Environmental Modelling and Software 147 (2022) 105234

Contents lists available at ScienceDirect

Environmental Modelling and Software

journal homepage: www.elsevier.com/locate/envsoft

SDSim: A generalized user friendly web ABM system to simulate spatiotemporal distribution of species under environmental scenarios

João Bioco [a,c,*], Fernando Cánovas [b], Paula Prata [a,c], Paulo Fazendeiro [a,c]

[a] *Cloud Computing Competence Centre (C4-UBI), Universidade da Beira Interior, Covilhã, Portugal*
[b] *Universidad Católica de Murcia, Murcia, Spain*
[c] *Instituto de Telecomunicações, Covilhã, Portugal*

ARTICLE INFO

Keywords:
Agent-based modelling and simulation
Distribution of species
Environmental modelling

ABSTRACT

This paper presents the Agent-Based Modelling System of spatial distribution of species *SDSim*. *SDSim* is an agent-based modelling system designed to simulate spatial distribution of species and populations for conservation and management purposes. *SDSim* gives to modellers the ability to simulate movements and colonization patterns of species given locations under study and a set of eco-geographical variables in which species depends on.

1. Introduction

Nowadays species distribution models (SDMs) are widely used by researchers in ecology and biology to perform studies for both conservation and management purposes. Ecological events can be described using these approaches, particularly shifting edge ranges observed under climate change scenarios (Edwards and Richardson, 2004; Hazen et al., 2013; Brown et al., 2016). In fact, recent climate changes have been considered one of the main causes affecting the geo-graphical distribution and persistence of species and populations (Heikkinen et al., 2006; Moore, 2003; Parmesan and Yohe, 2003; Walther et al., 2002; Kwon et al., 2015). Knowledge of species distribution becomes critical, even for economical reasons. Furthermore, public institutions and administrations are starting to grasp the benefits of resorting to SDMs for managing biological resources.

SDMs distinguishes the occurrence-environment relationships in order to project the distribution of species in different environment scenarios (past, present or future) (Meynard et al., 2019; Araújo and New, 2007).

The number of studies that are focused on SDMs, has increased considerably cf. (Guisan and Zimmermann, 2000; Barbet-Massin et al., 2018; Norberg et al., 2019). Normally, SDMs implicates the use of real distribution data of a species (occurrence data) in an environment described by a set of eco-geographical variables (EGVs), that calibrates the model to predict accurately future environment scenarios. Occurrence data are then used to validate and test the predictive power of the model. Generally, real data are unreliable to test SDMs, and in many cases it fails to give a full picture of the species' distribution (Meynard et al., 2019; Hirzel et al., 2001). To address these issues, virtual species are widely applied in SDMs (Miller, 2014).

Generating virtual species consists of simulating the distribution of a species knowing the occurrence-environment relationship. However, generating virtual species and SDMs are two different approaches since virtual species simulation involves exclusively the necessary steps to create virtual species (Meynard et al., 2019).

Alongside with the traditional modelling approach, agentbased modelling and simulation approaches (ABM) have been quite used in ecology, mainly due to their ability to simulate in a more realistic way the dynamics of different ecosystems resulting from the behavior and local interactions between individual entities (agents), e.g. (Jaxa-Rozen et al., 2019). Several studies have implemented ABM to build species distribution models capable of emulating the behavior and dynamics of real ecosystems across space and time (Manson et al., 2020.; Janssen et al., 2020). There are several solutions that implement ABM-based simulations to predict the spatial distribution of species and populations (e.g. (Pepin et al., 2017; Reuter et al., 2016; Xing et al., 2017; Parry et al., 2017)). However, most of them have a fundamental limitation because they simulate a particular species for which that ABM has been specifically designed, preventing for a more general use (Williams, 2021; Coakley et al., 2012; Collier et al., 2015). Furthermore, these frameworks require advanced programming skills in order to fit them properly (Grimm et al., 2006).

The approach presented in this paper goes one step further in a different direction in the sense that it allows the simulation of the species

---

* Corresponding author. Cloud Computing Competence Centre (C4-UBI), Universidade da Beira Interior, Covilhã, Portugal.
E-mail addresses: joao.bioco@ubi.pt (J. Bioco), fcanovas@ucam.edu (F. Cánovas), pprata@di.ubi.pt (P. Prata), fazendeiro@ubi.pt (P. Fazendeiro).

https://doi.org/10.1016/j.envsoft.2021.105234
Received 4 June 2021; Received in revised form 22 October 2021; Accepted 22 October 2021

geographical range in a continuum time, spreading over a predicted environment. As a result a novel generalized user friendly web ABM system is introduced in this study. The Species Distribution Simulator *SDSim* allows any modeller without any programming skills to model and simulate distribution of species and populations in real or potential environmental scenarios. *SDSim* was designed to study species distribution in an environmental landscape based on a set of parameters entered by the researcher.

In a nutshell, SDSim starts by enabling users to produce a SDM based on the species response functions to each of the EGVs that influence the distribution of a particular species. A Gaussian distribution function was chosen as the response function to each eco-geographical variable (EGV). Therefore, user provides as input data every EGV as raster maps, and for each EGV, a mean and a standard deviation defining a normal distribution hypothesized as optimal for the distribution of that species[1]. Response functions are combined (in an additive or multiplicative model), to create a SDM that represents environmental suitability for the species. This virtualization of the species distribution proceeds constrained by the environmental suitability but in close agreement with the species' life cycle algorithm. After a defined quantity of a virtual species is placed in random or selected *a priori* locations of a loaded landscape, a simulation can start. At each iteration, the virtual species will promote colonization of those locations deemed as suitable. Moreover, *SDSim* is able to monitor and analyse the evolution of the species spatiotemporal distribution in a landscape by using a visual component.

2. Related work

There are different software packages used to model the distribution of species. To the best of our knowledge, most of them are based on the R package (Muscarella et al., 2014; Golding et al., 2018; Kass et al., 2018; Naimi and Araújo, 2016; Thuiller et al., 2009), but there are also some offering specially tailored graphical user interfaces (GUI), e.g. (Brown et al., 2017). Usually, these packages project the past, current and future scenarios of the distribution of species. The methods applied to predict the distribution of species can be divided in two categories: machine learning and statistical methods. These methods are calibrated by a set of predictor variables and a sample of the known distribution of the species (presence-only or presence-absence data). In order to evaluate the prediction performance, these software implement some widely used performance measures such as area under the ROC curve (AUC), True Skill Statistics (TSS), sensitivity, specificity, and others. Predicting results are heavily influenced by the data quality (bias, poor data quality, etc). For this reason, several studies adopt the use of virtual species in order to have the full control of the relationship between the virtual species and the environment (EGVs) (Meynard et al., 2019).

There are several software packages developed to generate virtual species, e.g. (Duan et al., 2015; Leroy et al., 2016; Phillips et al., 2006; Qiao et al., 2016). Generally, these software packages receive as input data the environmental variables related to the species under study and the response functions that describe this relationship, and produce as the output the environment suitability (suitability map) for the species. This suitability map is then converted into potential presence/absence map of the species in that location. (Meynard et al., 2019). Different techniques are implemented to generate presence/absence map, see (Leroy et al., 2016). Normally, a sampling of the presence/absence points is saved to be used in a SDM as presence-only or presence-absence data.

Generally speaking, these software are bounded to predicting the areas in which a species may or may not occur. A biologically guided simulation process in which it is possible to observe how the species spreads spatially in the environment over time, following a life cycle, is usually absent. Therefore, obtaining information regarding where a species may or may not occur, may be insufficient, particularly for management, species conservation, and resource optimization. For example, based on the information regarding how a species of plants is able to swiftly occupy some locations, better transplanting strategies could be adopted to optimize the allocated resources.

The presented web-based software solution, *SDSim*, while sharing the common concerns of the available software packages for modeling the distribution of species, provides components that empower the user to visualize and analyse how species spreads spatially in the environment at each time interval.

