

# MetaIBM: A Python-based library for individual-based modelling of eco-evolutionary dynamics in spatial-explicit metacommunities

Jian-Hao Lin<sup>a</sup>, Yu-Juan Quan<sup>b</sup>, Bo-Ping Han<sup>a,\*</sup>

<sup>a</sup> Department of Ecology, Jinan University, Guangzhou, 510632, China

<sup>b</sup> Department of Computer Science, Jinan University, Guangzhou, 510632, China

## ARTICLE INFO

### Keywords:

Community assembly  
Eco-evolutionary processes  
Individual-based model  
Landscape network  
Metacommunity  
Python library

## ABSTRACT

Individual-based modelling (IBM) is a powerful tool for simulating complex biological communities. By defining a population as comprising individuals that differ from one another, IBM can simulate the assembly and organisation of complex communities under various eco-evolutionary processes in a large spatial scale, with tremendous variables or parameters considered simultaneously. IBM disentangles a complex system into various sub-systems interacting with each other, allowing us to develop a unified library with a modular design for a wide range of complex scenarios in community assembly. In such a library, a number of parameters-controlled processes can be primitively coded as the sub-systems (or sub-models). Here, we released a Python-coded library as a framework for Metacommunity Individual-based Modelling (MetaIBM). As an open-source library, the MetaIBM has several merits, including: (a) it can be used to simulate a wide range of ecological problems of metacommunities. The metacommunity landscape and its environment gradients can be designed flexibly by users. Users can selectively turn off or on and set up parameters-controlled ecological processes according to their needs. (b) It adopts optimised algorithms and adapts to the high-performance computing devices, so that the users can explore a wide range of parameters space synchronously within a reasonable time. (c) It can be used to simulate a group of communities with up to millions of unique individuals, which is an originally plain portrayal of natural communities. To guide potential users, we provided the source codes of the library and a user manual. In the present article, we gave four examples to demonstrate how to design and model a metacommunity using the MetaIBM, simulating the community assembly in an islands-mainland model under the metacommunity framework with (a) neutral assumptions, (b) niche assumptions, (c) slow evolution scenarios, (d) rapid evolution scenarios. The examples showed that the MetaIBM can efficiently fit the community assembly, and reveal several intrigued species diversity patterns under the interaction of evolutionary processes and dispersal processes at multiple scales. The MetaIBM will be continuously maintained and updated to provide more functions in the future.

## 1. Introduction

In the last three decades, ecologists have long paid attention to biodiversity patterns from local to regional scales, but been challenged by increasing complexity in ecological processes at an extending spatial scale (Ricklefs, 1987; Chase, 2003; Leibold et al., 2004; Cornell and Harrison, 2014; Fukami, 2015; Urban et al., 2020). The island biogeography theory (IBT) calls the ecologists to connect local processes to regional patterns for the first time (MacArthur and Wilson 1967). The size of islands and the dispersal between barriers could affect the habitat variety, extinction rate and immigration, leading to the change of

equilibrium in the species richness and species turnover. Similarly, the metapopulation theory describes how much the proportion of islands does a species can occupy, considering the dynamics of populations on multiple island-like habitats that are connected by migration of individuals (Levins, 1969, 1970; Hanski 1998). Inspired by IBT unexpectedly, Hubbell and Bell generalised it into an individual-based neutral theory, which explicitly distinguishes the dynamics of local communities and metacommunities (Hubbell, 2001; Bell, 2001; Chave, 2004; Alonso et al., 2006). Furthermore, metacommunities where species and population in a patchy landscape are linked by dispersal, has become a fundamental conceptual unit and framework in modern

\* Corresponding author.

