of ocean waves parameters," *Renewable Energy*, vol. 47, pp. 1–8, nov 1 2012. 251

[620] European Union, "Ocean Energy: Promising new technologies to help Europe achieve its ambitious climate goals | H2020 | Results Pack | CORDIS | European Commission," https://cordis.europa.eu/article/id/422014-ocean-energy-promising-new-technologies-to-help-europe-achieve-its-ambitious-climate-goals, 2020, [Online; accessed 2020-12-07]. [Online]. Available: https://cordis.europa.eu/article/id/422014-ocean-energy-promising-new-technologies-to-help-europe-achieve-its-ambitious-climate-goals 251

[621] A. Muzathik, W. Nik, M. Ibrahim, and K. Samo, "Wave energy potential of peninsular Malaysia," *ARPN Journal of Engineering and Applied Sciences*, vol. 5, no. 7, pp. 11–23, 7 2010. 251

[622] IRENA (2020), "Innovation outlook: Ocean energy technologies," Abu Dhabi, Tech. Rep., 2020. 251

[623] M. Melikoglu, "Current status and future of ocean energy sources: A global review," *Ocean Engineering*, vol. 148, pp. 563–573, jan 15 2018. 251

[624] B. Johnson and E. Cotilla-Sanchez, "Estimating the impact of ocean wave energy on power system reliability with a well-being approach," *IET Renewable Power Generation*, vol. 14, no. 4, pp. 608–615, mar 16 2020. 251

[625] D. Magagna and A. Uihlein, "Ocean energy development in Europe: Current status and future perspectives," *International Journal of Marine Energy*, vol. 11, pp. 84–104, sep 1 2015. 251

[626] J. Widén, N. Carpman, V. Castellucci, D. Lingfors, J. Olauson, F. Remouit, M. Bergkvist, M. Grabbe, and R. Waters, "Variability assessment and forecasting of renewables: A review for solar, wind, wave and tidal resources," *Renewable and Sustainable Energy Reviews*, vol. 44, pp. 356–375, apr 1 2015. 251

[627] S. C. James, Y. Zhang, and F. O'Donncha, "A machine learning framework to forecast wave conditions," *Coastal Engineering*, vol. 137, pp. 1–10, jul 1 2018. 251, 252

[628] C. Sweeney, R. J. Bessa, J. Browell, and P. Pinson, "The future of forecasting for renewable energy," *WIREs Energy and Environment*, vol. 9, no. 2, mar 9 2020. 251

[629] F. O'Donncha, Y. Zhang, B. Chen, and S. C. James, "Ensemble model aggregation using a computationally lightweight machine-learning model to forecast ocean waves," *Journal of Marine Systems*, vol. 199, p. 103206, nov 1 2019. 251

[630] D. Avila, G. N. Marichal, I. Padrón, R. Quiza, and Á. Hernández, "Forecasting of wave energy in Canary Islands based on Artificial Intelligence," *Applied Ocean Research*, vol. 101, p. 102189, aug 1 2020. 251, 252

[631] S. M. Mousavi, M. Ghasemi, M. D. Manshadi, and A. Mosavi, "Deep learning for wave energy converter modeling using long short-term memory," *Mathematics*, vol. 9, no. 8, p. 871, apr 2 2021. 251

[632] X. Feng, G. Ma, S. F. Su, C. Huang, M. K. Boswell, and P. Xue, "A multi-layer perceptron approach for accelerated wave forecasting in Lake Michigan," *Ocean Engineering*, vol. 211, p. 107526, sep 1 2020. 251, 252

[633] M. A. Desouky and O. Abdelkhalik, "Wave prediction using wave rider position measurements and NARX network in wave energy conversion," *Applied Ocean Research*, vol. 82, pp. 10–21, jan 1 2019. 252

[634] R. Bell and B. Kirtman, "Seasonal forecasting of winds, waves and currents in the North Pacific," *Journal of Operational Oceanography*, vol. 11, no. 1, pp. 11–26, jan 2 2018. 252

[635] C. W. Zheng, C. Y. Li, X. Chen, and J. Pan, "Numerical Forecasting Experiment of the Wave Energy Resource in the China Sea," *Advances in Meteorology*, vol. 2016, pp. 1–12, aug 15 2016. 252

[636] J. Jeon and J. W. Taylor, "Short-term density forecasting of wave energy using ARMA-GARCH models and kernel density estimation," *International Journal of Forecasting*, vol. 32, no. 3, pp. 991–1004, jul 1 2016. 252

[637] H. Wang, Z. Lei, X. Zhang, B. Zhou, and J. Peng, "A review of deep learning for renewable energy forecasting," *Energy Conversion and Management*, vol. 198, p. 111799, oct 15 2019. 252

[638] A. Mosavi, M. Salimi, S. F. Ardabili, T. Rabczuk, S. Shamshirband, and A. R. Varkonyi-Koczy, "State of the art of machine learning models in energy systems, a systematic review," *Energies*, vol. 12, no. 7, p. 1301, apr 4 2019. 252

[639] R. Lou, Z. Lv, S. Dang, T. Su, and X. Li, "Application of machine learning in ocean data," in *Multimedia Systems*, vol. 1. Springer Science and Business Media Deutschland GmbH, feb 14 2021, p. 3. 252

[640] J. Berbić, E. Ocvirk, D. Carević, and G. Lončar, "Application of neural networks and support vector machine for significant wave height prediction," *Oceanologia*, vol. 59, no. 3, pp. 331–349, jul 1 2017. 252

[641] T. Sadeghifar, M. Nouri Motlagh, M. Torabi Azad, and M. Mohammad Mahdizadeh, "Coastal Wave Height Prediction using Recurrent Neural Networks (RNNs) in the South Caspian Sea," *Marine Geodesy*, vol. 40, no. 6, pp. 454–465, nov 2 2017. 252

[642] P. Serras, G. Ibarra-Berastegi, J. Sáenz, and A. Ulazia, "Combining random forests and physics-based models to forecast the electricity generated by ocean waves: A case study of the Mutriku wave farm," *Ocean Engineering*, vol. 189, p. 106314, oct 1 2019. 252

[643] N. K. Kumar, R. Savitha, and A. Al Mamun, "Ocean wave height prediction using ensemble of Extreme Learning Machine," *Neurocomputing*, vol. 277, pp. 12–20, feb 14 2018. 252

[644] L. Cornejo-Bueno, P. Rodríguez-Mier, M. Mucientes, J. Nieto-Borge, and S. Salcedo-Sanz, "Significant wave height and energy flux estimation with a Genetic Fuzzy System for regression," *Ocean Engineering*, vol. 160, pp. 33–44, jul 15 2018. 252

[645] L. Cornejo-Bueno, J. Nieto-Borge, P. García-Díaz, G. Rodríguez, and S. Salcedo-Sanz, "Significant wave height and energy flux prediction for marine energy applications: A grouping genetic algorithm – Extreme Learning Machine approach," *Renewable Energy*, vol. 97, pp. 380–389, nov 1 2016. 252

[646] J. Oh and K.-D. Suh, "Real-time forecasting of wave heights using EOF – wavelet – neural network hybrid model," *Ocean Engineering*, vol. 150, pp. 48–59, feb 15 2018. 252

[647] M. Pirhooshyaran and L. Snyder, "Forecasting, hindcasting and feature selection of ocean waves via recurrent and sequence-to-sequence networks," *Ocean Engineering*, vol. 207, p. 107424, jul 1 2020. 252

[648] R. Lou, W. Wang, X. Li, Y. Zheng, and Z. Lv, "Prediction of Ocean Wave Height Suitable for Ship Autopilot," *IEEE Transactions on Intelligent Transportation Systems*, 2021. 252

[649] P. Izquierdo, C. Guedes Soares, J. C. Nieto Borge, and G. R. Rodríguez, "A comparison of sea-state parameters from nautical radar images and buoy data," *Ocean Engineering*, vol. 31, no. 17-18, pp. 2209–2225, dec 1 2004. 253

[650] J. Pastor and Y. Liu, "Wave Climate Resource Analysis Based on a Revised Gamma Spectrum for Wave Energy Conversion Technology," *Sustainability*, vol. 8, no. 12, p. 1321, dec 14 2016. 253

[651] G. Reikard, P. Pinson, and J.-R. Bidlot, "Forecasting ocean wave energy: The ECMWF wave model and time series methods," *Ocean Engineering*, vol. 38, no. 10, pp. 1089–1099, jul 1 2011. 254, 260

[652] S. Nosratabadi, A. Mosavi, P. Duan, P. Ghamisi, F. Filip, S. S. Band, U. Reuter, J. Gama, and A. H. Gandomi, "Data science in economics: Comprehensive review of advanced machine learning and deep learning methods," *Mathematics*, vol. 8, no. 10, pp. 1–25, oct 1 2020. 255

[653] F. Shaheen, B. Verma, and M. Asafuddoula, "Impact of Automatic Feature Extraction in Deep Learning Architecture," in *2016 International Conference on Digital Image Computing: Techniques and Applications, DICTA 2016*. Institute of Electrical and Electronics Engineers Inc., dec 22 2016. 255

[654] C. Wang, J. Yang, L. Xie, and J. Yuan, "Kervolutional neural networks," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2019-June. IEEE Computer Society, jun 1 2019, pp. 31–40. 255

[655] Z. Luo, X. Cai, K. Tanaka, T. Takiguchi, T. Kinkyo, and S. Hamori, "Can We Forecast Daily Oil Futures Prices? Experimental Evidence from Convolutional Neural Networks," *Journal of Risk and Financial Management*, vol. 12, no. 1, p. 9, jan 8 2019. 255

[656] A. Mosavi, Y. Faghan, P. Ghamisi, P. Duan, S. F. Ardabili, E. Salwana, and S. S. Band, "Comprehensive review of deep reinforcement learning methods and applications in economics," *Mathematics*, vol. 8, no. 10, p. 1640, oct 1 2020. 256

[657] I. Mitiche, A. Nesbitt, S. Conner, P. Boreham, and G. Morison, "1d-CNN based real-time fault detection system for power asset diagnostics," *IET Generation, Transmission and Distribution*, vol. 14, no. 24, pp. 5816–5822, dec 18 2020. 257

[658] F. Tian, J. Wang, and F. Chu, "Improved multi-strategy harris hawks optimization and its application in engineering problems," *Mathematics*, vol. 11, no. 6, 2023. [Online]. Available: https://www.mdpi.com/2227-7390/11/6/1525 343, 344

[659] H. Lu, X. Wang, Z. Fei, and M. Qiu, "The effects of using chaotic map on improving the performance of multiobjective evolutionary algorithms," *Mathematical Problems in Engineering*, vol. 2014, 2014. 343

[660] H. Nunes, P. Silva, J. Pombo, S. Mariano, and M. Calado, "Multiswarm spiral leader particle swarm optimisation algorithm for pv parameter identification," *Energy Conversion and Management*, vol. 225, p. 113388, 12 2020. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0196890420309249 344

[661] H. G. Nunes, J. P. Portugal, J. A. Pombo, S. J. Mariano, and M. R. Calado, "Parameter Estimation of Per-Unit Photovoltaic Models Using Optimization

Algorithms: Comparative Study," *Studies in Systems, Decision and Control*, vol. 212, pp. 157–195, 2022. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-031-07512-4_6 345

[662] M. Georgioudakis and V. Plevris, "A comparative study of differential evolution variants in constrained structural optimization," *Frontiers in Built Environment*, vol. 6, p. 552667, 7 2020. 345

[663] O. Yurur, C. H. Liu, and W. Moreno, "Modeling battery behavior on sensory operations for context-aware smartphone sensing," *Sensors (Switzerland)*, vol. 15, no. 6, pp. 12 323–12 341, 2015. 358

[664] S. Papaefthymiou, E. G. Karamanou, S. a. Papathanassiou, and M. P. Papadopoulos, "A wind-hydro-pumped storage station leading to high RES penetration in the autonomous island system of Ikaria," *IEEE Transactions on Sustainable Energy*, vol. 1, no. 3, pp. 163–172, 2010. 359

[665] T. Ma, H. Yang, L. Lu, and J. Peng, "Pumped storage-based standalone photovoltaic power generation system: Modeling and techno-economic optimization," *Applied Energy*, vol. 137, pp. 649–659, 2014. 359

[666] J. Razak, Y. Ali, and M. Alghoul, "Application of crossflow turbine in off-grid pico hydro renewable energy system," *Recent Advances in Applied Mathematics*, pp. 519–526, 2010. 360

[667] International Renewable Energy Agency (IRENA), "Hydropower," Abu Dhabi, Tech. Rep., 2012. 360

[668] C. S. Kaunda, C. Z. Kimambo, and T. K. Nielsen, "Potential of Small-Scale Hydropower for Electricity Generation in Sub-Saharan Africa," *ISRN Renewable Energy*, vol. 2012, pp. 1–15, 2012. 361