3. Simulator highlights

SDSim is a software tool allowing users to monitor interactively the movement of a species across any real or putative environmental scenario introduced by a researcher. Modellers can easily analyse from the beginning of the simulation, how an ecological system is capable to spread spatially and behave. In the *SDSim* web application, landscape is defined by a set of EGVs encoded in raster maps that are made up as a matrix of pixels, also referred to as cells. Simulation outputs depend on a set of parameters that provides a flexible virtual framework to define a colonization pattern for virtually any species (Bioco et al., 2020): (i) an initial distribution of patches (set of cells) from which the simulation can start, (ii) the cell capacity (the maximum number of specimens allowed in the cell), (iii) three life cycle parameters to define species distribution arrangement (birth rate, death rate and spread rate), (iv) any of multiplicative or additive approaches to model habitat suitability, (v) type of Moore approach to simulate neighbourhood expansion pattern through a landscape, and (vi) a stopping criterion. Table 1 summarizes input data and parameters of the *SDSim* web application.

3.1. Habitat suitability function

Habitat suitability functions become crucial to provide a realistic simulation scenario. In general terms, species show habitat preferences. For example, in a location where environmental conditions are not suitable for a certain species, the probability of colonization and expansion should be lower. *SDSim* landscape can be characterized by a set of EGVs, relying on the researcher's knowledge about a species ecology, which ones could limit, or otherwise promote, a species distribution. From a species' perspective the overall suitability of the region is determined by the probability of occurrence of the species, given the

Table 1
Description of all input data/parameters of *SDSim*.

Input data/Parameters	Description
Species name	Name of a species.
Initial population	Initial distribution of patches (set of cells) generated randomly across a landscape.
Number of iterations (ticks)	Number of the simulated epochs.
Refresh Frequency	Frequency of generated output (e.g. every ten iterations).
Life cycle parameters	Birth rate or survival percentage. Death rate or extinction percentage. Spreading capacity rate or colonization percentage.
Type of model	Types of aggregation operators (multiplicative or additive) to model habitat suitability.
Type of neighbourhood	Kind of approach to simulate expansion pattern (Moore or Weighted Moore).
Eco-geographical variables (EGVs)	Set of independent EGVs to fit habitat suitability.
Mean (μ) and standard deviation (σ)	Standard parameters that define a normal distribution for each EGV. Mean values represent optimal conditions for that species.
Stopping criterion	Maximum number of iterations or stability criterion.

---

[1] Alternatively, if presence data are available the referred means and standard deviations can be automatically calculated, upon user request, using these real data. Whenever there is occurrence data available SDSim gives a glimpse of model performance by calculating the ROC curve and the corresponding AUC.

environmental conditions in that location (Elith and Leathwick, 2009). It is the result of the aggregation of local EGV values that influence its life cycle. In *SDSim* each EGV is characterized by a normal distribution around a hypothesized optimal value for that species. Mean and standard deviation should be provided for each variable according to the preferences of that species studied in the simulation. Alternatively, in *SDSim* these values can be automatically calculated by providing the coordinates of the occurrences of the species observed directly on the terrain.

As a result, habitat suitability would be internally calculated in a map, which values will be normalized in a closed interval from 0 (unsuitable) to 1 (optimal). These values are obtained through probability density functions (Parzen, 1962), that incorporate EGV in each map location as arguments. *SDSim* standardizes each EGV map (Andrews and Mallows, 1974): $x_{i'} = (xi - \mu)/\sigma$, where xi is the value of an EGV in that location, μ is the mean (i.e., EGV's optimal suitability value for a species) and σ represents the standard deviation for that EGV map. *SDSim* implements two different model aggregation operators that can be selected by the researcher in order to compute an overall suitability map for a given species: additive and multiplicative. Additive option is a straightforward implementation of a generalized additive model (GAM). Therefore, habitat suitability map is obtained by adding EGV values in each raster cell after applying a probability density function. Under multiplicative option, habitat suitability map is obtained by a strict Archimedean triangular norm. In this case, habitat suitability is more restrictive because product t-norm produces stronger conjunction of probabilistic values. Multiplicative model might become useful in particular circumstances (e.g., invasive species, colonization of clonal organisms).

3.2. *General workflow*

In order to perform a simulation, users should set all required simulation input data and parameters showed in Table 1. *SDSim* allows users to create both real and simulated scenarios, by uploading a different set of EGVs as raster maps for a defining mean and standard deviation in order to build a suitability function associated with each EGV for that species.

SDSim allows users to select a type of neighbourhood from any of Moore (1962) or weighted Moore (Lipowski and Lipowska, 2012) options. Assuming a regular grid of a raster map, by using Moore's neighbourhood, species will expand homogeneously to their eight neighbours. When using weighted Moore option, each one of the eight neighbouring cells will receive a number of transferred individuals directly proportional to its suitability. *SDSim* generates a suitability map based on EGV maps and every specific probability density function. The more suitable is an adjacent cell for a species, the more likely it will be chosen for range expansion of that species. Each cell would contain two values representing the capacity and the suitability (fitness) of the habitat. Initially, population patches are randomly placed in a landscape before starting a simulation. During a simulation, SDSim calculates the difference between previous and current states at each iteration by the sum of the cell-by-cell differences between successive states of the system, see Algorithm 1.

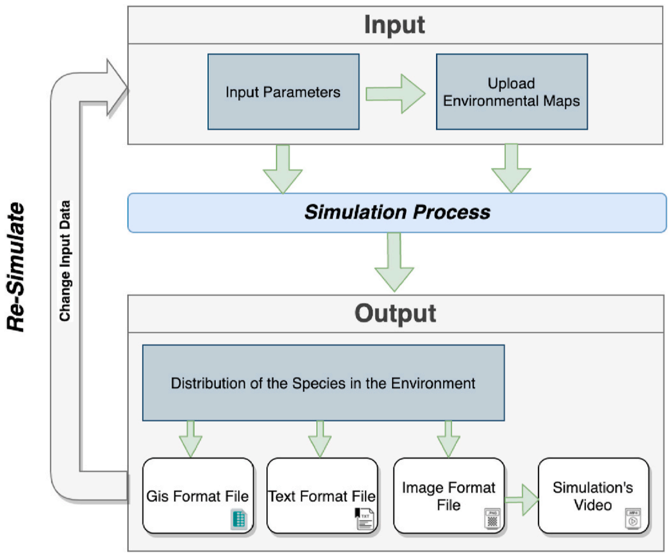


Fig. 1. SDSim General workflow, showing a standard procedure that should be followed by a user to perform a simulation in SDSim.

The simulation can be stopped when the system reaches a stable state or when a maximum number of iterations is completed. In this regard the system is said to be stable when the difference between two consecutive states of the simulation converges to zero.

SDSim saves an output of a simulation according to an interval previously defined by the user at each iteration or after a given number of epochs. At the end of a simulation, *SDSim* produces as output the species distribution map in three file formats: (i) text file, (ii) ASCII Grid file, (iii) image file; and a video file, see Fig. 1. For future reference the text file containing the values of all parameters used in the simulation is also available.

Algorithm 1. SDSim General Algorithm

Algorithm 1 *SDSim* General Algorithm

```
procedure SIMULATION
    Initialization of the general parameters, see Table 1
    EGV_data = readEcoGeographicalVariables(filepath)
    conuerted_EGV = convertEGVs(EGV_data, $\mu$, $\sigma$)
    suitability_map = generateSuitabilityMap(conuerted_EGV , probabilityDensityF unction)
    normalised_suitability_map = normalizeValues(suitability_map, model_type)
    patches = createPatches(normalised_suitability_map)
    Copy patches into preuious_patches
    generateInitialPopulation(preuious_patches,   agents_quantity)
    repeat
        for each patch P of patches do
            Find the neighbours of P
            LIFE_CYCLE(P , preuious_P , birth_rate, death_rate, spread_rate, P _neighbours)
        end for
        if mod(t, output_interual) == 0 then
            Create the species distribution file at time t
        end if
        Calculate cell-by-cell differences between patches and preuious_patches
        Copy patches into preuious_patches
    until stopping criterion
end procedure
```

Table 2
Description of the main functions implemented in the *SDSim*.