E-mail address: [tbphan@jnu.edu.cn](mailto:tbphan@jnu.edu.cn) (B.-P. Han).

ecology (Leibold et al., 2004; Mouquet and Loreau, 2002). Moreover, when ecology meets evolution, evolving metacommunity ecology has emphasised the importance of micro-evolution or rapid evolution in a regional scale, which usually leads to a monopolisation effect of a founder species (Thompson, 1998; Hendry and Kinnison, 1999, 2001; Hairston et al., 2005; Urban and Skelly, 2006; Loeuille and Leibold, 2008; Urban and DeMeester, 2009; De Meester et al., 2016).

Individual-Based Modelling (IBM) can be a really handful tool to deal with that complexity in natural communities (DeAngelis and Grimm, 2014; Grimm and Railsback, 2005; Grimm et al., 2017; Boyd et al., 2020b). First, when building an individual-based model, ecologists can easily reduce that complexity in modelling by disentangling a complex community into a few sub-models or various ecological processes. In classical ecology, community dynamics are usually formulated by using differential equations or probabilistic equations (Gotelli, 1995; May and McLean, 2007). However, as the species number in a community increases, more and more parameters and variables could matter to describe the community precisely, but these equations could not be easily solved using an analytical solution because of the difficulties in deducing. For such cases, simulation would be the only way to powerfully study the complex community. By using IBM, we can simulate any complex communities, in which tremendous parameters and variables can be considered simultaneously, and the computational calculations can be performed in a reasonable amount of time. Second, the variation among individuals within a population can be considered in IBM. In classical community ecology, population is the most practical, central, and unifying concept and individual is an impractical concept (Berriman 2002). However, population, a higher level of abstraction above individual, depicts the average properties of a group of individuals of the same species but loses the information of uniqueness of individuals. In many scenarios, such an information loss is acceptable, while the others do not, for example, the rapid evolution due to the genetic variation within any species of a community (Thompson, 1998; Hairston et al., 2005; Barrett and Schlüter, 2008). Theoretically, IBM can easily handle additional information among individuals by describing a population made up of individuals that may differ from one another (Uchmański and Grimm, 1996; Boyd et al., 2020a). Moreover, another tough utilisation of an individual-based approach is the computational problem of keeping records and tackling the individual-based datasets, which seems to have been impossible in the past. Fortunately, with the rapid development of computing hardware (e.g., high-performance computing devices) and algorithm (e.g., large-scale parallel computing in CPU or GPU), these interdisciplinary methods between computer science and ecology makes the tough tasks possible, despite the computational expensiveness of the approach still (Judson, 1994; Aumann, 2007; Winsberg, 2019). Simulation approaches can usually be an effective way to explore unknown and complex questions at the very beginning of a study in a new field. The simulating results of IBM help us get an intuitive understanding of the complex communities in a brand-new scientific study and, further researches can be designed and conducted more efficiently.

The IBM could be a general tool or approach paving the way for describing complex biological communities. Ecologists have been toughly seeking for general patterns, rules, and even laws in community ecology for decades, however painfully, with little exciting success. Community ecology was even considered a mess, with so much contingency that similar patterns could be generated by different rules and the same underlying rules could lead to different patterns, depending upon the details (Lawton, 1999). Additionally, although community patterns have been summarised to be generally driven by four basic processes: selection, drift, speciation, and dispersal, it seems that a vast number of other processes could also affect the community in details, but we usually do not know which process is important or not in a specific case (Vellend, 2010). This complexity has led tremendous intricate debates among community ecologists including, for examples, (a) whether a community is a coherent unit or a random assemblage