[669] M. Bencherif, B. N. Brahmi, and A. Chikhaoui, "Optimum Selection of Wind Turbines," *Science Journal of Energy Engineering*, vol. 2, no. 4, p. 36, 2014. 361

[670] S. Mandal, B. K. Das, and N. Hoque, "Optimum sizing of a stand-alone hybrid energy system for rural electrification in Bangladesh," *Journal of Cleaner Production*, vol. 200, pp. 12–27, 2018. 361, 364

[671] M. Potli, A. Mallikarjuna, J. C. Balachandra, and V. N, "Optimal Sizing of Wind/-Solar/Hydro in an Isolated Power System using SMFFT based Cuckoo Search Algorithm," in *2016 IEEE 6th International Conference on Power Systems (ICPS)*. New Delhi: IEEE, 2016. 361

[672] Y. Riffonneau, S. Bacha, F. Barruel, and S. Ploix, "Optimal Power Flow Management for Grid Connected PV Systems With Batteries," *IEEE Transactions on Sustainable Energy*, vol. 2, no. 3, pp. 309–320, 2011. 362, 365

[673] K. Saurav, H. Bansal, M. Nawhal, V. Chandan, and V. Arya, “Minimizing energy costs of commercial buildings in developing countries,” *2016 IEEE International Conference on Smart Grid Communications, SmartGridComm 2016*, pp. 637–642, 2016. 363

[674] OECD, *Taxing Energy Use 2018: Companion to the Taxing Energy Use Database*. Paris: OECD Publishing, feb 14 2018. 364

Appendix A

SASMA a novel surrogate assisted metaheuristic

Having described the holistic approach and key aspects behind the proposed SASMA in sub-section 4.5.3, with a focus on the novelties regarding the DB management strategy and the surrogate building approach, it is now useful to introduce a more in-depth account of the flow between the main mechanisms illustrated in Figure 4.11. Namely, by comprising all the involved variables and the individual stages needed to adapt the original SMA to its SAEA form in pseudo code (as shown in Algorithm 1).

Algorithm 1 Pseudo Code of the Surrogate Assisted Slime Mould Algorithm (SASMA) and the employed parameters

Input: $nPop = 30$; $nDims = 30$ or $nDims = 100$; $maxFEs$, f, lb, ub are test function/problem dependent; $iter = FEs = 0$; $z = 0.03$ (SMA parameter); $lengthDB = 1000$; $\phi_{\min} = 0.35$ and $\phi_{\max} = 0.95$; $\alpha_{\max} = 0.305$ and $\gamma = 1.5 \times 10^{-3}$; $lenghtXtrain_{lb} = \min(\text{length}(Xtrain^{space})) = nPop$ and $lenghtXtrain_{ub} = \max(\text{length}(Xtrain^{space})) = 5 \times nPop$; $\varepsilon_{\min} = 0.01$ and $\varepsilon_{\max} = 0.1$

Initialize (using LHS) the population, X, where the search agents or positions are denoted as $X_i,\ i \in [1, nPop]$; $X^{temp} = X$

Evaluate the real fitness of the initialized agents $S_i = f(X_i)$; $X_{best} | \ best = \arg \min_{i \in [1,nPop]} (S_i)$; $SmellIndex = \text{Sort}(S) \ | \ SmellIndex(i) \leq SmellIndex(i+1),\ i \in [1, nPop]$; $DF = bF = \min_{i \in [1,nPop]} (S_i)$; $wF = \max_{i \in [1,nPop]} (S_i)$

Store this information into the DB $\left(X^{DB}, fitness^{DB}, age^{DB}\right)$, according to the flowchart in Figure 4.12

Assemble the training set, X^{train} using Equations 4.18-4.21; Compute the spread, σ, using Equation 4.22; Compute the error goal, ε_{RBF}, using Equation 4.23; Build the cubic RBF surrogate (approximation) model, $\hat{f}$

while $(FEs < maxFEs)$ **do**

  if $(iter > 0)$ **then**

    $SmellIndex = \text{Sort}(S) \ | \ SmellIndex(i) \leq SmellIndex(i+1),\ i \in [1, nPop]$

    $bF = \min_{i \in [1,nPop]} (S_i) = SmellIndex(1)$; $wF = \max_{i \in [1,nPop]} (S_i) = SmellIndex(nPop)$

  end if

  if $(SmellIndex(1) < DF)$ **then**

    $DF = SmellIndex(1)$

  end if

  for $i = 1 : nPop$ **do**

    Compute the weight of each individual slime mould, W_i, using Equation 4.13

  end for

  Compute the auxiliary SMA variables: a, b

for $i = 1 : nPop$ **do**
  Calculate the variables p, v_b, v_c
  Update each swarm agent (position) and store it in a temporary variable, X_i^{temp}, according to Equation 4.12
  Bound checking of the updated position X_i^{temp}, using Equation 4.14
end for
$realFE_{counter} = 0$
for $i = 1 : nPop$ **do**
  Compute the surrogate value for each updated position, $\hat{f}\left(X_i^{temp}\right)$
  if $\left(\hat{f}\left(X_i^{temp}\right) < S_i\right)$ **then**
    The updated position is deemed as "promising", so it is accepted, $X_i = X_i^{temp}$
    The real fitness value is therefore computed, $S_i = f\left(X_i\right)$
    $realFE_{counter} = realFE_{counter} + 1$
  end if
end for
if $(realFE_{counter} == 0)$ **then** ▷ none of the updated positions were deemed as "promising"
  for $i = 1 : nPop$ **do**
    Accept the updated position regardless, $X_i = X_i^{temp}$ ▷ as in the original SMA code
    Compute its real fitness and store the value, $S_i = f\left(X_i\right)$
    $realFE_{counter} = realFE_{counter} + 1$
  end for
end if
for $i = 1 : nPop$ **do**
  Compute the variables $S\left(X_i\right), D\left(X_i\right), \phi$, according to Equations 4.15 and 4.16 and the inline ϕ equation in sub-section 4.5.3.1
  Compute the dual based merit metric, $f_{\text{merit}}\left(X_i\right)$, Equation 4.17
end for
Update the DB, i.e., $\left(X^{DB}, fitness^{DB}, age^{DB}\right)$, based on the f_{merit} and S, according to several stages depicted in the flowchart in Figure 4.12
Assemble the new training set, X^{train} using Equations 4.18-4.21; Update the spread, σ, and the error goal, ε_{RBF}, using Equation 4.22 and Equation 4.23, respectively; Update the cubic RBF surrogate (approximation) model, $\hat{f}$
$FEs = FEs + realFE_{counter}$
$iter = iter + 1$
end while
Output: X_{best}; $f\left(X_{best}\right)$

A.1 Case Study: an in-depth look at mathematical test functions

As mentioned in section 4.5.4.1, the mathematical test functions used as the Case Study I, i.e., its expression (formula), name, optimization range and global minimum are shown in Table A.1.

Table A.1: Case Study I: Benchmark test functions.

Function	Name	Range (n)	f_{min}
$F_1(x) = \sum_{i=1}^{n} x_i^2$	Sphere	$[-100, 100]$	0
$F_2(x) = \sum_{i=1}^{n} \lvert x_i \rvert + \prod_{i=1}^{n} \lvert x_i \rvert$	Schwefel 2.22	$[-10, 10]$	
$F_3(x) = \sum_{i=1}^{n} \left(\sum_{j=1}^{i} x_i \right)^2$	Schwefel 1.2	$[-100, 100]$	
$F_4(x) = \max_{i=1,\ldots,n} \lvert x_i \rvert$	Schwefel 2.21	$[-100, 100]$	
$F_5(x) = \sum_{i=1}^{n-1} \left[100 \left(x_{i+1} - x_i^2 \right)^2 + \left(x_i - 1 \right)^2 \right]$	Rosenbrock	$[-30, 30]$	
$F_6(x) = \sum_{i=1}^{n} \lvert x_i + 0.5 \rvert^2$	Step	$[-100, 100]$	
$F_7(x) = \sum_{i=1}^{n} i x_i^4 + \alpha_i, \;\; \alpha_i = rand([0, 1])$	Quartic	$[-1.28, 1.28]$	
$F_8(x) = \sum_{i=1}^{n} -x_i \sin\left(\sqrt{\lvert x_i \rvert} \right)$	Schwefel 2.26	$[-500, 500]$	$-418.9829 \times nDims$
$F_9(x) = \sum_{i=1}^{n} \left[10 + x_i^2 - 10\cos(2\pi x_i) \right]$	Rastrigin	$[-5.12, 5.12]$	0
$F_{10}(x) = -20e^{-0.2\sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}} - e^{\frac{1}{n}\sum_{i=1}^{n}\cos(2\pi x_i)} + 20 + e$	Ackley	$[-32, 32]$	
$F_{11}(x) = 1 + \sum_{i=1}^{n} \frac{x_i}{4000} - \prod_{i=1}^{n} \cos\left(\frac{x_i}{\sqrt{i}} \right)$	Griewank	$[-600, 600]$	
$F_{12}(x) = \frac{\pi}{n} \{ 10\sin(2\pi y_1) + \sum_{i=1}^{n-1} \left(\begin{array}{c} (y_i - 1)^2 \left[1 + 10\sin^2(\pi y_{i+1}) \right] \\ + (y_n - 1)^2 \end{array} \right) \} + \sum_{i=1}^{n} u(x_i, 10, 100, 4)$ $y_i = 1 + \frac{x_i + 1}{4}$ $u(x_i, a, k, m) = \begin{cases} k(x_i - a)^m, x_i > a \\ 0, -a \le x_i \le a \\ -k(x_i + a)^m, x_i < -a \end{cases}$	Penalized 1	$[-50, 50]$	
$F_{13}(x) = 0.1 \{ \sin^2(3\pi x_1) + \sum_{i=1}^{n} \left(\begin{array}{c} (x_i - 1)^2 \left[1 + \sin^2(3\pi x_i + 1) \right] \\ + (x_n - 1)^2 \left[1 + \sin^2(2\pi x_n) \right] \end{array} \right) \} + \sum_{i=1}^{n} u(x_i, 5, 100, 4)$	Penalized 2	$[-50, 50]$	
$F_{14}(x) = \sum_{i=1}^{n} i x_i^2$	Ellipsoid	$[-100, 100]$	
$F_{15}(x) = \sum_{i=1}^{n} \left[10 + z_i^2 - 10\cos(2\pi z_i) \right] + f_{bias}$ $z = (x - o) * M$ $f_{bias} = -330$	Shifted Rotated Rastrigin (F10 in [379])	$[-5, 5]$	-330
$F_{16}(CF) =$ [f_1, f_2 = Ratrigin's Function f_3, f_4 = Weierstrass Function f_5, f_6 = Griewank's Function f_7, f_8 = Ackley's Function f_9, f_{10} = Sphere Function] $f_{bias} = 120$ $\sigma_i = 1, i = 1, ..., n$ $\lambda = [1, 1, 10, 10, 5/60, 5/60, 5/32, 5/32, 5/100, 5/100]$ M_i are all identity matrices	Rotated Hybrid Composition (F16 in [379])	$[-5, 5]$	120
$F_{17}(CF) =$ [f_1, f_2 = Ackley's Function f_3, f_4 = Ratrigin's Function f_5, f_6 = Sphere Function f_7, f_8 = Weierstrass Function f_9, f_{10} = Griewank's Function] $f_{bias} = 10$ $\sigma_i = [1, 2, 1.5, 1.5, 1, 1, 1.5, 1.5, 2, 2]$ $\lambda = [10/32, 5/32, 2, 1, 10/100, 5/100, 20, 10, 10/60, 5/60]$ M_i are all rotation matrices.	Rotated Hybrid Composition (F19 in [379])	$[-5, 5]$	10

A.2 Optimizations Results for 100D

The descriptive statistics and the results of the Friedman tests and non-parametric Wilcoxon for the 100-dimensional case are presented in Tables A.2 and A.3. The corresponding convergence curves are shown in Figure A.1 for the unimodal test functions, Figure A.2 for the multimodal test functions, and Figure A.3 for the highly complex multimodal test functions.

Table A.2: Mean rank scores (based on the Friedman Test) of the SASMA and the benchmark SAEAs on 100D mathematical test functions.