Function	Description
findNeighbours	Identifies neighbouring cells.
lifeCycle	Quantifies species capacity in each cell according to a birth rate, death rate and spread rate.
readEcoGeographicalVariables	Loads EGVs.
convertEGVs	Standardization of EGVs.
generateSuitabilityMap	Fits a normal distribution as response function for each EGV and then combine such response functions by adding or multiplying.
normalizeValues	Normalizes habitat suitability map into a [0,1] interval.
generateInitialPopulation	Generates a defined capacity for a species distributed in random patches.
createDistributionFile	Outputs species distribution every user-defined interval.

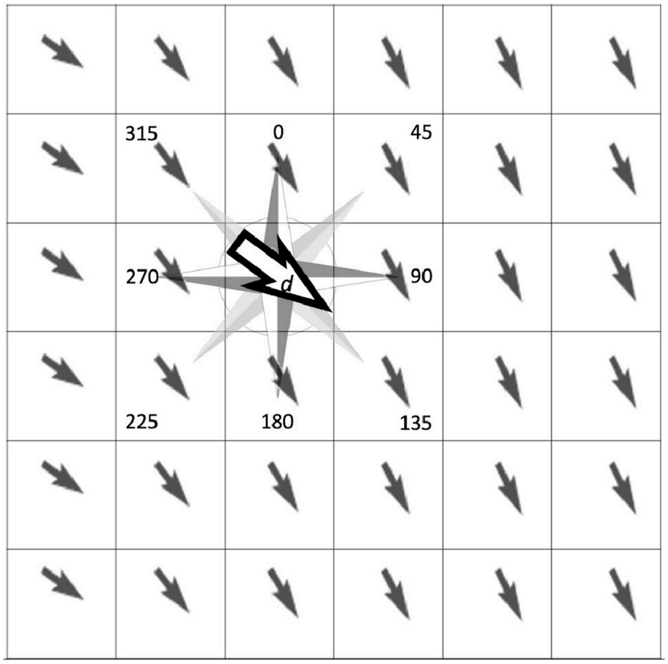


Fig. 2. Example of a dEGV where arrows represent direction in degrees clockwise from the geographic North. The highlighted cell has a direction *d*, see (1), and 8 neighbours. Each neighbour has a relative direction towards the cell ranging from 0^o to 315^o.

3.3. Species life cycle algorithm

Species life cycle algorithm includes three steps: reproduction, expansion/colonization, and death, see Algorithm 2. This algorithm tries to imitate a natural biological generation so that species may stay in habitats that are more appropriate for them to colonize, establish and finally expand.

During the reproduction phase, species spreads spatially according to a birth rate or survival percentage in each raster cell across a landscape. During the death phase, a percentage of a cell capacity would disappear according to a death rate or extinction percentage. During the expansion phase, a percentage of a cell capacity is transferred to their neighbour cells according to a spread rate or colonization percentage. An optimal strategy for a species in a particular and suitable landscape will ensure an optimal colonization of that particular setting. Table 2 summarizes the main functions implemented in *SDSim* ABM.

3.4. Direction constraints

Directional EGV (dEGV) can be used to promote or otherwise limit movement of a species in certain spatial directions. Typically a dEGV is characterized by two components that together represent a vector field magnitude (intensity) and direction. Each component should be represented by a raster map. As an example for marine data, currents should be split in magnitude or intensity, and direction. For intensity preferences of the species, the treatment is similar to the one previously described, i.e. given the intensity optimal parameters (mean and standard deviation) for a given species, the corresponding density value is an input to the model aggregation which determines the overall suitability map for the species. Direction component will be likewise integrated in the model once suitability is computed. Direction map should include values expressed in degrees clockwise from the geographic North. Thus each map cell has a direction value d, $0 \leq d < 360$, representing the flow or **directional movement**. Species are able to move from the core cell to any of its 8 neighbouring cells. Each neighbour has a **relative direction** towards the core cell, as depicted in Fig. 2.

In order to compute direction of each neighbour, *SDSim* computes the difference between the relative direction of the neighbour and the direction of movement for that cell:

$$\Delta_{\mathrm{k}} = \min\left(|\mathrm{d} - \mathrm{d_k}|, |\mathrm{d} + 360 - \mathrm{d_k}|\right), \tag{1}$$

where $k \in \{1, \ldots, 8\}$ stands for the neighbour cell and d_k is the relative direction of such neighbour.In Fig. 2 those directions are represented as values in each one of the neighbours of their core cell. Hence $\forall k \in \{1, \ldots, 8\}$: $0 \leq \Delta_k \leq 180$. Values are then normalized to the unit interval and incorporated into the model aggregation. Species expansion will be favoured by the overall direction information codified in every dEGV, i.e., most of the transferred individuals will migrate from a core cell to neighbouring ones that are well aligned with the direction of movement.

Let us return to the intensity component of a dEGV. Besides being treated as a mere EGV representing the preference of a species for a given value (e.g. some species might prefer waters with less current intensity) it worthwhile to notice that most times there is a physical facet attached to it, which should also have implications to the species fixation in a locus or to the speed of range expansion (e.g. some species travel easily if the current speed is higher). In such cases we propose the introduction of a *momentum* (species-dependent) constant, $m_S \in [0, 1]$, constraining the number of individuals that are transferred to neighbouring cells, N_{spread}, in the following way:

$$N_{spread} = N \times spR + I \times m_S, \tag{2}$$

where N is the cell's number of individuals, spR denotes the species' spread ratio as introduced before and $I \in [0, 1]$ is the normalized intensity value observed at the cell level.

Algorithm 2. Species life cycle

Algorithm 2 Species life cycle

```
procedure LIFE_CYCLE($self, birth\_rate, death\_rate, spread\_rate, neighbours$)
    # reproduction phase
    $self.quantity += self.quantity \times birth\_rate \times self.suitability$
    # death phase
    $self.quantity -= self.quantity \times death\_rate \times (2 - self.suitability)$
    # expansion phase
    $spread = self.quantity \times spread\_rate$
    $self.quantity -= spread$
    for $neighbour$ in $neighbours$ do
        $neighbour.quantity += spread \times \mathrm{suitability}(neighbour) / \mathrm{suitability}(neighbours)$
    end for
end procedure
```

Thus the intensity of a directional EGV constrains the spread of the species not only in what regards "preferences" of the species but also by applying the modelled traction forces. Notice that some species heavily depend on such variables to perform their spreading while others can expand its range even if they are positioned in a zero-intensity region, hence the need for the introduction of a momentum constant.

4. Web interface

SDSim is available online at https://sdsim.it.ubi.pt. It provides a user friendly Web interface that allows users to perform their simulations avoiding local installation. In order to get access to the application, users request an account by sending an email to the SDSim administrator, providing a username to associate that account. After an account is created, the user receives a notification with credentials to log in and access available services of *SDSim* application that are presented in the main screen (see Fig. 3).

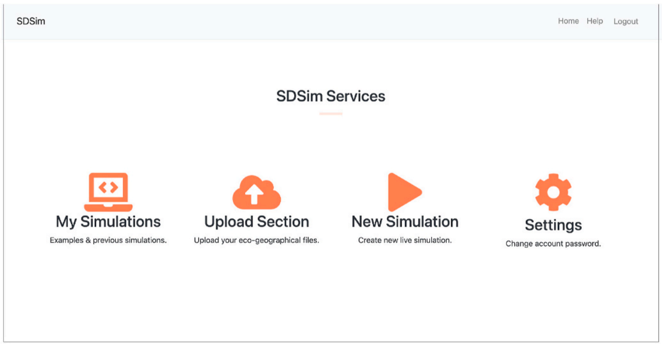
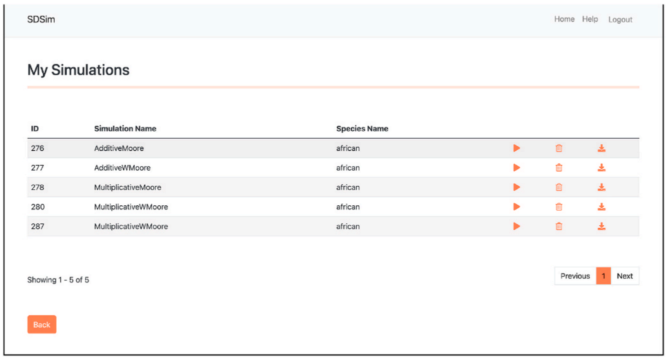


Fig. 3. *SDSim* available services through main screen.