(Forbes, 1887; Clements, 1916; Gleason, 1926; Kingsolver and Robert, 1991; Kingsland, 2005); (b) whether we should study an ecological system in a local scale or a regional scale (Ricklefs, 1987; Levin, 1992; Chave, 2013); (c) whether a community is of neutrality (stochastic process) or niche (deterministic process) and so on (Hubbell, 2001; Bell, 2001; Whitfield, 2002; Leibold and McPeek, 2006; Wennekes et al., 2012). Both arguments in each debate above seems plausible, but as coin has two sides, it is hard to simultaneously incorporate these two opinions into a single traditional model. Yet, IBM can just involve everything into a single model by building a number of sub-models representing each process. Despite these complexities in real communities, it is still believed that, indeed, there do exist general patterns in the world. Therefore, a general strategy would be that we can just straightforward simulate the natural communities to look into what is really happening as a controller and an observer of the modelling system. In other words, instead of directly finding the general patterns by the data painstakingly collected from the fields and experiments, a challenging attempt would be that we could build a general tool that is able to simulate a wide range of complex scenarios in the communities, at least, as far as possible. Philosophically, IBM is an approach of systematic thinking by disentangling a giant complex system into various sub-systems interacting with each other (i.e., the sub-models in the IBMs) from the idea of system science. It contains both the advantages of reductionism and holism, which are two opposite philosophical modes of thinking in modern science (Wiegert, 1988; Wilson, 1988; Bergandi and Blandin, 1998). On the one hand, in the reductionism or individualism aspect, we can monitor or even manipulate the behaviours of each single individual in details to understand complex communities. On the other hand, from a holistic perspective, some unexpected results of the simulation can emerge, only when all the individuals interact with each other as a whole system. By applying IBM, instead of arguing such these knotty debates or arguing which process is important or not, we can simply involve everything (e.g., both local process and regional processes; both neutrality or niche; both ecological and evolutionary processes) into a single IBM to conduct simulations. These simulations would inspire us what is really important or not. Recently, several new patterns in metacommunity ecology have been explored and discovered by IBM (Urban and DeMeester, 2009; Vanoverbeke et al., 2016; Leibold et al., 2019). To facilitate these explorations, a unified library for building an IBM conveniently and efficiently is needed, so that a number of parameters-controlled processes are primitively coded as sub-models into the toolbox of the library.

Thus, inspired by the IBM from Urban and DeMeester (2009), Vanoverbeke et al. (2016) and Leibold et al. (2019), we aim to develop a Metacommunity Individual-based Modelling framework (i.e., MetaIBM in abbreviation), which is a Python-based library and a spatial explicit, agent-based, forward-in-time simulator for metacommunity landscape. The framework is suitable for various scenarios, where the heterogeneous landscapes (including the size and its spatial arrangement or structure), the environmental parameters, the environmental gradients, the species traits, and parameters-controlled individuals' behaviours (including colonisation, dispersal, selection, reproduction & mutation, dormancy, and germination) can be flexibly set by the users to fits their needs. The MetaIBM was developed to have the following merits: (1) it is able to model a flexibly designed landscape and a wide range of ecological and evolutionary scenarios by simply turning on or off each process, which is parameters-controlled; (2) if high performance computing devices (HPC) are available, it can help users spontaneously run a numbers of simulation trials to explore the results in a wide range of parameters spaces; (3) it can simulate up to millions of individuals simultaneously in a simulation within a reasonable time. In this article, we described the major properties of each biological or ecological process in the library in the main document with additional details in the user's manual. Then we offered four examples to demonstrate how to modelling the dynamics of biological communities using the library. We simulated the community assembly in an islands-mainland model with

(1) neutral assumptions, (2) niche assumptions, (3) slow evolution scenarios and (4) rapid evolution scenarios. Finally, we discussed the necessities in IBM and the advantages of using the library in the field of community ecology.

## 2. Materials and methods

Following previously IBM (Urban and De Meester, 2009; Vansverbeke et al. 2016; Leibold et al. 2019), we developed the Python-based library (i.e., MetaIBM) for a wider range of scenarios in modern community ecology. The library simulates the demographic, ecological and evolutionary processes of populations of multi-species in a metacommunity, in which individuals interact with each other by dispersal. In the library (Fig. 1, Table S1 in the Appendix 1), the landscape includes three spatial scales from locally to regionally: habitats, patches, and global scales (see Appendix 5 for details). Habitats scales, the fundamental units of the landscape in the model, functions as local communities in the metacommunity and each habitat is attributed to its own environments. At the patch scales, a patch is composed of several habitats and each patch in the landscape is labelled with a pair of unique X-Y coordinates. Finally, the global scale was defined as an entire metacommunity containing an unlimited number of patches. By spatial-explicitly parameterising the environmental variables among habitats, users can generate an environmentally heterogeneous landscape. Within habitats, individuals undergo colonisation, germination, growth, mortality, reproduction, mutation, dormancy, and disturbance. The migration of individuals in the metacommunity occurs during dispersal, which is controlled by a function of any landscape-specific and species-specific parameters. These eco-evolutionary processes are simulated over a given time-step set by users. Depending on the studying objects, users can decide whether each process is executed or not and set the parameters of these processes. All the setting details can be available in the user's manual.