Friedman Test		GORS-SSLPSO	SAE-ARFS	SHPSO	TL-SSLPSO	TLSAPSO	SASMA
Mean Rank	F1	2.31	5.23	2.77	3.91	5.77	1.00
	F2	2.11	3.71	4.00	4.17	6.00	1.00
	F3	2.26	5.49	3.46	3.40	5.40	1.00
	F4	3.54	4.80	2.00	3.71	5.94	1.00
	F5	3.00	5.00	2.00	4.00	6.00	1.00
	F6	1.43	5.14	1.91	3.89	5.86	2.77
	F7	2.86	5.00	2.14	4.00	6.00	1.00
	F8	2.17	3.57	5.77	3.26	5.23	1.00
	F9	2.00	5.00	3.94	3.06	6.00	1.00
	F10	4.06	4.94	2.00	3.00	6.00	1.00
	F11	2.37	6.00	2.86	4.03	4.74	1.00
	F12	2.71	5.00	2.29	4.00	6.00	1.00
	F13	2.91	5.00	2.09	4.00	6.00	1.00
	F14	2.06	5.09	2.94	4.00	5.91	1.00
	F15	2.40	5.20	1.06	2.54	5.31	4.49
	F16	3.09	4.00	1.06	2.00	5.34	5.51
	F17	4.63	4.74	2.80	2.69	5.14	1.00

[1] According to the Friedman test the best mean rank (of the 35 runs) for each test function is highlighted in grey.

,

Table A.3: Statistical results of the proposed SASMA and the benchmark SAEAs on 100D mathematical test functions.

Test Function/Metric		GORS-SSLPSO		SAE-ARFS		SHPSO		TL-SSLPSO		TLSAPSO		SASMA
F1	Mean	1.429E+03	+	1.065E+05	+	6.281E-01	+	7.452E+03	+	1.243E+05	+	9.287E-29
	St. Dev.	4.299E+03		1.121E+04		1.953E-01		3.197E+03		1.531E+04		2.344E-28
	Min	1.069E-04		7.513E+04		2.863E-01		2.312E+03		9.326E+04		3.395E-39
F2	Mean	6.761E+31	+	2.640E+09	+	1.826E+15	+	3.132E+12	+	2.640E+38	+	6.749E-15
	St. Dev.	4.000E+32		1.449E+10		8.767E+15		1.799E+13		7.400E+38		1.480E-14
	Min	1.052E+02		2.908E+02		1.578E+02		2.439E+02		3.492E+23		1.595E-20
F3	Mean	1.442E+05	+	4.064E+05	+	2.504E+05	+	2.285E+05	+	3.963E+05	+	9.798E-16
	St. Dev.	7.018E+04		6.533E+04		5.408E+04		5.307E+04		7.280E+04		4.145E-15
	Min	4.291E+04		2.742E+05		1.417E+05		1.212E+05		2.500E+05		2.371E-33
F4	Mean	7.814E+01	+	8.319E+01	+	4.313E+01	+	7.757E+01	+	9.266E+01	+	7.664E-14
	St. Dev.	3.531E+00		2.529E+00		5.503E+00		4.761E+00		3.424E+00		1.466E-13
	Min	6.999E+01		7.760E+01		3.173E+01		6.773E+01		8.292E+01		1.066E-16
F5	Mean	2.020E+05	+	2.977E+08	+	4.839E+04	+	2.669E+06	+	6.314E+08	+	9.651E+01
	St. Dev.	3.715E+04		6.150E+07		3.117E+04		2.023E+06		8.800E+07		1.046E+01
	Min	1.254E+05		1.881E+08		1.328E+04		6.084E+05		4.451E+08		4.343E+01
F6	Mean	1.426E+03	+	1.084E+05	+	6.607E-01	-	7.415E+03	+	1.233E+05	+	1.750E+01
	St. Dev.	3.544E+03		1.120E+04		3.117E-01		2.875E+03		1.226E+04		6.667E+00
	Min	8.937E-05		8.731E+04		2.933E-01		3.330E+03		8.511E+04		9.927E-02
F7	Mean	1.826E+00	+	3.976E+02	+	1.133E+00	+	1.070E+01	+	1.070E+03	+	2.193E-03
	St. Dev.	3.753E-01		6.799E+01		4.985E-01		7.321E+00		1.514E+02		1.736E-03
	Min	1.146E+00		2.651E+02		5.251E-01		3.419E+00		7.234E+02		6.724E-05
F8	Mean	1.946E+04	+	2.264E+04	+	3.159E+04	+	2.211E+04	+	3.020E+04	+	3.056E+03
	St. Dev.	1.683E+03		1.095E+03		1.512E+03		2.012E+03		1.420E+03		4.036E+03
	Min	1.567E+04		2.035E+04		2.854E+04		1.865E+04		2.790E+04		1.903E-01
F9	Mean	3.447E+02	+	1.042E+03	+	7.663E+02	+	5.165E+02	+	1.423E+03	+	4.582E-06
	St. Dev.	6.295E+01		5.005E+01		1.009E+02		6.831E+01		6.909E+01		1.280E-05
	Min	2.129E+02		9.500E+02		5.722E+02		4.049E+02		1.235E+03		0.000E+00
F10	Mean	1.861E+01	+	1.951E+01	+	4.199E+00	+	1.575E+01	+	2.039E+01	+	1.537E-14
	St. Dev.	5.462E-01		2.553E-01		5.341E-01		1.145E+00		1.293E-01		3.398E-14
	Min	1.670E+01		1.897E+01		3.444E+00		1.284E+01		2.006E+01		4.441E-16
F11	Mean	1.570E+01	+	9.917E+02	+	9.750E-01	+	6.393E+01	+	9.979E+01	+	0.000E+00
	St. Dev.	4.100E+01		8.318E+01		4.646E-02		2.747E+01		1.390E+01		0.000E+00
	Min	7.129E-02		8.201E+02		8.485E-01		2.568E+01		7.361E+01		0.000E+00
F12	Mean	2.859E+02	+	5.301E+08	+	2.988E+03	+	1.379E+06	+	1.392E+09	+	5.005E-01
	St. Dev.	6.397E+02		1.440E+08		8.741E+03		1.759E+06		2.812E+08		4.198E-01
	Min	1.425E+01		2.186E+08		8.739E+00		3.766E+04		6.339E+08		4.151E-04
F13	Mean	6.285E+04	+	1.173E+09	+	2.135E+04	+	5.801E+06	+	2.573E+09	+	6.895E+00
	St. Dev.	2.217E+04		2.290E+08		5.463E+04		5.531E+06		4.791E+08		4.174E+00
	Min	1.120E+04		6.274E+08		1.272E+02		4.045E+05		1.666E+09		5.662E-02
F14	Mean	7.918E+03	+	4.391E+06	+	1.662E+04	+	3.397E+05	+	5.543E+06	+	9.533E-26
	St. Dev.	1.857E+04		4.554E+05		3.097E+03		1.615E+05		6.585E+05		3.178E-25
	Min	1.041E+03		3.447E+06		1.041E+04		1.650E+05		4.357E+06		1.228E-40
F15	Mean	1.464E+03	-	2.480E+03	+	1.172E+03	-	1.491E+03	-	2.519E+03	+	2.315E+03
	St. Dev.	1.058E+02		2.535E+02		9.584E+01		1.312E+02		2.345E+02		9.103E+01
	Min	1.272E+03		1.995E+03		1.047E+03		1.223E+03		2.017E+03		2.080E+03
F16	Mean	6.388E+02	-	7.446E+02	-	3.989E+02	-	4.961E+02	-	9.037E+02	=	9.161E+02
	St. Dev.	6.302E+01		7.359E+01		2.929E+01		5.488E+01		9.710E+01		8.486E+01
	Min	5.078E+02		6.188E+02		3.517E+02		3.895E+02		7.458E+02		7.173E+02
F17	Mean	1.483E+03	+	1.482E+03	+	1.409E+03	+	1.407E+03	+	1.500E+03	+	9.000E+02
	St. Dev.	4.642E+01		3.113E+01		3.634E+01		3.282E+01		3.949E+01		0.000E+00
	Min	1.364E+03		1.428E+03		1.331E+03		1.333E+03		1.435E+03		9.000E+02
	Win	15		16		14		15		16		
	Tie	0		0		0		0		1		
	Lose	2		1		3		2		0		

[1] According to the Wilcoxon signed rank test at 5% significance level, the symbols "+", "=" and "−" symbolize that the performance of SASMA is better, similar, or worse than that of other SAEAs, respectively. In addition, the best results for all metrics (from the 35 runs) for each test function are highlighted in grey.

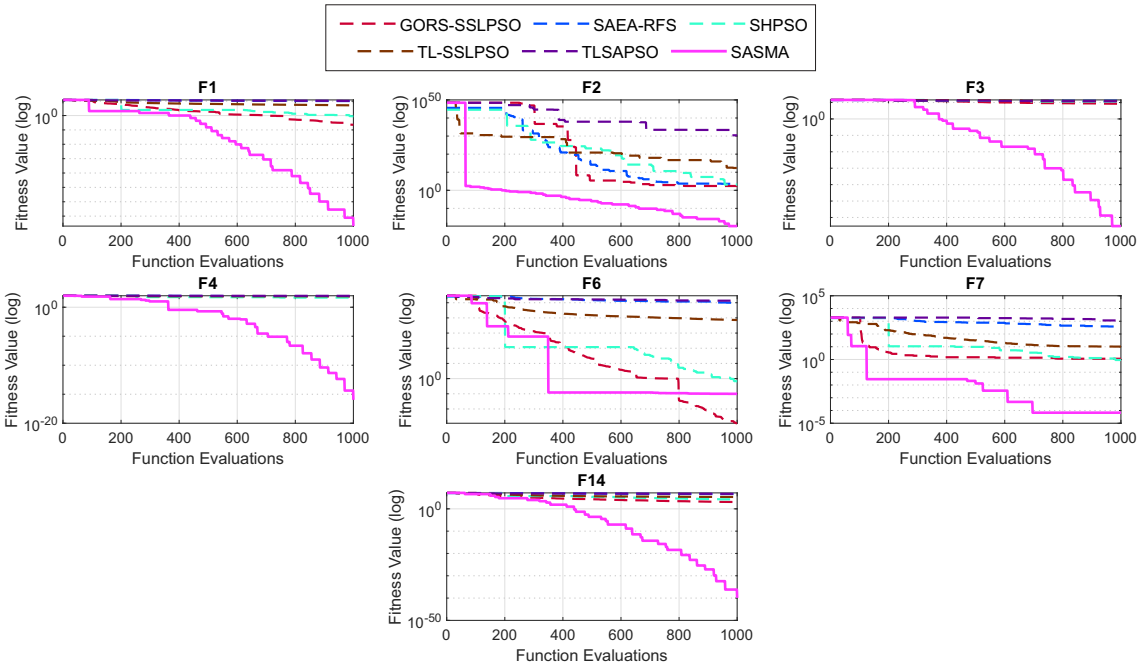


Figure A.1: Convergence curves (fitness value): GORS-SSLPSO (dashed red), SAE-ARFS (dashed blue), SHPSO (dashed cyan), TL-SSLPSO (dashed brown), TLSAPSO (dashed purple) and SASMA (solid pink) on 100D unimodal test functions (F1-F4,F6,F7 and F14).

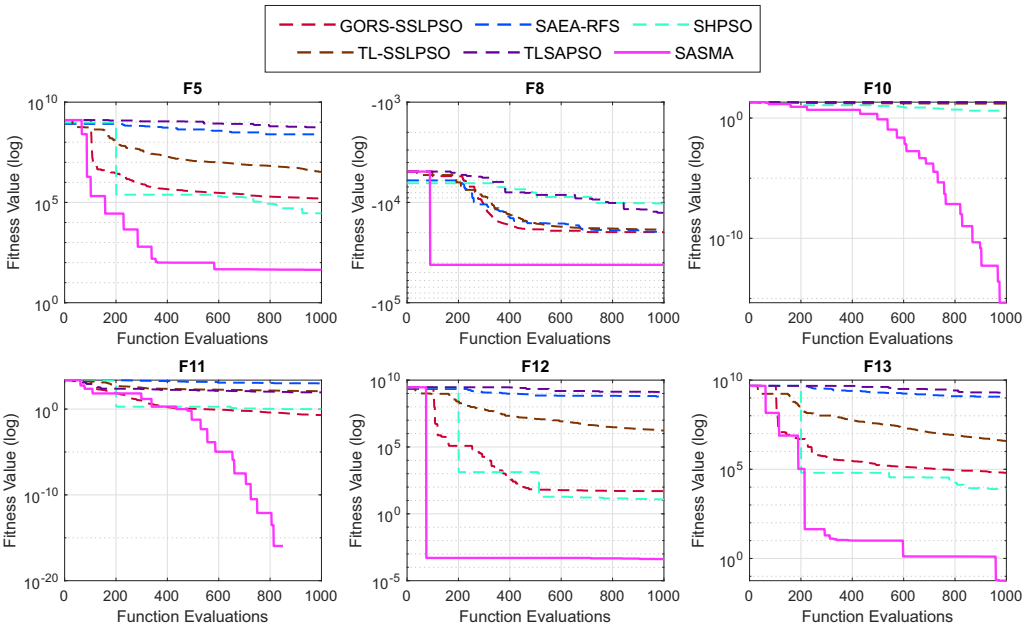


Figure A.2: Convergence curves (fitness value): GORS-SSLPSO (dashed red), SAE-ARFS (dashed blue), SHPSO (dashed cyan), TL-SSLPSO (dashed brown), TLSAPSO (dashed purple) and SASMA (solid pink) on 100D multimodal test functions (F5,F8 and F10-F13).