My Simulations

ID	Simulation Name	Species Name			
276	AdditiveMoore	african	▶	🗑	⤓
277	AdditiveWMoore	african	▶	🗑	⤓
278	MultiplicativeMoore	african	▶	🗑	⤓
280	MultiplicativeWMoore	african	▶	🗑	⤓
287	MultiplicativeWMoore	african	▶	🗑	⤓

Showing 1 - 5 of 5

Fig. 4. Options included in section "My simulations".

In the section *"My Simulations"*, users can access to details from previous simulations, including simulation parameters and output data (simulation results). Users can download output data as set of raster maps containing the state of the simulation at different time steps, figures of the state of the simulation at each time step, the video file that shows the spatial distribution evolution, and a performance metric graph (receiver operating characteristic curve - ROC curve) if a species occurrences file was provided. The user can also perform new simulations based on previous ones or remove them (see Fig. 4). In this section, users can find an initial set of simulation examples to explore and get acquaintance with this system.

In the *"Upload Section"*, users are able to upload all necessary EGVs and dEGVs in form of raster maps in order to describe the landscape that is intended to simulate. *SDSim* does not use any specific datum or projection, relying on user needs to make such decisions. At this point, the Web application accepts only *ASCII Grid* raster format (please see GDAL library for raster and vector geospatial data formats at https://gdal.org/). In addition, the user can also upload a comma separated values (.csv) file containing the coordinates of the occurrences (and/or absences) of any sampled species. This file facilitates the user's work in such a way that he/she does not need to provide the mean and standard deviation for each EGV in order to estimate each probability density function. The *SDSim* performs all the necessary calculations to provide the mean and standard deviation for each EGV depending on the occurrences that the user has uploaded.

Presence/absence file can be uploaded to *SDSim*, allowing a numerical comparison between the results of different simulations based on the correct classification of presence/absence locations showing a ROC curve and the area under the curve (AUC) at the end of the simulation.

In the section *"Simulation"*, users can start a new simulation by filling a form with all the required parameters, as described in Table 1.

Users should follow all these steps after authentication to perform a simulation:

1. Access the section *"Upload"* and add EGVs (see Fig. 5).

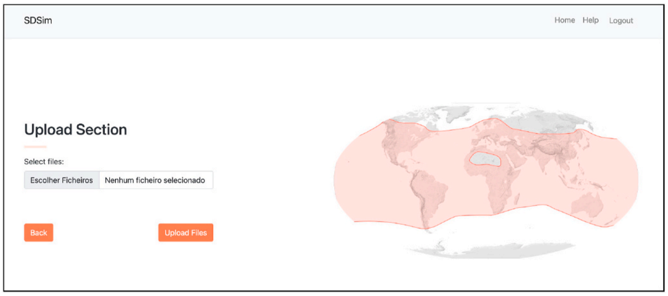
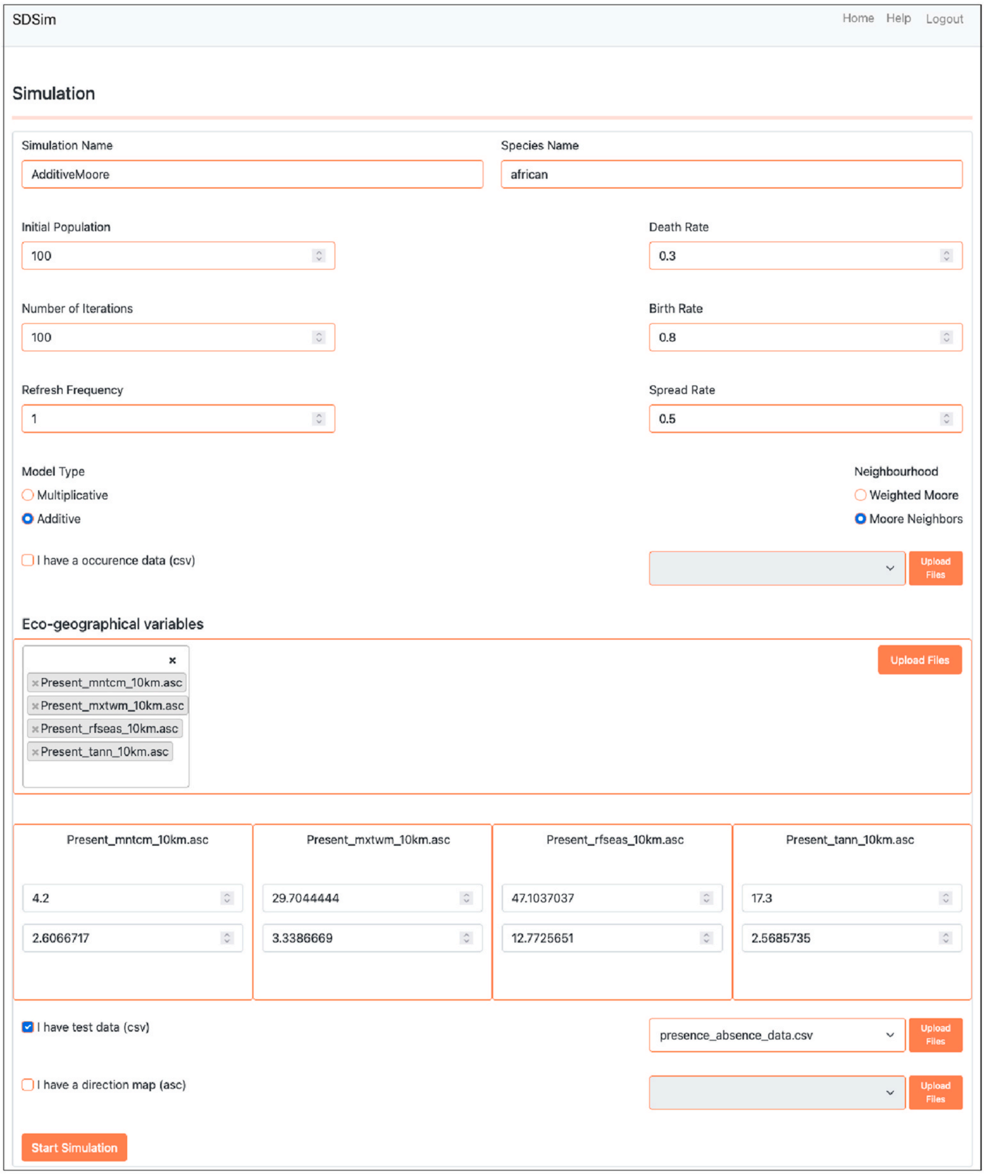


Fig. 5. *SDSim* "Upload" section.



Simulation

Simulation Name: AdditiveMoore
Species Name: african
Initial Population: 100
Death Rate: 0.3
Number of Iterations: 100
Birth Rate: 0.8
Refresh Frequency: 1
Spread Rate: 0.5

Model Type
○ Multiplicative
● Additive

Neighbourhood
○ Weighted Moore
● Moore Neighbors

☐ I have a occurence data (csv)

Eco-geographical variables
× Present_mntcm_10km.asc
× Present_mxtwm_10km.asc
× Present_rfseas_10km.asc
× Present_tann_10km.asc

Present_mntcm_10km.asc	Present_mxtwm_10km.asc	Present_rfseas_10km.asc	Present_tann_10km.asc
4.2	29.7044444	47.1037037	17.3
2.6066717	3.3386669	12.7725651	2.5685735

☑ I have test data (csv): presence_absence_data.csv
☐ I have a direction map (asc)

Fig. 6. Example form to introduce *SDSim* parameters.