### 2.1. Model assumptions

The MetaIBM can be employed in modelling a wide range of scenarios in community ecology with four fundamental assumptions: (1) island-mainland model assumption; (2) limited dispersal assumption; (3) pre-emptive competition assumption; (4) additive genetic effects (Table 1).

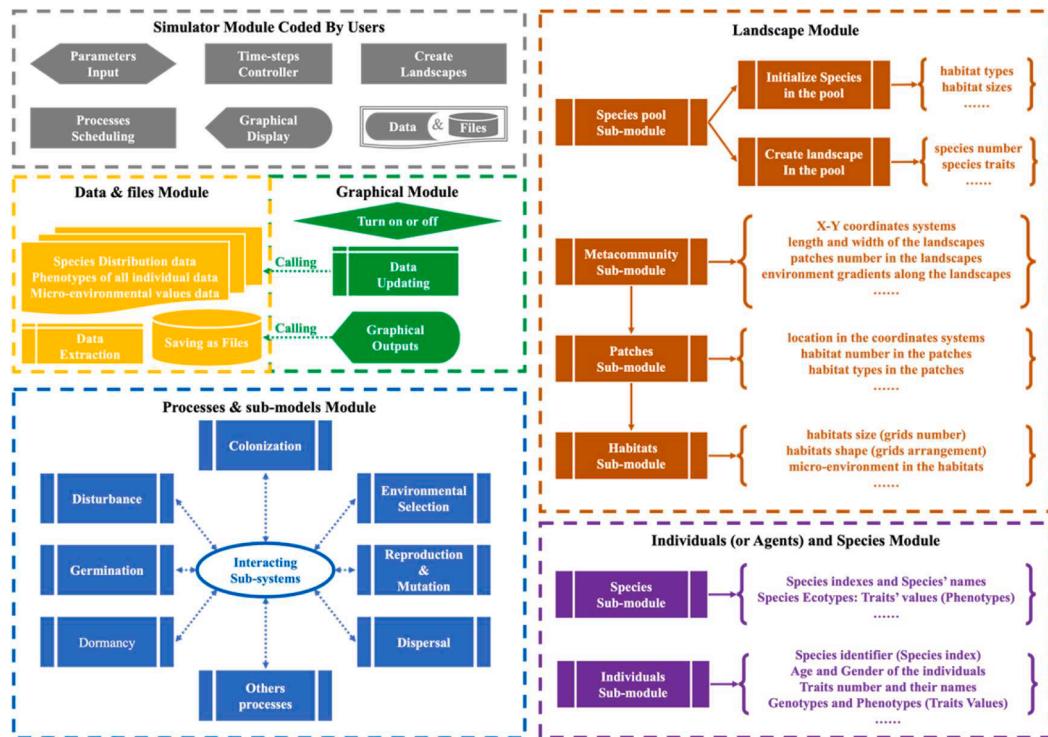
### 2.2. The landscapes and its environment

The MetaIBM provides a grid-based model, in which pre-emptive competition is assumed such that a grid can only be occupied by one individual because of limited resources (Chesson and Warner, 1981). We defined each grid as a *microsite*. Each habitat, the fundamental units of the landscape, consists of a set of microsites, the number of which can be set by users. Any specific habitat shapes can also be achieved by setting

**Table 1**

The details of the model assumptions used in the MetaIBM.