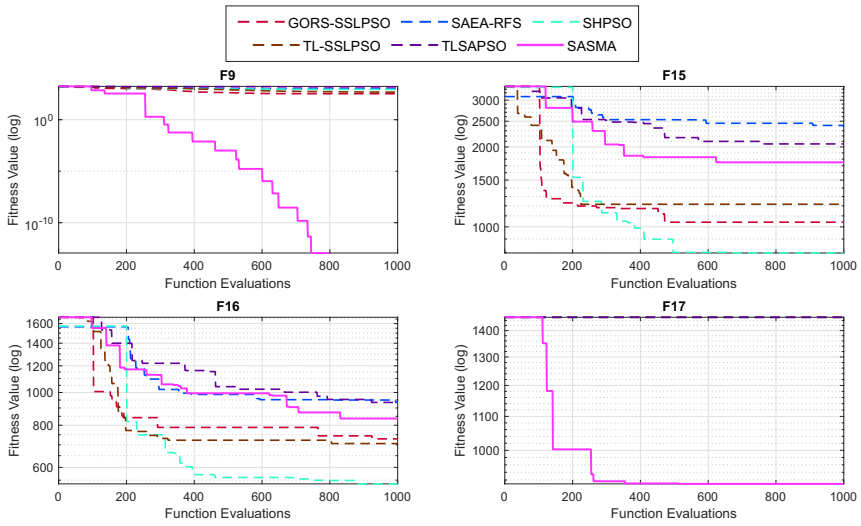


Figure A.3: Convergence curves (fitness value): GORS-SSLPSO (dashed red), SAE-ARFS (dashed blue), SHPSO (dashed cyan), TL-SSLPSO (dashed brown), TLSAPSO (dashed purple) and SASMA (solid pink) on 100D very complicated multimodal test functions (F9 and F15-F17).

Appendix B

A modified Harris Hawks Optimization (mHHO)

This appendix introduces the Harris Hawks Optimization (HHO) algorithm, outlining its search space exploration and exploitation mechanisms — while assuming familiarity with the notation for position updates and auxiliary variables previously defined in sub-section 4.5.2 — and presents the proposed modifications (compiled in the pseudo-code form in Algorithm 2) designed to enhance its effectiveness in solving a concrete OPF problem.

B.1 Harris Hawks Optimization (HHO)

Inspired by the cooperative foraging activities of Harris hawks, HHO algorithm was developed by Heidari et al. [535] and it replicates the unique hunting prey behavior of this species via several exploration and exploitation mechanisms that comprise the two main solution search stages. In the optimization context, the "prey" is the best candidate solution after each iteration (best position as of now), based on an objective function value, $f_{obj}\left(X_i\right)$, while individual hawks, X_i, are the evolving candidate solutions, (from a population of n_{pop} individuals defined across all the problem dimensions, n_{dims}). Transition between search stages and the mechanisms to conduct the search space are described below.

B.1.1 From Exploration to Exploitation

The concept of prey's escape energy, $E_{escprey}$, is determined iteratively using Equation B.1. It works as a control variable that determines a binary decision. That is, to continue the global exploration stage $|E_{escprey}| \geq 1$, seeking solution diversification, or to otherwise conclude and move to the local exploitation stage $|E_{escprey}| < 1$, i.e., traversing the search space only in the vicinity of a promising solution.

$$E_{escprey}\left(iter\right) = 2E_0\left(iter\right)\left(1 - \frac{iter}{iter_{\max}}\right) \quad \text{(B.1)}$$

Where $E_0\left(iter\right)$ is an iteratively random number in the interval $(-1, 1)$, denoting a decay trend in the escape energy as the iteration number, $iter$, grows is the current iteration number; and $iter_{\max}$ is the maximum number of iterations.

B.1.2 Exploration stage

In this stage, the algorithm promotes the use of randomized operators to thoroughly probe distant problem space-regions. So, the ith hawk updated position, $X_i\,(iter+1)$, is computed as given in Equation B.2, where q, r_1, r_2, r_3 and r_4 are random numbers with a uniform distribution between 0 and 1, with q representing a control variable that ensures a balanced exploration update strategy, considering both the individual and the (hawk) population "experiences"; establishing that half of the times, the new position is based on its current distance (per dimension) to a fellow hawk (chosen randomly from the current population), $X_{rand}\,(iter)$ with $\{rand \in \mathbb{N} | 1 \leq rand \leq n_{pop}\}$; while for the other half , the hawk will travel accordingly to the current prey's (best) position, $X_{prey}\,(iter)$, i.e., a population "memory" of the best recorded candidate solution, $X_{prey}\,(iter) \rightarrow \min f_{obj}\,(X_i)$; the current average of position of all hawks, $X_{mean}\,(iter) = \frac{1}{n_{pop}} \sum_{i=1}^{n_{pop}} X_i\,(iter)$; and the search space upper and lower bounds, ub and lb, respectively.

$$X_i\,(iter+1) = \begin{cases} X_{rand}\,(iter) - r_1\,|X_{rand}\,(iter) - 2r_2 X_i\,(iter)|\,, & q \geq 0.5 \\ X_{prey}\,(iter) - X_{mean}\,(iter) - r_3\,(lb + r_4\,(ub - lb))\,, & q < 0.5 \end{cases} \tag{B.2}$$

B.1.3 Exploitation stage

After finding a promising prey, this more intricate stage, is responsible for narrowing in the search to the surroundings of $X_{prey}\,(iter)$, and as a result, find the objective's function optimal value. The same is achieved through soft and hard besiege mechanisms, based (again) on the control variable, $E_{escprey}\,(iter)$. As the name suggests, these two mechanisms translate different encircling strategies of the prey, and each one is subdivided into two position update equations, including one with "progressive rapid dives". Its use is based on a new parameter, r, a random number in the interval $(0, 1)$, indicating the prey's escape chances during the current besiege stage.

B.1.3.1 Soft Besiege and Soft Besiege with Progressive Rapid Dives

This represents the initial phase of the exploitation stage, where hawks actively pursuit their prey, which still has some escape energy,$|E_{escprey}| \geq 0.5$. Where for $r \geq 0.5$, the update is given by Equation B.3, while when $r < 0.5$, the updated position is given by Equation B.4, where a progressive rapid dive is implemented. This besiege strategy is centered around the use of Lévy flights, Equation B.5 and Equation B.6, that mimic the irregular leapfrog movements of the prey, and so the hawks are able to "anticipate" a change in the prey's position $X_{prey}\,(iter)$. Where J is a random number in the interval $(0, 2)$; S is a random vector with size $1 \times n_{dims}$, in the interval $(0, 1)$; $LF\,(x)$ is the function modeling the Lévy flights; μ and v are normally distributed random vectors with a $1 \times n_{dims}$ size; β is a constant usually set to equal to 1.5; and lastly, Γ, is the gamma function.

$$X_i\left(iter+1\right) = X_{prey}\left(iter\right) - X_i\left(iter\right) - E_{escprey}\left(iter\right)\left|JX_{prey}\left(iter\right) - X_i\left(iter\right)\right| \quad \text{(B.3)}$$

$$X_i\left(iter+1\right) = \begin{cases} \Upsilon_i = X_{prey}\left(iter\right) - E_{escprey}\left(iter\right) \\ \qquad\qquad \left|JX_{prey}\left(iter\right) - X_i\left(iter\right)\right|, & \text{if } f_{obj}(\Upsilon_i) < f_{obj}\left(X_i\left(iter\right)\right) \\ Z_i = \Upsilon_i + S \times LF\left(n_{dims}\right), & \text{if } f_{obj}(Z_i) < f_{obj}\left(X_i\left(iter\right)\right) \end{cases} \quad \text{(B.4)}$$

$$LF\left(x\right) = \frac{\mu\sigma}{100|v|^{\beta^{-1}}} \quad \text{(B.5)}$$

$$\sigma = \left(\frac{\Gamma\left(1+\beta\right)\sin\left(\frac{\pi\beta}{2}\right)}{\Gamma\left(\frac{1+\beta}{2}\right)\beta 2^{\left(\frac{\beta-1}{2}\right)}}\right)^{\beta^{-1}} \quad \text{(B.6)}$$

B.1.3.2 Hard Besiege and Hard Besiege with Progressive Rapid Dives

When the prey's energy is drained below the 0.5 threshold, because of the different besiege mechanisms. If $r \geq 0.5$, the update is given by B.7, while when $r < 0.5$, the updated position is given by B.8, and the rapid dive now favors a position update equation, which considers the population average position, $X_{mean}\left(iter\right)$, rather than the individual one, $X_i\left(iter\right)$.

$$X_i\left(iter+1\right) = X_{prey}\left(iter\right) - E_{escprey}\left(iter\right)\left|X_{prey}\left(iter\right) - X_i\left(iter\right)\right| \quad \text{(B.7)}$$

$$X_i\left(iter+1\right) = \begin{cases} \Upsilon_i = X_{prey}\left(iter\right) - E_{escprey}\left(iter\right) \\ \qquad\qquad \left|JX_{prey}\left(iter\right) - X_{mean}\left(iter\right)\right|, & \text{if } f_{obj}(\Upsilon_i) < f_{obj}\left(X_i\left(iter\right)\right) \\ Z_i = \Upsilon_i + S \times LF\left(n_{dims}\right), & \text{if } f_{obj}(Z_i) < f_{obj}\left(X_i\left(iter\right)\right) \end{cases} \quad \text{(B.8)}$$

B.1.4 Modifications

The multiple HHO update position mechanisms are a core strength of the algorithm, however, further refinements can be made to enhance its standardized features to better capture the global "prey" and escape the local optima trap [658]. With this purpose, Equation B.1 is modified, to ensure that exploitation and exploration can occur throughout the iterative process with highly arbitrary patterns, translating a smoother transition between the exploration and exploitation stages. To achieve this, chaotic maps with its powerful diversification features are explored in [659], thus, leading to a "less regular"

non-monotonic behavior, given by Equations B.9 and B.10, while maintaining the overall declining trend of the function envelope, and so the transition between stages. These modifications result in a tweaked pattern, Figure B.1, where for the considered number of iterations, the original versus the modified escape energy behavior is shown, together with a band around the 0.5 threshold (transition between exploitation mechanisms).

$$Cmap_{Log}\left(iter+1\right) = 4\,Cmap_{Log}\left(iter\right) \times \left(1 - Cmap_{Log}\left(iter\right)\right) \tag{B.9}$$

$$E_{escprey}\left(iter\right) = 2\left(1 - \frac{iter}{iter_{\max}}\right) \times Cmap_{Log}\left(iter+1\right)\cos\left(\pi iter\right) \tag{B.10}$$

Where $Cmap_{Log}\left(iter\right)$ is the logistic map function computed at each iteration [660], with its initial value being $Cmap_{Log}\left(0\right) = 0.3$.

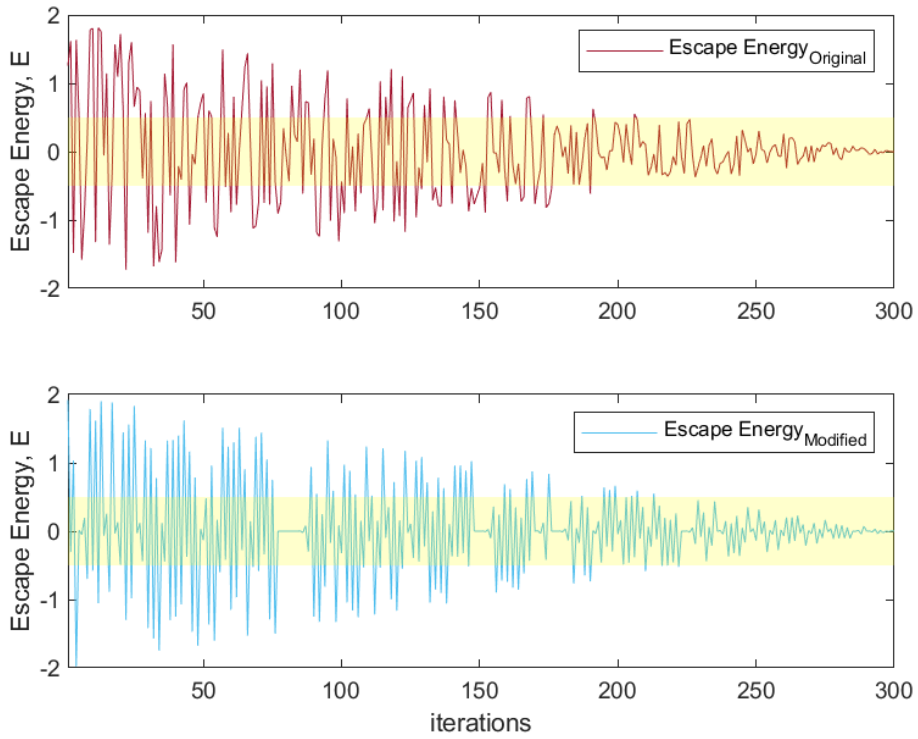


Figure B.1: $E_{escprey}$ modified behavior illustration.