2. Return to the main screen, and access the *Simulation* section where users should complete all the required parameters, including the selection of the corresponding EGVs and their parameters to estimate each probability density function (see Fig. 6).
3. Results (species distribution maps and outputs) can be managed in a gallery of images after the completion of the simulation (see Fig. 7). Users can also see the video that shows how the species' temporal distribution evolved (see Fig. 8).
4. A simulation can be saved (initial parameters and results) in order to reformulate any experiment and refine new ones based on previously stored simulations.

5. Case study

Honeybee *Apis mellifera* L. is distributed in a wide area. Populations of the Iberian Peninsula are of particular interest because of the hypothesized hybrid status of *A. m. iberiensis* (Engel, 1999). In agreement with this hypothesis, *A. m. iberiensis* distribution reflects colonization movements associated to an environmental gradient, probably showing selective factors of climatic tolerance (Cánovas et al., 2014).

In order to compare results, both against previous simulation studies (Bioco et al., 2018) and against real experimental data collected on the field (Cánovas et al., 2014), the spatial distribution of the African lineage of the Iberian honeybee *A. m. iberiensis* in the Iberian Peninsula was modelled and simulated in *SDSim*. Landscape was described by four main EGVs that were previously found to have an important role to shape present spatial distribution of this species: maximum temperature of the warmest month (*mxtrm*), minimum temperature of the coldest month (*mntcm*), rainfall seasonality (*rfseas*) and average annual temperature (*tann*) (Cánovas et al., 2014). EGVs were obtained from the climatic atlas of the World (Hijmans et al., 2005). Raster maps had a dimension of 196 × 122 cells.

To illustrate the range of freedom that is provided to the user, we report four simulations covering the different combinations between the available types of neighbourhood (Moore and Weighted Moore) and model aggregation (multiplicative and additive). The overall input parameters were the same for all four simulations, see Table 3.

Simulation results after 100 iterations are showed in Fig. 9. The left column of the figure presents the results for the Moore neighbourhood (the amount of transferred individuals from each cell is equally

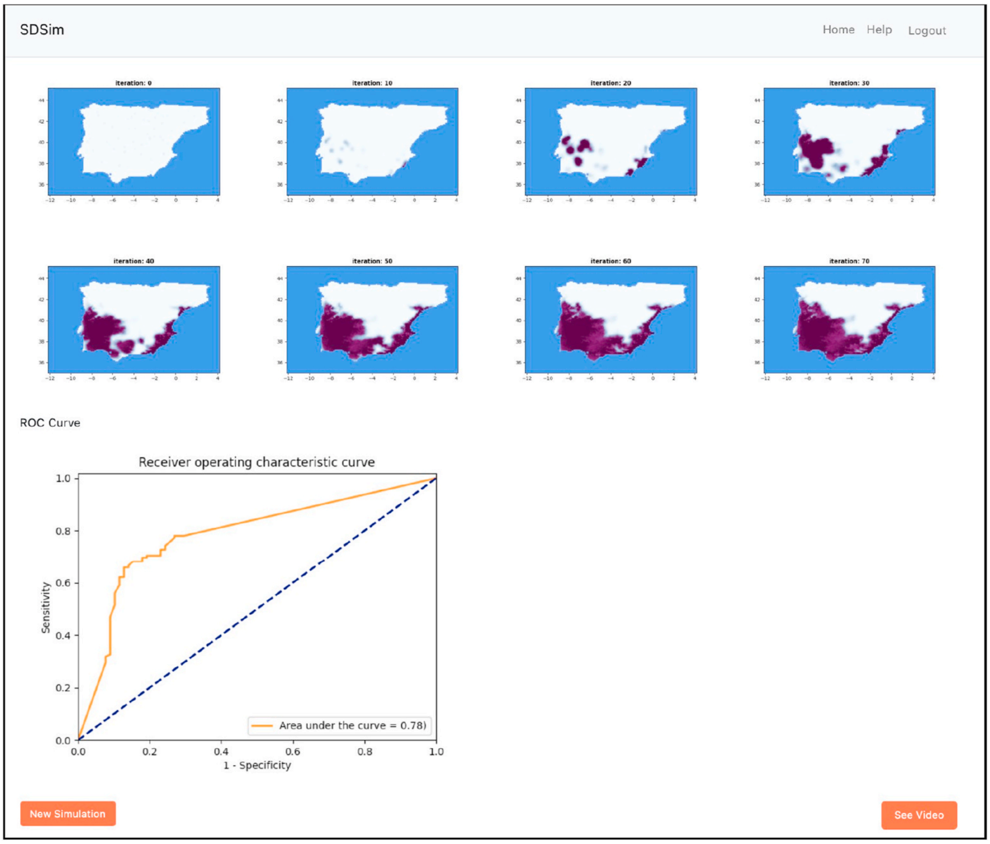


Fig. 7. Simulation results are shown in form of images that later can be exported as raster maps.

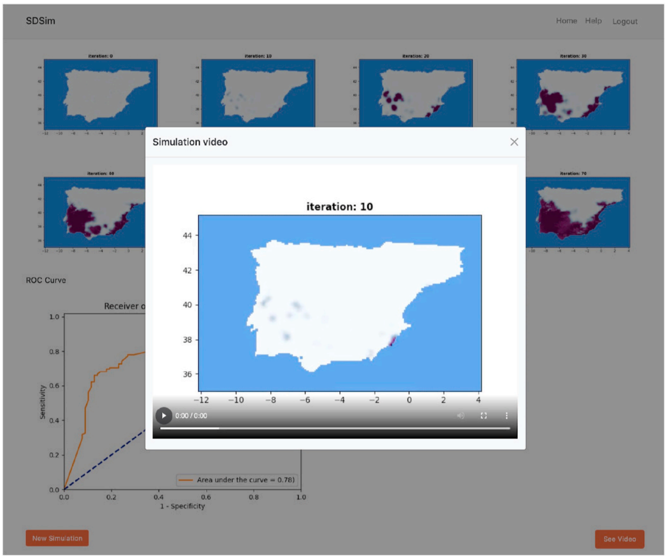


Fig. 8. Videos enable the visualization of the evolution of a simulation.

Table 3
Values of the parameters of the case study.

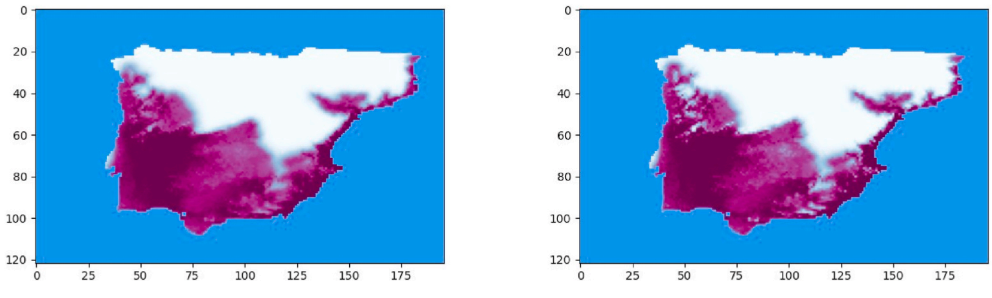
Input data/Parameters	Values
Species name	African-bees
Initial population	100
Life cycle parameters	Birth rate (0.8) Death rate (0.3) Spreading capacity (0.5)
EGVs	Maximum temperature of the warmest month (mxtwm); Rainfall seasonality (rfseas) Average annual temperature (tann); Minimum temperature of the coldest month (mntcm)
Mean (μ) $\pm$ *standard deviation* (σ)	29.7 $\pm$ 3.3; 47.1 $\pm$ 12.8; 17.3 $\pm$ 2.5; 4.2 $\pm$ 2.6
Stopping criterion	Number of iterations (100)

distributed by the eight neighbouring cells). In Fig. 9(a) the type of aggregation is additive, i.e. as a result we get generalized additive model for the four considered EGVs. In Fig. 9(c) the aggregation is multiplicative, thus heavily penalizing poor values in any of the EGVs. The column on the right presents the results for the same variation in the model aggregation operator but for the case of the weighted Moore neighbourhood in this case the neighbouring cells receive a quantity of transferred individuals directly proportional to their suitability, i.e., neighbours with better suitability receive a greater quantity of species than the less suitable ones.