Assumption	Description
Island-mainland model	Based on the island-mainland model in biogeography, the model simulates the community assembly over a group of islands influenced by a mainland. The islands collectively form a metacommunity, with each island acting as a patch in the metacommunity system.
Limited dispersal	The model assumes that simulated perennial species can produce propagules, and only their offspring propagules can disperse over the landscape.
Pre-emptive competition	In the model, each grid can only support the survival of one individual and propagules can only colonise empty grids. In a community, individuals interact and compete only indirectly with each other via their response to environmental conditions.
Additive genetic effects	The determination of the phenotype by the genotype of an individual in the model follows the principle of additive genetic effects.



**Fig. 1.** The modules designed in MetaIBM. The MetaIBM are designed to involve six major modules in the Python library including: Simulator module, data & files module, graphical module, processes & sub-models module, landscapes module, individuals (or Agents) and species module. Notably, the simulator module is need to be coded by the users as a main function via calling the primitive functions from the library in the framework (An example in the Appendix 4).

the spatial arrangement of the microsites (for now, in the source code). Based on the environmental parameters of the habitat set by the users (mean and variation), the environmental values of the microsites in that habitat were set to be normally distributed. The MetaIBM can support unlimited types of landscape environments (e.g., temperature, humidity etc.) set by users. It can also allow the users to dynamically alter the environmental values during a simulation over time, which can address some questions on, for examples, daily, seasonal, or yearly environmental changes. Users can also create heterogeneous landscapes across habitats by setting different environmental parameters. The environments in a habitat can influence the demographic processes of individuals in that habitat and such a fitness landscape can demographically affect all the individuals across habitats in the metacommunity.

Patches, consisting of a set of habitats (which may or may not be environmentally different), are the secondary spatial unit, and each patch is labelled with a pair of unique X-Y coordinates. The topographic structure among patches in the metacommunity can also be set by users, expressed as an adjacency matrix with  $N \times N$  dimensions (Harary, 1962), if the patch number is equal to  $N$ . By default, the patches are topographically fully connected and the elements in the matrix indicate the Euclidean distance between each pair of patches according to their X-Y coordinates. When barriers between patches are considered, users can set other topography structure. Users can also automatically generate some classic topographic network structures of patches based on some classic algorithms in graph theory in the library (e.g., Steiner tree, minimum spanning tree, travelling salesman problem, small world network), which, for examples, could be useful in the study of corridors planning.

Finally, the global scale of the metacommunity model comprises a set of patches (Leibold et al., 2004). For a scenarios-driven study in ecology, the MetaIBM has a great degree of flexibility in landscape settings and can offer a quick insight into how the landscape influences the diversity patterns, which could be extremely tough and expensive to monitor in an empirical field.

### 2.3. Individuals' attributes

MetaIBM considers the complexity of individuals-level interactions in the natural communities. Each individual in a population owns its unique attributes including *species identifier*, *gender*, *age*, *phenotypes* (*traits values*), and *genotypes*. Users can set any number of traits for an individual they are interested in (e.g., optimum temperature and humidity). For each trait of an individual, we assumed  $L$  bi-allelic additive genes coded as 1 s and 0 s as genotypes and the phenotype was genetically based and calculated as the mean of the genotypes plus a stochastic Gaussian variable indicating a non-genetic phenotypic variation. Furthermore, the fitness landscape is an interaction between the traits' values of an individual and the values of environmental parameters in a habitat, which can influence the individual-levels demographic processes.

### 2.4. Environmental selection processes

The fitness landscape affects an individual demographically through the environmental selection process, after which the individual might die or survive, and the selection process is controlled by a survival rate calculated as a Gaussian distribution between traits values (phenotype) and micro-environment in the microsites as follow (Vanoverbeke et al., 2016):

$$S_i = (1 - d) \times \prod_{traits\_num} \exp - \left( \frac{P_i - E_m}{\omega} \right)^2,$$

where  $d$  is the baseline mortality rate;  $\omega$  is the width of stabilising selection; for each trait,  $P_i$  is the phenotype of an individual; and  $E_m$  is the

environmental values of a microsites the individual lived in. In the case above, a niche-based model is assumed and users can also choose other distributions; for example, a uniform distribution indicating that the survival rate is always a constant and neutral dynamics are assumed.