Besides, another contentious issue with HHO are its diversity mechanisms (during exploration), crucial to avoid local traps (optimums), again as recognized in [658]. The original mechanisms, as given in Equation B.2, bases its functioning on the use of four random numbers in a position update scheme that is either dependent on a randomly chosen hawk from the current population (half of the times), which can include the position being updated (itself), X_i, or a position update based on the distance from the currently best

position, X_{prey}, to the mean position per every dimension of the entire hawk population, X_{mean}, and then to a newly created position (the other half). As such, one can argue that the diversification stage can be enhanced by considering additional randomly (entirely new) search space locations, rather than simply use random numbers as scale changes applied to all the dimensions of the calculated differences, and so all referencing current swarm positions, which most likely is limiting the exploration stage somewhat.
Taking this into consideration, the modified HHO, will consider the powerful exploration capabilities of the population-based metaheuristic DE, namely its mutation and crossover mechanisms [661]. With the first, usually resorting to a series of vector additions/differences between randomly selected positions and/or existing positions, thus, resulting in a (single) new mutated position (often referred to as "donor vector"). While the second mechanism, subjects the current position to a dimension-wise crossover with this donor vector. To control this binomial crossover, a user-defined (crossover) factor, $CR = 0.3$, is compared for each ith position and jth dimension (at each iteration) against a uniformly distributed random number, r_0, in the range $[0, 1]$; and guaranteeing through an extra condition, based on a position index $\{j_{rand} \in \mathbb{N} \mid 1 \leq j_{rand} \leq n_{pop}\}$, that the updated position contains at least one copied dimension from the mutated position, and so enforcing the targeted diversification.
Nevertheless, for the mutation process, and unlike the traditional approach, this work proposes a scheme with two mutations (resulting in two donor vectors), formulated inside the crossover piece-wise function B.11. Where, the first is computed according to the popular "DE/rand/1" mutation variant [662], considering a mutation factor $F_1 \in]0, 2]$ and three randomly initialized positions (per dimension j), $X_{rand1,j}$, $X_{rand2,j}$ and $X_{rand3,j}$, as given by B.11a; whereas the second, is inspired by the "DE/best/1" [662], and so as the name suggests it considers the current best position, $X_{prey,j}\,(iter)$, but instead of the two randomly generated positions, in the term multiplied by the mutation factor F_2, only one is considered. That is, $X_{rand4,j}$. While the current hawk position for the jth dimension, $X_{i,j}\,(iter)$, takes the place of the other. With this modification, we ensure that the current hawk position (commonly known as "target vector") also contributes for the updated position $X_i\,(iter + 1)$ via the second mutated position, as given by B.11b.

$$X_{i,j}\,(iter + 1) = X_{rand1,j}\,(iter) + F_1\,(X_{rand2,j}\,(iter) - X_{rand3,j}\,(iter))\,, \quad \text{if } r_{0,j} \leq CR \vee j = j_{rand} \tag{B.11a}$$

$$X_{i,j}\,(iter + 1) = X_{prey,j}\,(iter) + F_2\,(X_{rand4,j}\,(iter) - X_{i,j}\,(iter))\,, \quad \text{if } r_{0,j} > CR \vee j \neq j_{rand} \tag{B.11b}$$

B.1.5 mHHO: The Pseudo-code

,

Algorithm 2 Pseudo-code of mHHO algorithm

Input: population size n_{pop}, maximum number of iterations $iter_{\max}$, number of problem dimensions n_{dims}, problem lower lb and upper ub bounds, crossover factor CR

Output: Prey's location $X_{prey}\left(iter_{\max}\right)$, its objective value $f_{obj}\left(X_{prey}\left(iter_{\max}\right)\right)$, and convergence curve

Randomly initialize the population of hawks, $X_i\left(0\right)\ \forall\, i \in (1, n_{pop})$
while $(iter \leq iter_{\max})$ **do**
    for $i = 1$ to n_{pop} **do**
        for $j = 1$ to n_{dims} **do**
            if $(X_{i,j} < lb)$ **then**
                $X_{i,j} = lb$
            else if $(X_{i,j} > ub)$ **then**
                $X_{i,j} = ub$
            end if
        end for
    end for
    for $i = 1$ to n_{pop} **do**
        Calculate the objective value of the hawks, $f_{obj}\left(X_i\right) \forall\, i \in (1, n_{pop})$
    end for
    Find the current best hawk (lowest f_{obj}), $prey = \arg\min_i f_{obj}\left(X_i\left(iter\right)\right) \forall\ i \in (1, n_{pop})$, and set it as the prey's most promising location, $X_{prey}\left(iter\right)$
    for $i = 1$ to n_{pop} **do**
        Compute the iterative chaotic number, $Cmap_{Log}\left(iter + 1\right)$, using Equation B.9
        Update the prey's escape energy, $E_{escprey}\left(iter\right)$, using Equation B.10
        if $(|E_{escprey}\left(iter\right)| \geq 1)$ **then**
            Update the hawk's positions, using Equation B.11
        else if $(|E_{escprey}\left(iter\right)| < 1)$ **then**
            if $(r \geq 0.5$ and $|E_{escprey}\left(iter\right)| \geq 0.5)$ **then**
                Update the hawk's positions, using Equation B.3
            else if $(r \geq 0.5$ and $|E_{escprey}\left(iter\right)| < 0.5)$ **then**
                Update the hawk's positions, using Equation B.7
            else if $(r < 0.5$ and $|E_{escprey}\left(iter\right)| \geq 0.5)$ **then**
                Update the hawk's positions, using Equation B.4, considering Equation B.5 and Equation B.6
            else if $(r < 0.5$ and $|E_{escprey}\left(iter\right)| < 0.5)$ **then**
                Update the hawk's positions, using Equation B.8, considering Equation B.5 and B.6
            end if
        end if
    end for
end while
return $X_{prey}\left(iter_{\max}\right)$ and $f_{obj}\left(X_{prey}\left(iter_{\max}\right)\right)$

Appendix C

Proposed Methodology for STPF: Pseudo-code

The following pseudo-code (Algorithm 3) provides a high-level execution outline of the methodology described in Section 7.2.2, illustrating the main stages from data selection to the final Short-Term Load Forecasting (STLF). Although not a conventional pseudo-code, it is designed to facilitate understanding of the overall workflow and logical structure of the proposed approach.

Algorithm 3 Ensembled DFNN-based Short-Term Load Forecasting (STLF) Procedure

Input: Load data, calendar and exogenous variables
Output: 24-hour ahead load forecasts

INIT
GET and format ISO-NE load data and relevant variables
COMPUTE ACF and PACF for time-series analysis
LIST suitable ARIMA/SARIMA models based on correlation and seasonality
SET datasets with different windowing (δ)
CALCULATE ADF test to determine integration order d
FIT Box-Jenkins models to training windows
COMPUTE BIC (Equation 7.7) to select top 3 models per window
DETERMINE and store 24h-ahead STLF using selected models (15 base learners)
DEFINE input dataset: concatenate ARIMA forecasts with exogenous/calendar variables and target
NORMALIZE input data using Equation 7.8 and prepare training/validation/testing sets
DEFINE DFNN hyperparameters:
Architecture: 3 hidden layers [20, 10, 5]
Optimizer: SCG
Performance metric: modPerfF (Equation 7.9)
Train/Validation split: 70%/30%
Activation: Sigmoid (hidden), Linear (output)
SET number of runs = 50
for each run **do**
    TRAIN DFNN offline
    while iterations $<$ 2500 **do**
        if validation error stops decreasing and counter $<$ 50 **then**
            Counter += 1
        else if counter $\geq$ 50 **then**
            break
        else
            Counter = 0
        end if
    end while

```
    COMPUTE STLF using trained DFNN on test set
    STORE predicted loads and forecasts
    UPDATE training/validation sets with new predictions
    if no more test weeks then
        break
    end if
end for
END
```

,

Appendix D

Using STLF as an exogenous input for STPF: a case study on the MIBEL data

As discussed in sub-section 7.3.1.4, load forecasts are considered correlated variables to electricity prices. Therefore, in the context of the sensitivity analysis, they were treated as intermediate objectives. These forecasts were subsequently used as input variables to help the artificial neural network (ANN) learn and represent the underlying relationship with the day-ahead spot prices in the MIBEL market. As such, this appendix chapter is dedicated to presenting the results derived from this analysis.

D.1 Intermediate STLF in Peninsular Spain and Portugal: Methodology

For the selected input variables, the consensus established in the field literature reviewed in section 7.2 was built upon. Therefore, in addition to lagged load samples and all the common calendar variables, the model also incorporated average lagged daily temperatures for the two largest population centers in each country. This inclusion was motivated by the well-known correlation between temperature and electricity consumption, often referred to as heating load.

Regarding the forecasting model, a multilayered FFNN similar to the one used for the main STPF task was employed—specifically, an ML model configured with the same hyperparameters outlined in Section 7.3.1.3.

D.2 Intermediate STLF in Peninsular Spain and Portugal: Results

The intermediate results exhibited strong performance, with average monthly MAPE values remaining below 3% and monthly RMSE values consistently under 190 MW, as shown in Table D.1.

The relatively low standard deviations and the narrow range between minimum and maximum errors further support the accuracy and robustness of the chosen approach and input variables for 24-hour ahead load prediction.

Figures D.1 and D.2 illustrate the forecasting accuracy for a test month in the Peninsular Portuguese and Spanish cases, respectively, confirming the reliability of the metrics through a strong alignment between the curves, with consistent patterns across most areas.

Table D.1: STLF of the Spanish and Portuguese System Load : Error Metrics Analysis.

Month	Region	MAPE (%)				RMSE [MW]			
		Mean	Min	Max	Std	Mean	Min	Max	Std
Feb 2019	Spain	1.67	1.54	1.80	0.07	150.22	138.10	160.75	5.76
	Portugal	1.67	1.54	1.80	0.07	150.22	138.10	160.75	5.76
May 2019	Spain	2.09	1.98	2.22	0.06	184.94	175.95	202.64	6.69
	Portugal	2.09	1.98	2.22	0.06	184.94	175.95	202.64	6.69
Aug 2019	Spain	2.41	2.07	2.55	0.10	174.70	154.43	188.70	7.86
	Portugal	2.41	2.07	2.55	0.10	174.70	154.43	188.70	7.86
Oct 2019	Spain	1.72	1.62	1.90	0.08	124.82	114.29	139.18	6.32
	Portugal	1.72	1.62	1.90	0.08	124.82	114.29	139.18	6.32

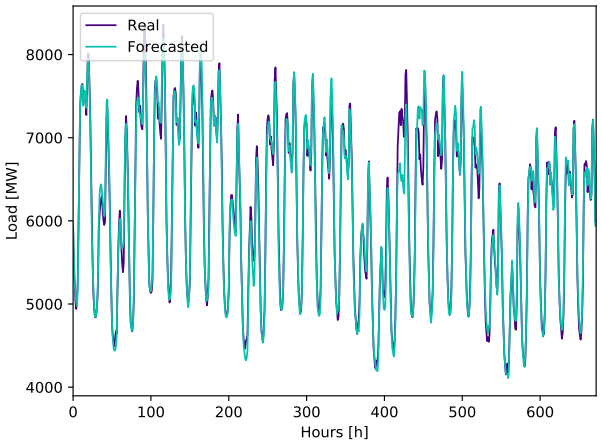


Figure D.1: Day-ahead load forecast [MW] of the Portuguese System for the month of February 2019: Real Load (in violet) vs Forecasted Load (in teal green).

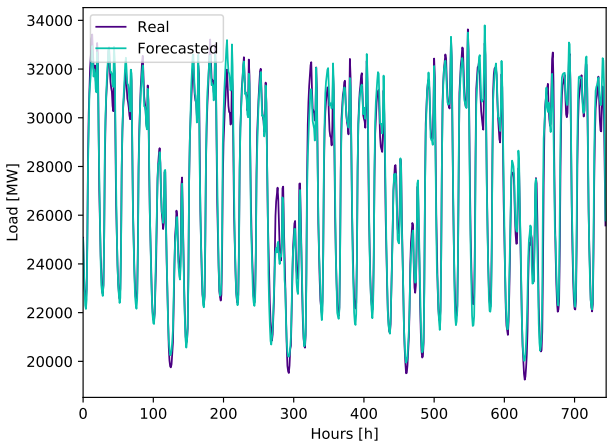


Figure D.2: Day-ahead load forecast [MW] of the Spanish System Load for the month of October 2019: Real Load (in violet) vs Forecasted Load (in teal green).