In order to evaluate the performance of the model we compared the simulation results against the 210 field observations (species' presence and absence csv file containing 210 locations latitude and longitude). In Fig. 10 we show a ROC curve and AUC values for the types of neighbourhood and model aggregation. The additive model seems to perform better than the multiplicative model, and in the additive model both Moore and weighted Moore neighbourhood showed similar final results ($AUC = 0.78$).

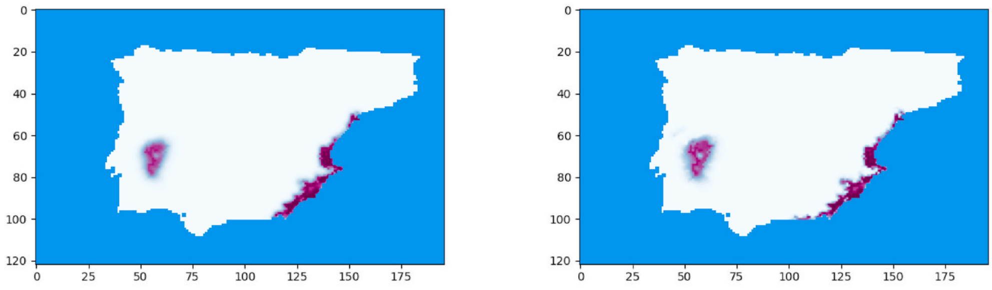
Considering the additive model in Fig. 11, detail differences can be followed as the species spreads spatially using Moore neighbourhood (on the left) and weighted Moore neighbourhood (on the right). These differences are more evident at these early stages of the simulation and are diluted once stabilization is reached. In Moore neighbourhood, a more regular deployment is able to produce almost circular regions, whereas weighted Moore neighbourhood made colonization more irregular, advancing faster in the track for more suitable regions.

Overall, the weighted Moore method seems to be a good choice for



(a) Additive + Moore

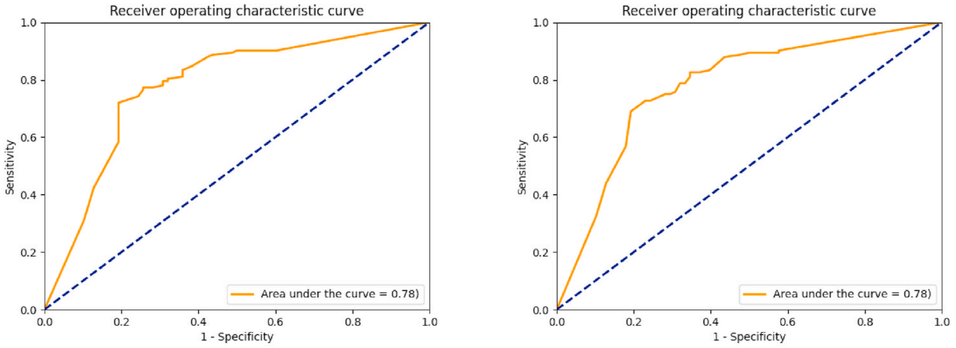
(b) Additive + weighted Moore



(c) Multiplicative + Moore

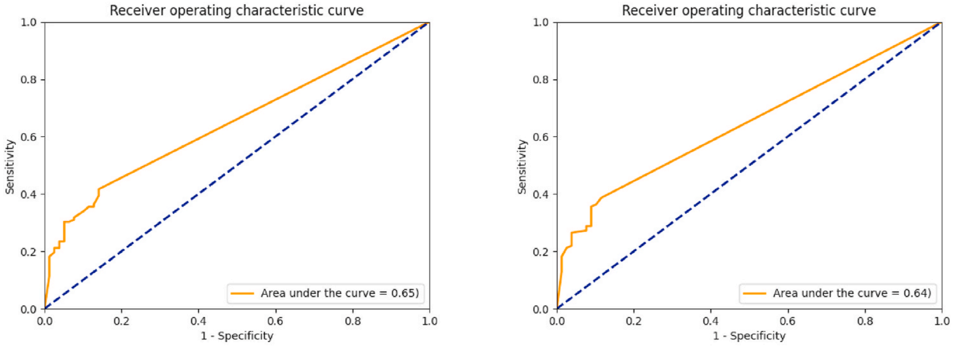
(d) Multiplicative + weighted Moore

Fig. 9. Evaluation of results from aggregation and neighbourhood operators. These snapshots were obtained after 100 iterations, with an initial population of 100 individuals (apiaries uniformly distributed over the Iberian Peninsula) and using the same birth, death and spread rates. Figures (a) and (b) were obtained with the additive aggregation operator while (c) and (d) are the corresponding ones for the multiplicative operator. Figures (b) and (d) were obtained with weighted Moore neighbourhood whereas figures (a) and (c) used the usual Moore neighbourhood.



(a) Additive + Moore

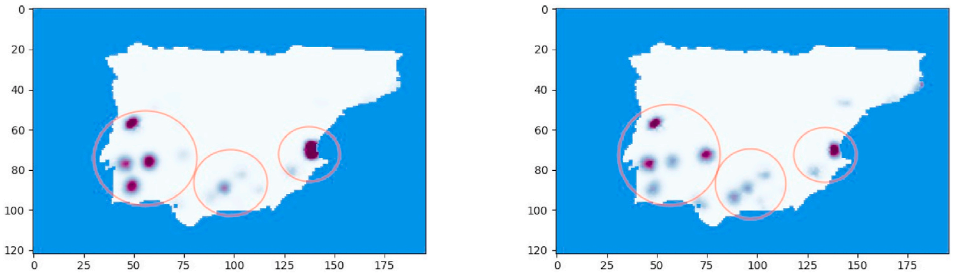
(b) Additive + weighted Moore



(c) Multiplicative + Moore

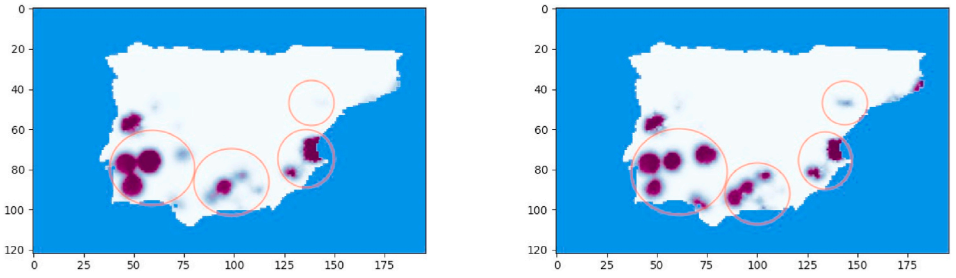
(d) Multiplicative + weighted Moore

Fig. 10. ROC curve and the AUC values for the types of neighbourhood and model aggregation.



(a) Additive + Moore (iteration 14)

(b) Additive + weighted Moore (iteration 14)



(c) Additive + Moore (iteration 35)

(d) Additive + weighted Moore (iteration 35)

Fig. 11. Differences between the evolution of the simulation with Moore neighbourhood (on the left) and the evolution with weighted Moore neighbourhood (on the right).