### 2.5. Reproduction and mutation processes

MetaIBM offers three options of reproduction mode for users: asexual reproduction, sexual reproduction or a mix of the two. The reproduction process is controlled by the birth rate parameter ( $b$ ) set by the users. The expectancy of offspring number ( $I$ ) in the metacommunity at each time step would be  $I = b \cdot N_f$ , where  $N_f$  is denoted as the number of all the female individuals in the metacommunity. In the asexual mode, all the individuals are set to be female and the genotype of the offspring is identical to that of its parent. In the sexual mode, Mendelian inheritance assumes that each value of the biallelic genes of the offspring is inherited from the female or male parents with the same probability of 50 %. In the mixed mode, a survival rate threshold is set by the users as a parameter and all the individuals in the fitness landscape will undergo asexual reproduction, if the survival rate of the individuals greater than the threshold otherwise; they undergo sexual reproduction, as some species are capable of both asexual and sexual reproduction, for example, cyclical parthenogenesis (Schön et al., 2009). Mutation also occurs while a bi-allelic gene's value is modified from 0 to 1 (vice versa) with a mutation rate ( $\mu$ ). The phenotype was then calculated based on its genotype and a non-genetic phenotypic variation. Above all, evolutionary processes including recombination in sexual reproduction or mutation occurs and users can control these processes by selecting reproduction mode or setting the parameters, mutation rate ( $\mu$ ). The offspring function as the propagules in the colonisation, dispersal, and germination processes later.

### 2.6. Colonisation and dispersal processes

Pre-emptive competition and lottery competition models are assumed such that only the empty microsites (grids) can be colonised by propagules randomly chosen from local or exotic offspring via dispersal processes (Smart, 1995; Chesson and Warner, 1981). Two dispersal pathways are assumed such that the propagules can disperse within a patch or disperse among patches in the metacommunity landscape. The dispersal process is controlled by the rate of dispersal within patch and dispersal among patches. Dispersal rate controls the total number of propagules available for dispersal. Within a patch, the propagules could disperse evenly between habitats. Dispersal among patches is in more complexity. For each specific pairwise patches in the metacommunity, the expectancy of emigrants or immigrants between the two patches between time-steps was controlled or calculated by a Markov process, with the consideration of landscape-specific parameters (distance between patches, barriers in the landscape), species-specific parameters (the long-distance ability of species) and a distance-migration equation (by default, migration reduces exponentially with the increase in distance). Moreover, the islands-mainland model can be constructed by building an initially occupied mainland as a species pool and a set of initially empty patchy islands of the metacommunity and the species colonised the islands from the mainland by propagules' rains (MacArthur and Wilson, 1967; Zobel et al., 1998; Fukami, 2005, 2015).