Additionally, one can notice that the largest errors tend to occur during peak load periods and on weekdays—an expected trait that may indicate an anomaly not explicitly captured by the provided context (input labeled data). ,

Appendix E

HES: proposed operational strategy, data analysis and modeling

To maintain focus and avoid overloading the main body of the thesis with extended discussions on the studied daily operation optimization of a hybrid energy system considering a short-term electricity price forecast (sub-section 7.3.2)— a topic related to EPSs but not directly tied to the forecasting task itself — this appendix provides all relevant supplementary implementation details. It includes the key principles and full formulation guiding the dispatch strategy, case study data analysis and the full HES modeling.

E.1 Proposed operation strategy

The overall operation strategy starts with a data gathering and features extraction stage (i), where the different time-series are analyzed and processed. Subsequently, the short-term price forecasting (ii) is carried out for a one-day ahead horizon, as explained in sub-section 7.3.2.1 and summarized in Figure 7.15.

Following the forecasting stage (ii), for each discrete-time step k, up to a total of KN steps (simulation time horizon), it is determined how much the non-dispatchable PV and wind generations can supply the existing load $P_D\,(k)$ given by $\varepsilon(k)$ (iv). This will result in situations where $\varepsilon(k) > 0$, i.e., energy deficit, which involves the discharge of the storage technologies or the purchase of energy from the grid. On other hand, when $\varepsilon(k) \leq 0$, i.e., energy surplus, requiring that the energy surplus is either stored in the storage technologies or sold to the grid.

Thereafter, the economic dispatch is carried out, i.e., allocating the available generation to supply the HES load, while minimizing the operation costs (v) and fulfilling the demand–supply balance (vi) and unit capacity limits constraints (x).

In (v) the ED stage is executed by linear programming, determining the output power of the different units that form the proposed HES, to fulfill the existing energy deviation $\varepsilon(k)$ in the most cost-effective manner. C_{solar}, C_{wind}, C_j and C_{grid} are the costs associated with the respective generation technologies and the grid electricity cost (energy price), respectively. In particular, C_j translates the storage unit and PSH cost as defined in Equation E.1 and are computed as a function of the state of charge, SOC_j, and an estimate of how much the SOC of the storage technologies varies, ΔSOC_j, as well as the forecasted prices, $\hat{p}\,(k)$, where the index j selects the battery bank for $j = 1$ and PSH as a source/sink for $j = 2$. In the case of the PSH, SOC_2 is defined as the ratio between the current volume of stored water and the reservoir maximum capacity. Equation E.1 was defined as a piecewise function where each sub-function has its own SOC_j, working range and signal of $\varepsilon(k)$. The considered SOC working range for both storage technologies were [0.15,1], implying that the

domain of ΔSOC_j, is $[-0.85, 0.85]$.
The marginal costs of the non-dispatchable generation are considered null. Regarding the grid cost, the same is computed as a function of the state of charge, SOC_j, and the electricity price forecast (predicted value $\hat{p}$, predicted daily average and maximum predicted value) and is given by Equation E.2. This cost was also defined as a piecewise function where each sub-function is defined over an interval formed by the signal of $\varepsilon(k)$ and $\hat{p}\,(k)$. The inclusion of the electricity price forecast information on the formulation allows to gauge the future trend of electric energy prices, which is of high interest for a player with this type of generation mix. Since it allows to dispatch the stored energy when high electricity prices are expected and, on the opposite, to take advantage of off-peak periods for storing energy when lower electricity prices are expected.
For instance, in the presence of energy deficits, $\varepsilon(k) > 0$, the $SOC_j(k)$ is inversely proportional to the attributed grid cost, furthermore a normalized price information $\frac{1}{24}\sum_{j=1}^{24}\hat{p}_j$ $/\max\left(\{\hat{p}_1 ... \hat{p}_{24}\}\right)$ regarding the considered day will translate how the average compares with the maximum forecasted value, finally the predicted $\hat{p}\,(k)$ price for the current hour is compared with its daily average, decreasing the grid cost if $\hat{p}\,(k) < \frac{1}{24}\sum_{n=1}^{24}\hat{p}_n$ and otherwise increasing it, in turn for the energy surplus situation, $\varepsilon(k) \leq 0$, the $SOC_j(k)$ is now directly proportional to the attributed grid cost and instead of adding the factor regarding the difference between the current dispatch price and its daily average, the same is now subtracted.
From point (vi) forward, we focus the attention on the set of operational and system constraints to which the ED is subject. The first system operation constraint is the demand–supply balance equation (vi), translating the need for the non-dispatchable generation, $P_{\text{solar}}\,(k)$ and $P_{\text{wind}}\,(k)$, coupled with the sum of the output power from the PSH and battery bank units, $Pg_j\,(k)$, as well as the power flow with the upstream grid, $P_{\text{grid}}\,(k)$, to match at all times the HES load, $P_{\text{D}}\,(k)$. In addition, it was established that for energy deficit occurrences, the output power $Pg_j\,(k)$ and $P_{\text{grid}}\,(k)$ have by convention a positive signal (vii), i.e., power is being injected into the buses (HES), contrarily, when $P_{\text{grid}}\,(k) < 0$ and $Pg_j\,(k) < 0$ power is now being injected into the grid and being consumed for pumping up the water (upstream reservoir) and/or to charge the storage bank (viii). Besides, an additional constraint concerning the operating range of the PSH unit is defined, imposing 10% of the rated power of the turbine, as minimum operating point both for turbine and pump modes (vii-b and viii-b). The last constraints (x) ensure that the power generation limits, firstly for battery bank, translates the lower and upper charge and discharge power, and secondly the generation/consumption bounds for the PSH unit, which are a function of the current SOC_2 and its generator rated power, $P_{\text{hydrorated}}$.

$$C_j\left(SOC_j(k), \Delta SOC_j(k), \hat{p}\left(k\right)\right) = \begin{cases} 2, & \text{if } \left(SOC_j(k) \leq 0.15 \wedge \varepsilon(k) > 0\right) \vee \left(SOC_j(k) = 1 \wedge \varepsilon(k) \leq 0\right) \\ \dfrac{\left(1 - SOC_j(k) + \Delta SOC_j(k)\right)\left(\frac{1}{24}\sum_{j=1}^{24}\hat{p}_j\right)}{\max\left(\{\hat{p}_1, \ldots, \hat{p}_{24}\}\right)}, & \text{if } 0.15 \leq SOC_j(k) \leq 1 \wedge \varepsilon(k) > 0 \\ \dfrac{SOC_j(k)\left(\frac{1}{24}\sum_{j=1}^{24}\hat{p}_j\right)}{\max\left(\{\hat{p}_1, \ldots, \hat{p}_{24}\}\right)}, & \text{if } SOC_j(k) < 1 \wedge \varepsilon(k) \leq 0 \end{cases} \tag{E.1}$$

$$C_{\text{grid}}\left(\hat{p}\left(k\right), SOC_j(k)\right) = \begin{cases} \hat{p}_1, \ldots, \hat{p}_{24}, & \text{if } \varepsilon(k) > 0 \wedge \hat{p}\left(k\right) > 0 \\ \min\left(\{SOC_1(k),\ SOC_2(k)\}\right) \cdot \dfrac{\frac{1}{24}\sum_{j=1}^{24}\hat{p}_j}{\max\left(\{\hat{p}_1, \ldots, \hat{p}_{24}\}\right)} - \dfrac{\hat{p}\left(k\right) - \frac{1}{24}\sum_{j=1}^{24}\hat{p}_j}{\max\left(\{\hat{p}_1, \ldots, \hat{p}_{24}\}\right)}, & \text{if } \varepsilon(k) \leq 0 \wedge \hat{p}\left(k\right) > 0 \\ 0, & \text{if } \hat{p}\left(k\right) \leq 0 \end{cases} \tag{E.2}$$

Lastly, several tweaks were also considered: for small discharge depths and state of charges close to 1 the cost of both technologies is similar; with state of charges close to 1 and as the discharge depth increases, the cost of using the battery bank is higher. This is especially important, since the battery bank is a more sensible storage technology, i.e., its preferable to avoid full discharges cycles, even if it requires charging cycles more frequently.
A concise pseudo-code representation of the dispatch problem is provided in Algorithm 4. Although not intended as an executable sequence, the structure was chosen for ease of understanding and to clearly convey the logical flow of the approach.

Algorithm 4 Operational strategy.

init: Data gathering and features extraction
price forecasting: $\{\hat{p}_1, \hat{p}_2, ..., \hat{p}_{24}\}$
for $k = 1$ to K_{N} **do**
    $\varepsilon(k) = P_{\mathrm{D}}(k) - (P_{\text{solar}}(k) + P_{\text{wind}}(k))$
    minimize:

$$C_{\text{solar}} P_{\text{solar}}(k) + C_{\text{wind}} P_{\text{wind}}(k) + C_{\text{grid}}(\varepsilon(k)) P_{\text{grid}}(k) + \sum_{j=1}^{2} C_j(SOC_j(k), \Delta SOC_j, \hat{p}(k)) Pg_j$$

    subject to system constraints:
    $P_{\text{solar}}(k) + P_{\text{wind}}(k) + P_{\text{grid}}(k) + \sum_{j=1}^{2} Pg_j(k) = P_{\mathrm{D}}(k)$
    if $\varepsilon(k) > 0$ **then**
        $Pg_1(k) > 0$
        $Pg_2(k) > 0.1 P_{\text{hydro rated}}$
        $P_{\text{grid}}(k) > 0$
    else
        $Pg_1(k) \leq 0$
        $Pg_2(k) \leq -0.1 P_{\text{hydro rated}}$
        $P_{\text{grid}}(k) \leq 0$
    end if
    additional constraints:

$$Pg_{j,\min}(k) \leq Pg_j(k) \leq Pg_{j,\max}(k)$$
$$P_{\text{grid}}(k) \leq P_{\text{grid, max}}$$

end for

To better highlight the co-domain of this cost function, Figure 2 is introduced, where four different initial SOCs are considered for an energy deficit situation (where a discharge of the storage technologies is expected). The chosen SOCs, were the upper limit, i.e., SOC = 1, followed by two intermediate levels (equally spaced), i.e., SOC = 0.5 and SOC = 0.75 and a near border limit of SOC = 0.85.
Figure E.1 clearly shows that the cost is inversely proportional to the current SOC and varies in a way that favors the use of PSH and storage bank technologies when not only its SOCs are higher but also when the required DOD is smaller. Although the function ranges between [−1,1.85], for energy deficit situations its respective cost is always positive, since discharge cycles have negative repercussions on these technologies' lifespan.

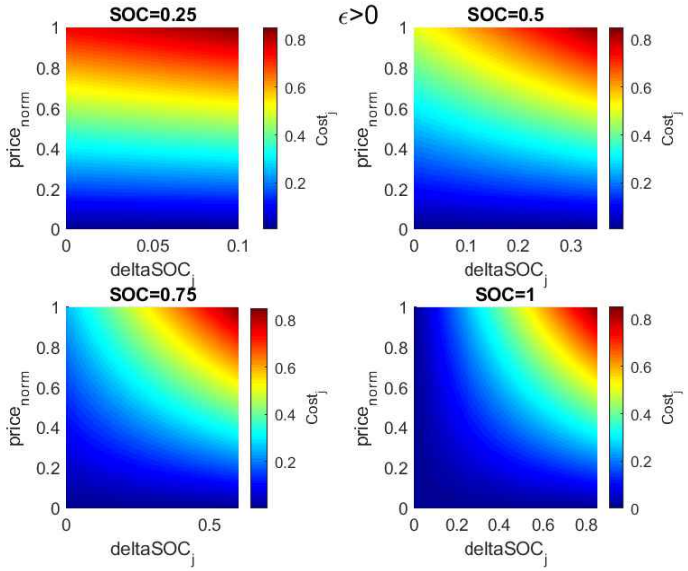


Figure E.1: Equation (2) co-domain illustration: PSH and battery bank costs for four different SOCs.

E.2 A brief statistical analysis of the datasets

Table E.1 provides the time-series descriptive statistics after the adjustments mentioned before, i.e., mean, median, standard deviation, upper and lower limits, as well as the shape related statistics, kurtosis and skewness for the hourly time-series. We can see a mean and median price around 3 cents per kWh, with a considerable standard deviation, in the order of 40% of the mean value (thus translating the difficulty for the forecasting task), a minimum negative value of 2 cents per kWh was recorded and the shape metrics reveal an approximately symmetric and leptokurtic distribution. The load was modeled with an average/median value of around 100 kW and presents also an approximately symmetric distribution, however kurtosis reveals a more compact distribution, i.e., platykurtic distribution.

Table E.1: Descriptive statistics of the different datasets.