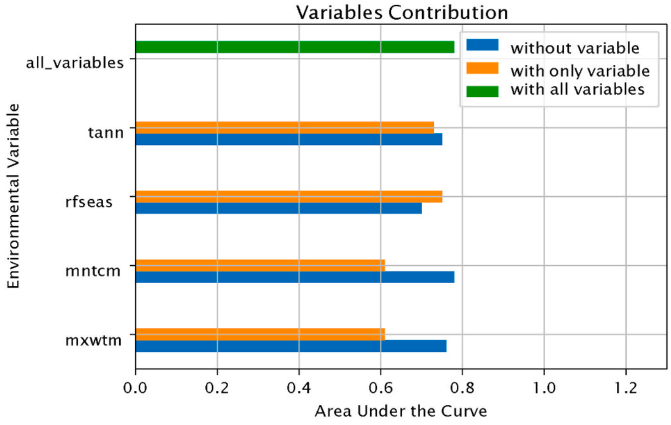


Fig. 12. Variables contribution.

this species, because it seems closer to an expected biological behavior such as the observed in previous studies (Cánovas et al., 2014). Species colonized with greater abundance in the most suitable areas, defined by a set of EGVs. Moreover, environmental suitability calculated using an additive model seems to be more appropriate according to what was observed for this species. However, choices regarding type of neighbourhood and aggregation strongly rely on the specific knowledge that users have about the behavior of each particular system. Furthermore, the great variability on the distribution that is possible to obtain for any given species (as is clearly exemplified in Fig. 9), far from being considered a drawback, is seen as a mean to easily test and evaluate different hypothesis on the species' mechanism of colonization.

Results of this case study are available to all users registered in the application, allowing access, exploration and reproduction of experiments available in the *SDSim* website.

The impact of each EGV was analysed in the simulation results, fixing the model (additive) and the neighbourhood (weighted Moore), and performing eight EGV combinations as follow. First we perform simulations with each EGV individually, and then removing each EGV, one at a time, from the complete set giving rise to the following combinations: *(mntcm, rfseas, tann)*, *(mxtwm, rfseas, tann)*, *(mxtwm, mntcm, tann)* and *(mxtwm, mntcm, rfseas)*. This comparison is summarised in Fig. 12 which depicts the AUC values for each experiment.

It is possible to notice that rainfall seasonality (*rfseas*) had the greatest impact in the overall results of the simulation, followed by the average annual temperature (*tann*). The minimum temperature of the coldest month (*mntcm*) and the maximum temperature of the warmest month (*mxtrm*) showed the least impact since, according to Fig. 12, if one of these variables is removed the quality degradation after stabilization is not significant. This result seems to indicate that the studied species can be very robust when faced with one-off sudden variations on the temperature as long as the average annual temperature stays within an admissible range.

6. Conclusion

In this paper, an ABM system was designed to model and simulate spatial distribution of biological species and populations by using a set of EGVs. *SDSim* computes a habitat suitability index, producing an internal species' suitability map. *SDSim* assumes a standard normal distribution as a probabilistic response function for each EGV. Those are later aggregated using an additive or a multiplicative model, chosen at users' will. Together with EGVs, *SDSim* receives a set of input parameters which fit ABM behavior and can be easily tuned. Therefore, any species could be modelled by using this ABM in different environmental scenarios. A simulation will always start by generating an initial occupation of the available habitat in random locations. At each iteration of the simulation, species will then find more suitable locations for them to survive, following a defined species life cycle with corresponding parameters. *SDSim* allows users to study and understand the species-environment relationship by changing the values of the environmental variables in order to verify the behavior of any simulation scenario. As illustrated in the reported case study, if in possession of real experimental data collected on the field, the Web interface provides supporting tools for assessing and compare a set of distinct hypothesis subsumed by the different experimental setups. Moreover, the flexibility and user friendliness of the deployed Web ABM system enables the visual analysis of the evolution of species in any hypothetical landscape without requiring any programming skills from the user.

SDSim will incorporate new features. The relationship between geological time and computational epochs will be useful for modellers to decide simulation time according to evolutionary biological time steps instead of stabilization of the simulation. Such ability will give extra tools to modellers who will be able to program simulations in evolutionary time rather than in computational one.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by Centro-01-0145-FEDER- 000019 - C4 - Centro de Competências em Cloud Computing, cofinanced by the European Regional Development Fund (ERDF) through the Programa Operacional Regional do Centro (Centro 2020), in the scope of the Sistema de Apoio à Investigação Científica e Tecnológica - Programas Integrados de IC&DT. This work was also funded by FCT/MCTES through national funds and when applicable co-funded EU funds under the project UIDB/50008/2020.

Authors thank the Department of Zoology and Anthropology from the University of Murcia in which the hypotheses proposed here with honeybees, were generated and tested in previously published scientific manuscripts, therefore providing a suitable case study to test SDSim.

The authors wish to thank Pedro Violante's contribution to the development of the front-end. Lastly, but not least, the authors gratefully acknowledge an anonymous reviewer for his/her thoroughly revision and insightful remarks.

References

Andrews, D.F., Mallows, C.L., 1974. Scale mixtures of normal distributions. J. Roy. Stat. Soc. B 36, 99–102.

Araújo, M.B., New, M., 2007. Ensemble forecasting of species distributions. Trends Ecol. Evol. 22, 42–47.

Barbet-Massin, M., Rome, Q., Villemant, C., Courchamp, F., 2018. Can species distribution models really predict the expansion of invasive species? PLoS One 13, e0193085.

Bioco, J., Silva, J., Canovas, F., Fazendeiro, P., 2018. A Cellular Automata Model of Spatio-Temporal Distribution of Species. In: International Conference on Advanced Intelligent Systems for Sustainable Development. Springer, pp. 118–128.

Bioco, J., Fazendeiro, P., Cánovas, F., Prata, P., 2020. Parameterization of an Agent-Based Model of Spatial Distribution of Species. In: Miraz, M.H., Excell, P.S., Ware, A., Soomro, S., Ali, M. (Eds.), Emerging Technologies in Computing. Springer International Publishing, Cham, pp. 251–260.

Brown, C.J., O'Connor, M.I., Poloczanska, E.S., Schoeman, D.S., Buckley, L.B., Burrows, M.T., Duarte, C.M., Halpern, B.S., Pandolfi, J.M., Parmesan, C., et al., 2016. Ecological and methodological drivers of species' distribution and phenology responses to climate change. Global Change Biol. 22, 1548–1560.

Brown, J.L., Bennett, J.R., French, C.M., 2017. Sdmtoolbox 2.0: the next generation python-based gis toolkit for landscape genetic, biogeographic and species distribution model analyses. PeerJ 5, e4095.

Cánovas, F., De la Rúa, P., Serrano, J., Galián, J., 2014. Analysis of a contact area between two distinct evolutionary honeybee units: an ecological perspective. J. Insect Conserv. 18, 927–937.

Coakley, S., Gheorghe, M., Holcombe, M., Chin, S., Worth, D., Greenough, C., 2012. Exploitation of High Performance Computing in the Flame Agent-Based Simulation Framework. In: 2012 IEEE 14th International Conference on High Performance Computing and Communication & 2012 IEEE 9th International Conference on Embedded Software and Systems. IEEE, pp. 538–545. IEEE.

Collier, N., Ozik, J., Macal, C.M., 2015. Large-scale agent-based modeling with repast hpc: a case study in parallelizing an agent-based model. In: European Conference on Parallel Processing, pp. 454–465. Springer.

Duan, R.Y., Kong, X.Q., Huang, M.Y., Wu, G.L., Wang, Z.G., 2015. Sdmvspecies: a software for creating virtual species for species distribution modelling. Ecography 38, 108–110.

Edwards, M., Richardson, A.J., 2004. Impact of climate change on marine pelagic phenology and trophic mismatch. Nature 430, 881–884.

Elith, J., Leathwick, J.R., 2009. Species distribution models: ecological explanation and prediction across space and time. Annu. Rev. Ecol. Evol. Syst. 40, 677–697. https://doi.org/10.1146/annurev.ecolsys.110308.120159.