### 2.7. Dormancy processes

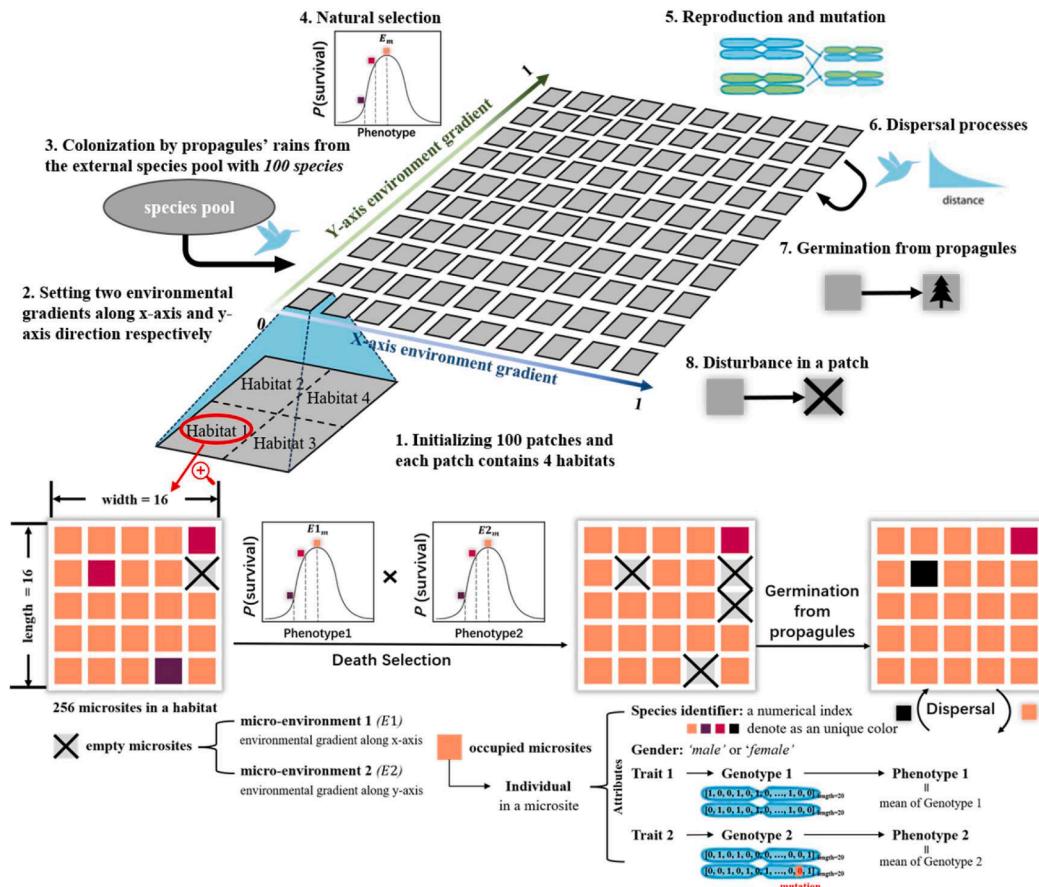
After dispersal and colonisation, the propagules that failed to colonise could be in dormancy. We also assumed that the resting propagules would not undergo environmental selection process.