Variable	Mean	Median	St. Dev	Minimum	Maximum	Skewness	Kurtosis
Prices [€/kWh]	3.17×10^{-2}	3.07×10^{-2}	1.28×10^{-2}	-2.00×10^{-2}	0.10	0.43	4.57
Load [€/kWh]	97.9	97.8	16.0	59.0	140.2	0.13	2.30
Wind Speed [m/s]	11.8	10.7	5.38	0.44	39.8	1.18	4.81
Irradiance [W/m²]	127.2	0.54	242.5	0	1193.3	2.17	6.94
Temperature [°C]	9.07	7.96	9.00	−12.0	36.2	0.37	2.41

With respect to the wind speed profile, we can see that it presents a considerable standard deviation (in the order of 46% of the average value) with an adjusted average speed close to 11 m/s and the shape metrics reveal a leptokurtic and highly positively skewed distribution. The irradiance time-series further accentuates these shape characteristics. Furthermore, an interesting situation, where the standard deviation is higher than the mean and the median, has a considerably distance from the mean value of 127.2 W/m, revealing

a very widely distributed data set. Meanwhile, temperature statistics reveal a time-series with relatively low mean temperatures and an approximately symmetrical distribution. Finally, we can also perceive that during the summer season, conditions are substantially more favorable for non-dispatchable generation.

E.3 Hybrid system modeling

In this section, the modeling of the several sources and coupling devices that form the HES is presented in a way that sufficiently captures its core operational behavior, without delving into detailed modeling aspects—such as transient state dynamics — that would introduce unnecessary complexity and computational cost for the intended simulation purposes. These sources will supply the existing electrical load, and include the battery bank, the pumped-storage hydropower, wind turbine, PV generator, and the power converters. In addition, the grid-connected topology enables the interaction with the upstream grid, which is considered as a perfect power source or sink, i.e., electric power is purchased from upstream grid to supply existing loads and/or charge storage units, otherwise in surplus situations to export (sell) the exceeding power to the upstream grid. The available storage capacity will play a vital role, providing the ability to respond peak time loads and store energy surplus during off-peak periods.

E.3.1 Kinetic Battery Model

The integration of a battery pack into the HES is highly recommended since it will store the surplus energy and supply the power deficit for the different operation conditions. The literature offers a comprehensive range of models to describe the behavior of different types of batteries under different operating conditions. Among them, four major groups stand out: electrochemical, stochastic, electrical and analytical models. In this work, authors employed the powerful and widely used analytical model: Kinetic Battery Model (KiBaM) [663], the same model that is used by the computational tools Homer® and Hibrid2®.

KiBaM models the total capacity of a battery in two reservoirs, separated by a conductance, as illustrated in Figure E.2. The available charge reservoir contains a fraction c of the total (nominal) capacity qmax (expressed by q1), responsible for the immediate supply of energy to the connected load. The bound charge reservoir has a fraction $1 - c$ of the total capacity (expressed by q2) and is responsible for supplying energy exclusively to the available charge reservoir. The rate of charge flow between reservoirs depends on the conductance parameter k_{batt}, as well as the difference between the two reservoirs height $(h_1 - h_2)$, with $h_1 = q_1/c$ and $h_2 = q_2/(1 - c)$. After periods of battery discharge/charge the reservoirs tend to balance, i.e., $h_1 = h_2$.

The flows of available and bound charge during a constant current discharge or charge (I), are computed using the differential equations system, as shown in Equation E.3. A full model description, including the maximum allowable battery charge and discharge

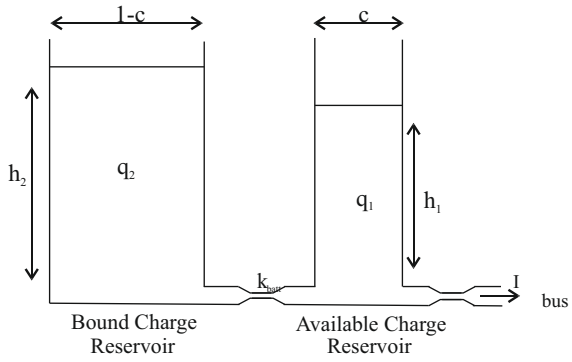


Figure E.2: The two-well kinetic battery model (KiBaM).

power equations, as well as the amount of available and bound energy at each time-step can be found in [50,51].

$$\begin{cases} \frac{dq_2}{dt} = k_{\text{batt}}\,(h_1 - h_2) \\ I = -\frac{dq_1}{dt} - \frac{dq_2}{dt} \end{cases} \quad \text{(E.3)}$$

At each time-step, the maximum amount of discharge/charge power is computed, as well as the available energy and bound energy and from these two we determine the total amount of energy (storage) $q(t) = q_1(t) + q_2(t)$, to then compute one of the most important battery parameters: its $SOC_{\text{batt}}\,(t) = q\,(t)/q_{\max}$.
In this work, specific model parameters were identified using the information provided by the manufacturers and are available in Table E.2. The battery bank has a capacity about 5 times larger than a low range electric vehicle battery and is, nowadays, easy to assemble with solutions like Tesla's© Powerwall and Powerpack (up to 210 kWh-AC) solutions.

Table E.2: KiBaM model: battery bank specifications.

Parameter	**Value**
Total Capacity (C)	131.3305 kWh
Charge Distribution Ratio (c)	0.6213
Rate Constant (k)	0.185

E.3.2 Pumped-Storage Hydropower

Hydropower generation with pumping is characterized by the presence of two reservoirs, one upstream and the other downstream. The basic premise behind this system is simple: in surplus off-peak periods, with a lower energy price, water is pumped upstream, then when the demand is higher (peak periods), and consequently the energy price is higher (more profitable), the stored water drives the turbines to produce energy. Given these characteristics, PSH has a very important load balancing role, since it allows a high penetration of intermittent sources.
As defined in the literature [664,665], the amount of power generated, P_{hydro} [W], and the

pumped flow rate, Q_{pump} [m^3s^{-1}], can be determined using Equations E.4 and E.5, respectively:

$$P_{\text{hydro}} = \mu_{\text{hydro}} \rho_{\text{water}} g h_{\text{net}} Q_{\text{turbine}}, \quad \text{(E.4)}$$

$$Q_{\text{pump}} = \mu_{\text{pump}}{}^{P_{\text{pump}}}/_{\rho_{\text{water}} g h_{\text{net}}} \quad \text{(E.5)}$$

where μ_{hydro} [%] and μ_{pump} [%] represent the turbine and pump energy conversion efficiencies, respectively; h_{net} [m] is the effective head; Q_{turbine} [m^3s^{-1}] is the turbined water flow; P_{pump} [W] is the pump power consumption; ρ_{water} is the water density, typically 1000 kgm^{-3} ($4\,^{\circ}\text{C}$); g is the gravitational acceleration, usually 9.81 ms^{-2}. Table E.3 displays all the remaining parameters associated with the hydropower turbine application.

Table E.3: PSH considered specifications.

Parameter	Value
h_{net}	25 m
$P_{\text{hydro installed}}$	150 kW
V_{total}	35,000 m^3

A Francis turbine is a good fit for energy demanding appliances. For this HES a mini hydropower application (<1 MW) is considered, feasible with low to medium head resources with relatively small dams, as Figure E.3 reveals [666]. China has exploited these mature small-scale ranges better than anyone else [667], proving the effectiveness of these technologies, leading the charts with around 200 GW of installed hydropower capacity.

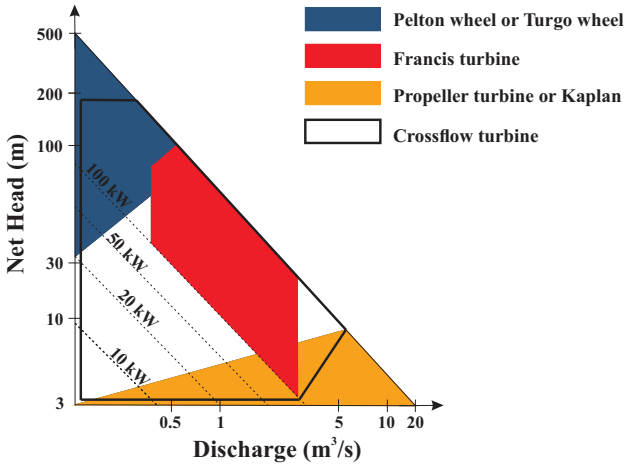


Figure E.3: Head-flow ranges of small hydro turbines.

Having selected a Francis turbine, and in order to replicate its operational conditions in the best manner, since the output rate depends heavily upon percent of actual flow to nominal flow, instead of using an approximate constant efficiency (such as the euro effi-

ciency standard), it was decided to do the modeling of the performance as a function of the turbine flow, as shown in [668]. For this purpose, a 5th degree polynomial approximation was made to the curve data points, with a R^2 = 0.9906, resulting in the following polynomial expression:

$$\begin{aligned}\mu_{\text{hydro}}(Q) = {} & 1.19 \times 10^{-10} Q^5 - 6.58 \times 10^{-8} Q^4 \\ & + 1.39 \times 10^{-5} Q^3 - 1.4 \times 10^{-3} Q^2 + 0.071\, Q - 0.637,\end{aligned} \tag{E.6}$$

Additionally, for the sake of simplification h_{net} [m] was considered constant despite the volume change in the reservoir, μ_{hydro} [%] and μ_{pump} [%] were considered equal. Finally, since PSH allows storage in an analogous form as the battery bank, the concept of SOC has also been attributed, and it is computed as the ratio between the current reservoir capacity and its maximum capacity allowance, i.e., $SOC_{\text{hydro}}(t) = V(t)/V_{\text{max}}$.

E.3.3 Wind Power

The wind energy conversion process is characterized by two specific curves associated with each wind turbine model. Power output of wind turbine generator at a specific site depends on wind speed at hub height and speed characteristics of the turbine, furthermore the wind speed needs to be constrained between a model's specific cut-in and cut-out speeds. One known as power curve relates the turbine output power with the wind speed, the other is the power efficiency curve relating the turbine output power with the available power in the wind (power coefficient).
In order to perform the wind speed mapping to turbine's output power, the authors used the approximated power curves provided by the manufacturers, as depicted in Figure E.4, despite the existence of linear, nonlinear and physical approximated models [669–671]. In this work, the adopted wind energy conversion system fell on a combination of Bergey BWC Excel-10 and Norwin 24-STALL-150 kW turbines, in order to a total an exact amount of 200 kW of installed capacity.

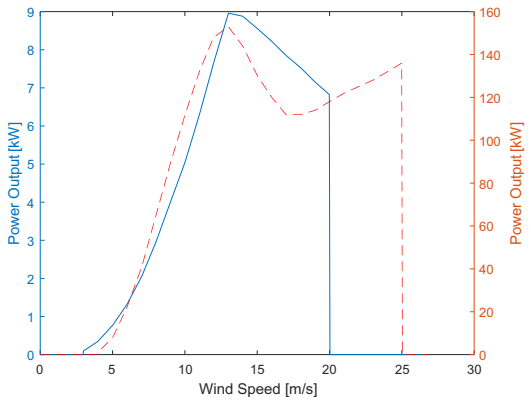


Figure E.4: Wind Turbines Power Curves: BWC Excel-10 (blue line) and Norwin 24-STALL-150 (dashed red line).

Additional details vis-à-vis important features can be seen in Table 4. The wind speed measured by the anemometer at a certain height needs to be corrected given the wind shear effect; to this end the power law profile was used, Equation E.7, with the power law exponent $\alpha = 1/7$:

$$\frac{u}{u_{\mathrm{r}}} = \left(\frac{z}{z_{\mathrm{r}}}\right)^{\alpha}, \quad \text{(E.7)}$$

where $u_{\mathrm{r}}\left[ms^{-1}\right]$ is the measured (available) wind speed at height $z_{\mathrm{r}}\,[m]$ and $u\left[ms^{-1}\right]$ is the wind speed at hub height $z\,[m]$ of the wind turbine.

Table E.4: Wind Turbine Modeling: Bergey and Norwin.

Parameter	**Value**	**Parameter**	**Value**
Rated Power (Norwin)	200 kW	Power Coefficient (C_p)	1/7
Hub Height (Norwin)	30 m	Hub Height (Bergey)	30 m
Rated Power (Bergey)	10 kW	Rated Power (Alt Bergey)	150 kW
Cut-in Speed (Norwin)	2.5 m/s	Cut-in Speed (Bergey)	4 m/s
Cut-out Speed (Norwin)	20 m/s	Cut-out Speed (Bergey)	25 m/s
# of Turbines (Norwin)	5	# of Turbines (Bergey)	1

E.3.4 Solar Photovoltaic

Solar PV systems generate electric power from solar energy through PV cells. As pointed out in 2, substantial penetration levels can be found both in conventional producers' portfolios, as well as in residential and business/industry appliances, seeking financial savings. To this end, its relatively portable and sizable structure, the increased technological maturity, hand-in-hand with targeted fiscal policies and other incentives, constituted unparalleled attractiveness factors. As such, it makes it one of the most consensual solutions for hybrid energy systems, with the disadvantage of only being able to operate for 8 h per day.