Engel, M.S., 1999. The taxonomy of recent and fossil honey bees (Hymenoptera: Apidae; apis). J. Hymenoptera Res. 8, 165–196.

Golding, N., August, T.A., Lucas, T.C., Gavaghan, D.J., van Loon, E.E., McInerny, G., 2018. The Zoon R Package for Reproducible and Shareable.

Grimm, V., Berger, U., Bastiansen, F., Eliassen, S., Ginot, V., Giske, J., Goss-Custard, J., Grand, T., Heinz, S.K., Huse, G., Huth, A., Jepsen, J.U., Jørgensen, C., Mooij, W.M., Müller, B., Pe'er, G., Piou, C., Railsback, S.F., Robbins, A.M., Robbins, M.M., Rossmanith, E., Rüger, N., Strand, E., Souissi, S., Stillman, R.A., Vabø, R., Visser, U., DeAngelis, D.L., 2006. A standard protocol for describing individual-based and agent-based models. Ecol. Model. 198, 115–126. https://doi.org/10.1016/j.ecolmodel.2006. 04.023. https://www.sciencedirect.com/science/article/pii/S0304380006002043.

Guisan, A., Zimmermann, N.E., 2000. Predictive habitat distribution models in ecology. Ecol. Model. 135, 147–186. https://doi.org/10.1016/S0304-3800(00)00354-9. https://www. sciencedirect.com/science/article/pii/S0304380000003549.

Hazen, E.L., Jorgensen, S., Rykaczewski, R.R., Bograd, S.J., Foley, D.G., Jonsen, I.D., Shaffer, S.A., Dunne, J.P., Costa, D.P., Crowder, L.B., et al., 2013. Predicted habitat shifts of pacific top predators in a changing climate. Nat. Clim. Change 3, 234–238.

Heikkinen, R.K., Luoto, M., Araújo, M.B., Virkkala, R., Thuiller, W., Sykes, M.T., 2006. Methods and uncertainties in bioclimatic envelope modelling under climate change. Prog. Phys. Geogr. 30, 751–777.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A., 2005. Very high resolution interpolated climate surfaces for global land areas. Int. J. Climatol.: A Journal of the Royal Meteorological Society 25, 1965–1978.

Hirzel, A.H., Helfer, V., Metral, F., 2001. Assessing habitat-suitability models with a virtual species. Ecol. Model. 145, 111–121.

Janssen, M.A., Pritchard, C., Lee, A., 2020. On code sharing and model documentation of published individual and agent-based models. Environ. Model. Software 134, 104873.

Jaxa-Rozen, M., Kwakkel, J.H., Bloemendal, M., 2019. A coupled simulation architecture for agent-based/geohydrological modelling with netlogo and modflow. Environ. Model. Software 115, 19–37.

Kass, J.M., Vilela, B., Aiello-Lammens, M.E., Muscarella, R., Merow, C., Anderson, R.P., 2018. Wallace: a flexible platform for reproducible modeling of species niches and distributions built for community expansion. Methods in Ecology and Evolution 9, 1151–1156.

Kwon, Y.S., Bae, M.J., Hwang, S.J., Kim, S.H., Park, Y.S., 2015. Predicting potential impacts of climate change on freshwater fish in korea. Ecol. Inf. 29, 156–165.

Leroy, B., Meynard, C.N., Bellard, C., Courchamp, F., 2016. virtualspecies, an r package to generate virtual species distributions. Ecography 39, 599–607.

Lipowski, A., Lipowska, D., 2012. Roulette-wheel selection via stochastic acceptance. Phys. Stat. Mech. Appl. 391, 2193–2196.

Manson, S., An, L., Clarke, K.C., Heppenstall, A., Koch, J., Krzyzanowski, B., Morgan, F., O'Sullivan, D., Runck, B.C., Shook, E., Tesfatsion, L., 2020. Methodological issues of spatial agent-based models. J. Artif. Soc. Soc. Simulat. 23, 3. https://doi.org/10.18564/jasss.4174. https://par.nsf.gov/biblio/10172443.

Meynard, C.N., Leroy, B., Kaplan, D.M., 2019. Testing methods in species distribution modelling using virtual species: what have we learnt and what are we missing? Ecography 42, 2021–2036.

Miller, J.A., 2014. Virtual species distribution models: using simulated data to evaluate aspects of model performance. Prog. Phys. Geogr.: Earth Environ. 38, 117–128. https://doi.org/10.1177/0309133314521448 arXiv: https://doi.org/10.1177/0309133314521448.

Moore, E.F., 1962. Machine Models of Self-Reproduction. In: Proceedings of Symposia in Applied Mathematics. American Mathematical Society, New York, pp. 17–33.

Moore, P.D., 2003. Back to the future: biogeographical responses to climate change. Prog. Phys. Geogr. 27, 122–129.

Muscarella, R., Galante, P.J., Soley-Guardia, M., Boria, R.A., Kass, J.M., Uriarte, M., Anderson, R.P., 2014. Enm eval: an r package for conducting spatially independent evaluations and estimating optimal model complexity for maxent ecological niche models. Methods in ecology and evolution 5, 1198–1205.

Naimi, B., Araújo, M.B., 2016. sdm: a reproducible and extensible r platform for species distribution modelling. Ecography 39, 368–375.

Norberg, A., Abrego, N., Blanchet, F.G., Adler, F.R., Anderson, B.J., Anttila, J., Araújo, M.B., Dallas, T., Dunson, D., Elith, J., et al., 2019. A comprehensive evaluation of predictive performance of 33 species distribution models at species and community levels. Ecol. Monogr. 89, e01370.

Parmesan, C., Yohe, G., 2003. A globally coherent fingerprint of climate change impacts across natural systems. Nature 421, 37–42.

Parry, H., Paull, C., Zalucki, M., Ives, A., Hulthen, A., Schellhorn, N., 2017. Estimating the landscape distribution of eggs by helicoverpa spp., with implications for bt resistance management. Ecol. Model. 365, 129–140.

Parzen, E., 1962. On estimation of a probability density function and mode. Ann. Math. Stat. 33, 1065–1076.

Pepin, K.M., Davis, A.J., VerCauteren, K.C., 2017. Efficiency of different spatial and temporal strategies for reducing vertebrate pest populations. Ecol. Model. 365, 106–118.

Phillips, S.J., Anderson, R.P., Schapire, R.E., 2006. Maximum entropy modeling of species geographic distributions. Ecol. Model. 190, 231–259. https://doi.org/10.1016/j.ecolmodel.2005. 03.026. https://www.sciencedirect.com/science/article/pii/S030438000500267X.

Qiao, H., Peterson, A.T., Campbell, L.P., Soberón, J., Ji, L., Escobar, L.E., 2016. Nichea: creating virtual species and ecological niches in multivariate environmental scenarios. Ecography 39, 805–813.

Reuter, H., Kruse, M., Rovellini, A., Breckling, B., 2016. Evolutionary trends in fish schools in heterogeneous environments. Ecol. Model. 326, 23–35.

Thuiller, W., Lafourcade, B., Engler, R., Araújo, M.B., 2009. Biomod–a platform for ensemble forecasting of species distributions. Ecography 32, 369–373.

Walther, G.R., Post, E., Convey, P., Menzel, A., Parmesan, C., Beebee, T.J., Fromentin, J. M., Hoegh-Guldberg, O., Bairlein, F., 2002. Ecological responses to recent climate change. Nature 416, 389–395.

Williams, R.A., 2021. User experiences using flame: a case study modelling conflict in large enterprise system implementations. Simulat. Model. Pract. Theor. 106, 102196.

Xing, L., Zhang, C., Chen, Y., Shin, Y.J., Verley, P., Yu, H., Ren, Y., 2017. An individual-based model for simulating the ecosystem dynamics of jiaozhou bay, China. Ecol. Model. 360, 120–131.