With this regard, a rich literature is available concerning the accurate behavior mapping of PV cells operating under different conditions. Among them we can highlight the commonly used single diode or two diode models, as well as three or more diodes and partial-shading models. However due to the computational effort they require, an alternative followed in HES modeling is to use a more synthesized model [604, 609]. PV cells are the main units of a PV generator, and are grouped to form panels or arrays.

Considering a PV generator formed by N_{S} modules connected in series, N_{P} connected in parallel, with a maximum power point $P_{\mathrm{MPP,STC}}$ [W] under STC conditions, i.e., the standard test condition temperature $T_{\mathrm{ref,\,STC}}$ and with an irradiance $G_{\mathrm{ref,\,STC}}$, with an efficiency rate μ_{MPP} [%] and a temperature coefficient of the open-circuit voltage $\alpha_{\mathrm{VOC}}\left[\mathrm{V\,^{\circ}C^{-1}}\right]$, then for a given solar irradiance G [$\mathrm{Wm^{-2}}$] the PV generator maximum power output P_{solar} [W] is given by Equation E.8 [672]:

$$P_{\text{solar}} = \mu_{\text{MPP}} \left(P_{\text{MPP,STC}} \frac{G}{G_{\text{ref, STC}}} \left(1 + \alpha_{\text{VOC}} \left(T_{\text{cell}} - T_{\text{ref, STC}}\right)\right) \right) N_{\text{S}} N_{\text{P}}, \quad \text{(E.8)}$$

where the cell temperature, Equation E.9 is a function of the ambient temperature T_{amb} [°C] and the nominal operating cell temperature NOCT [°C] under NOCT conditions i.e., with an air temperature $T_{\text{ref, NOCT}}$ and an irradiance $G_{\text{ref, NOCT}}$:

$$T_{\text{cell}}\,[^{\circ}\text{C}] = T_{\text{amb}} + \frac{G}{G_{\text{ref, NOCT}}} \left(\text{NOCT} - T_{\text{ref, NOCT}}\right), \quad \text{(E.9)}$$

It is also important to notice that PV systems are commonly coupled with a grid-connected inverter, that besides the DC/AC conversion task, employs an MPP-tracking algorithm to ensure that the maximum power is being extracted (its absence may account for losses up to 30%). The efficiency of these algorithms is nowadays well above 95%, and in this work a value of 99% was assumed, based on the Chint Power© CPS SC100KT datasheet. Additionally, to achieve a total of 75 kW, the V_{PM} voltage of each module, as well as the maximum input voltage (DC) of the grid-inverter, 880 V, for the CPS SC100KT, needs to be taken into account, therefore we end up with a configuration with $N_{\text{S}} = 28$ and $N_{\text{P}} = 9$. In this work, for simulation purposes, a HARP© NU-RD300S mono-crystalline silicon photovoltaic module, with an 18.3% conversion efficiency, was employed. The full PV panel specifications are presented Table E.5.

Table E.5: Specifications of the PV solar module.

Parameter	**Value**	**Parameter**	**Value**
Voltage Temp. Coefficient	-0.0029 V/°C	Nominal Cell Temp.	25 °C
Reference Temperature	25 °C	Irradiance (Test 1)	800 W/m^2
Irradiance (Test 2)	1000 W/m^2	Module Rated Power	300 W
Max Power Voltage	31.2 V	System Rated Power	75 kW
Operating Temp.	49 °C	Efficiency	99%
Module Efficiency	18.3%	Max System Voltage	880 V
Number of Modules	28	Number of Strings	9

E.3.5 Diesel Generator

Diesel generators, also known as diesel gensets are a stable and robust option to feed unserved loads, widely used as stand-alone application in isolated areas, especially in developing countries, where India stand outs [673], or in complementary forms, for instance coupled with PV systems and storage, reducing pollution, noise, the partial loads which reduces the already low efficiency of the generator and fuel consumption, which in addition, to direct savings, translates into indirect savings associated with costly diesel fuel transportation, especially in rural areas.

In this work, diesel generators are always seen as a backup option, given its weak thermal efficiency and its pollutant emissions, to ensure the HES supply safety. A linear approx-

imation, Equation (11), frequently used to approximate the amount of fuel the generator consumes, $F_{\text{consumption}}\left[\text{Lh}^{-1}\right]$ to generate electricity [670]:

$$F_{\text{consumption}} = F_0 \times Y_{\text{gen}} + F_1 \times P_{\text{gen}}, \quad \text{(E.10)}$$

where $F_0\left[\text{Lh}^{-1}\text{kW}^{-1}\right]$ stands for the fuel curve intercept coefficient, $Y_{\text{gen}}\,[\text{kW}]$ for the rated capacity of the generator, $F_1\left[\text{Lh}^{-1}\text{kW}^{-1}\right]$ is the fuel curve slope and $P_{\text{gen}}\,[\text{kW}]$ the generator power output. In this context, two diesel generators with a rated power, $Y_{\text{gen}} = 60\,\text{kW}$, more suitable for large energy deficits and one diesel generator with a rated power $Y_{\text{gen}} = 12\,\text{kW}$ for small deviations were integrated on the HES portfolio. To calculate the equation parameters, the approximate fuel consumption data at ¼, ½, ¾ and full load of real generators were used, as illustrated in Figure E.5, with the constant parameters defined as: $F_0 = 0.0481\text{Lh}^{-1}\text{kW}^{-1}$ and $F_1 = 0.2187\text{Lh}^{-1}\text{kW}^{-1}$ for the 60 kW generator, and $F_0 = 0.0532\text{Lh}^{-1}\text{kW}^{-1}$ and $F_1 = 0.2167\text{Lh}^{-1}\text{kW}^{-1}$ for the 12 kW generator.

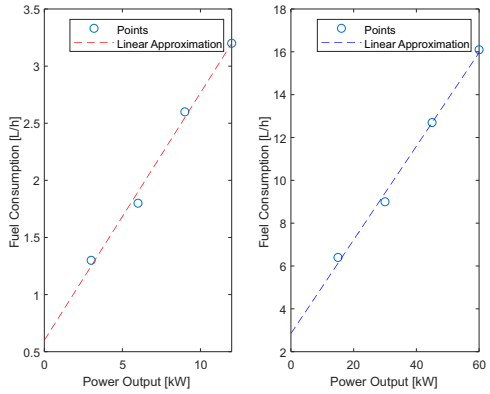


Figure E.5: Diesel Generator: fuel consumption curves as a function of power output.

As a side note, the use of rebated fuel for off-road applications is a common practice in most countries and is regulated within the European Union by Directive 92/81/EEC. Article 8(2) of this directive allows for reduced rates or exemptions, particularly for electricity production purposes. This type of fuel, often referred to as colored or dyed fuel, benefits from reduced taxation. For instance, in the UK, a VAT rate of 5% is applied to "red diesel," compared to the standard 20% VAT rate [674]. Consequently, for simulation purposes, a 10% discount was applied to end-user diesel prices — reflecting a prosumer benefiting from a special regime.

E.3.6 Power Electronic Converters

Besides energy generation sources and loads, power electronic converters are fundamental blocks in a HES. The possible configurations of the hybrid system are namely DC coupled, AC coupled and hybrid coupled system [605].

The AC/DC and DC/DC power converters were modeled according to their efficiency. The DC/AC electronic power converter was modeled based on [672], where the converter is characterized according to its efficiency as a function of the normalized rated power, $\frac{P_{\text{in}}}{P_{\text{rated}}}$, expressed by Equation E.11, therefore the only input parameter that characterizes the model is the converter rated power:

$$\eta_{\text{conv(DC/AC)}}\,[\%] = 1 - \frac{P_{\text{rated}}}{P_{\text{in}}}\left(0.0094 + 0.013\left(\frac{P_{\text{in}}}{P_{\text{rated}}}\right) + 0.04\left(\frac{P_{\text{in}}}{P_{\text{rated}}}\right)^2\right) \quad \text{(E.11)}$$

With regard to the DC/DC converter, the current high efficiency designs decrease the losses to the extent of achieving efficiencies close to 98/99%, as the bidirectional full SiC 200 kW DC/DC converter efficiency curve shows, therefore a constant value of $\eta_{\text{conv(DC/DC)}} = 98\%$ was considered reasonable.

E.4 A weekly results analys: costs, dispatch and STPF

To better understand the followed dispatch strategy, Figures E.6 and E.7 are presented to better highlight the proposed dispatch strategy, regarding a given weekly simulation, in this case for the chosen summer week. Figure E.6 illustrates the obtained generation mix profile for 168 h horizon and the cost functions for the hydro, storage bank and grid and, based on these, the cost-effective generation mix for this scenario was determined. As we can observe, this scenario presents an approximately constant PV generation and load profile, apart from the first day, where the almost null irradiance levels unveil the presence of a cloudy day. The high levels of wind generation can almost exclusively suppress the existing load for many hours, as such it allows to increase the SOC_2 and sales to the upstream grid. As intended, the battery bank has preferably an upper load balancing role within short periods of time. Moreover when the prices are expected to be low $\left(\hat{p}\,(k) - \frac{1}{24}\sum_{j=1}^{24}\hat{p}_j < 0\right)$ and $\varepsilon\,(k) \geq 0$, the strategy favors the energy buying, otherwise, when prices are higher $\left(\hat{p}\,(k) - \frac{1}{24}\sum_{j=1}^{24}\hat{p}_j \geq 0\right)$ and $\varepsilon\,(k) \geq 0$, PSH is dispatched. The plot shows the cost functions for the hydro, storage and grid and, based on these, the cost-effective generation mix for this scenario was determined.
The one day-ahead electricity prices forecast for this summer week (10–16 August) with an average daily MAPE of 5.49%., as well as the SOCs profile during the 168 h, can be seen in Figure E.7. Being evident that there is a surplus that allows the storage systems charging, and uncovers the intended tendency of the proposed strategy, in doing so in periods where low electricity prices are expected, as is the case with the abrupt SOC ramping around hour 80 h, in a clear contrast with what happens for instance in the first few hours of the simulation.

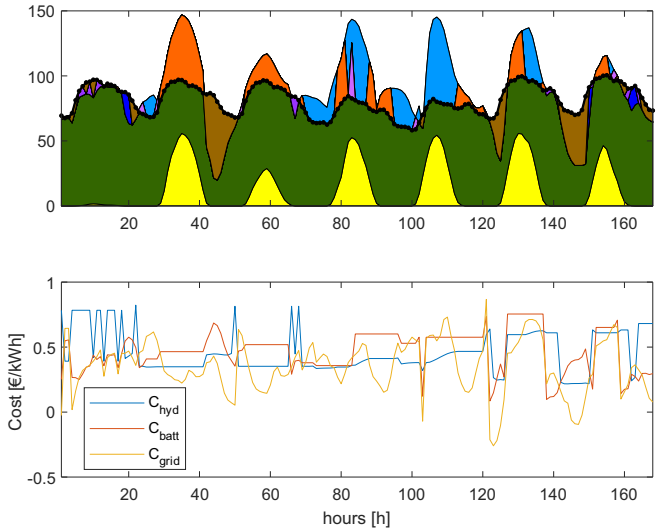


Figure E.6: Dispatched energy mix and its dispatch costs for the summer week. Where yellow, green and blue represent the PV, wind and hydro generation, respectively, light blue represents pumping periods, brown buying periods and orange selling periods. In turn, purple represents discharge periods, and lilac periods of charging the battery bank. Black represents the existing load level.

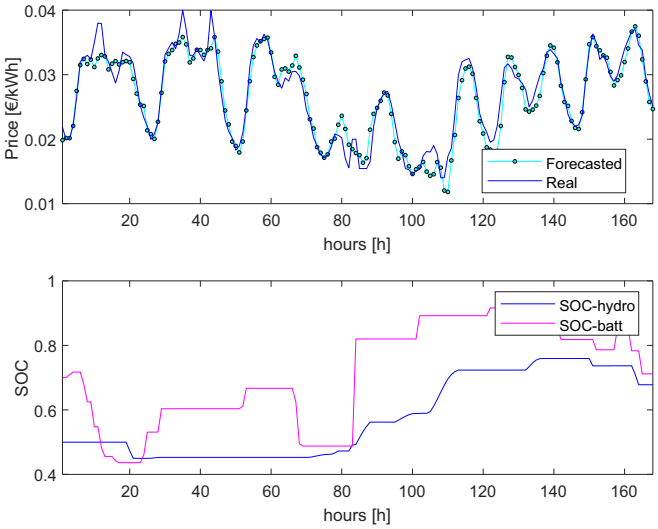


Figure E.7: Price Forecast vs. Real Prices and State of Charge evolution for the summer week.