## 2.8. Disturbance processes

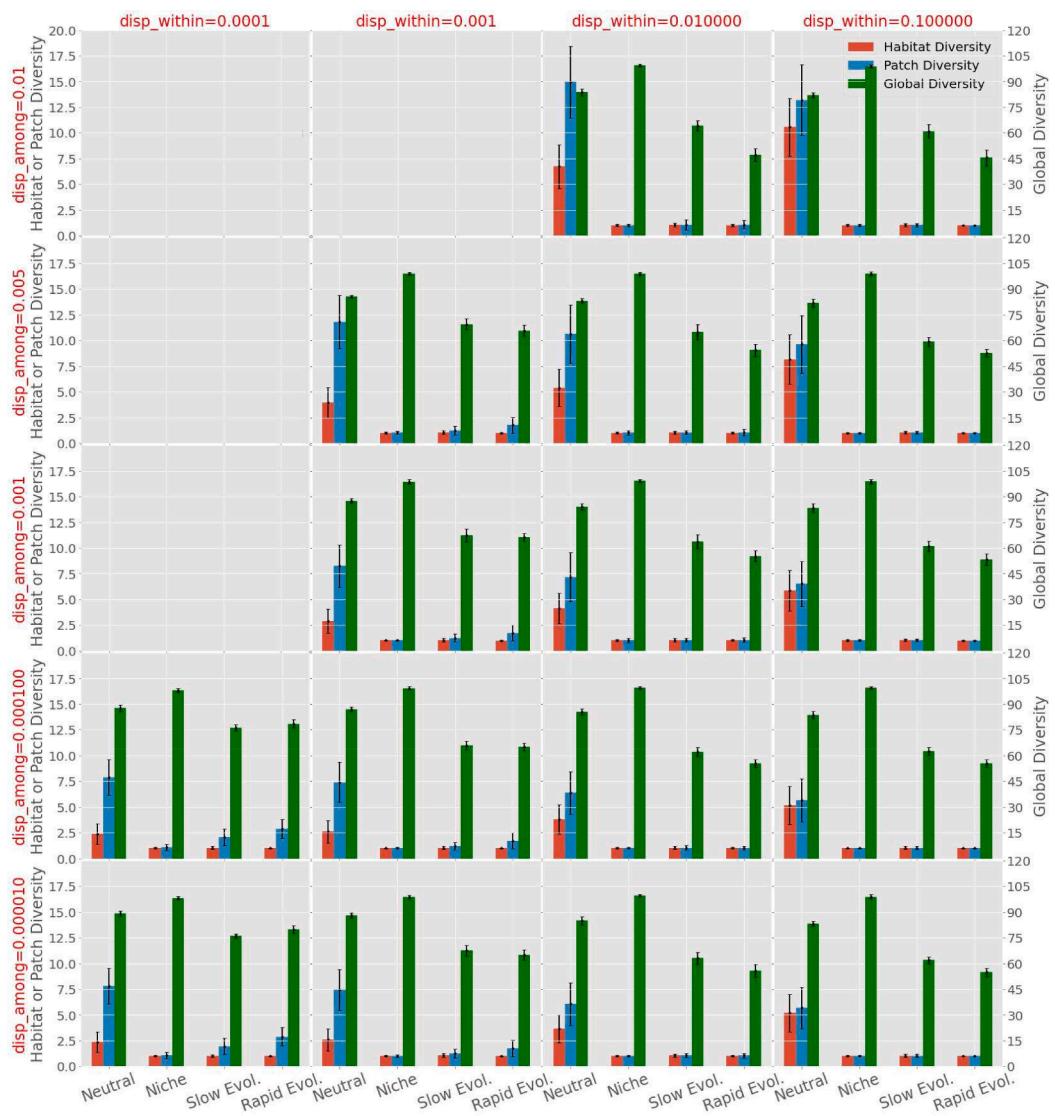
Disturbance processes may occur in habitats or patches scales, controlled by disturbance rate. When disturbance happened, all the exiting individuals in the habitat or patch and all the resting propagules would be dead (Leibold et al., 2019).

## 2.9. Modelling design

To elucidate how to build and run a simulation using the MetaIBM, we provided four examples simulating islands-mainland models with (a) neutral assumptions; (b) niche assumptions; (c) slow evolution scenarios; (d) rapid evolution scenarios respectively. In the context of the islands-mainland model (MacArthur and Wilson, 1967), we created an initially-occupied mainland with 100 species as an external species pool and an initially empty metacommunity with a spatially explicit structure including patch and habitat scales. To model the community assembly in the metacommunity, we simulated each sub-models in an order of colonisation, natural selection (environmental filtering), reproduction & mutation, dispersal process, germination, and disturbance processes for a time-step (Fig. 2). We simulated the model for 5000 time-steps to reach an equilibrium state, confirming with simulations of 20,000 time-steps.



**Fig. 2.** Illustration of the model design for the examples. The simulation, following an initialisation, runs the sub-models of colonisation, natural selection, reproduction & mutation, dispersal, germination, and disturbance processes in orders within a time-steps. An islands-mainland model was created with an initially-empty metacommunity with a spatially explicit structure including patch and habitat scales and an initially-occupied mainland with 100 species as an external species pool for the subsequent community assembly in the metacommunity. Two environmental parameters ( $X$  and  $Y$ ) have an environmental gradient ranging from 0 to 1 along the x-axis and y-axis respectively. There are 102,400 grids (microsites) in the global metacommunity and pre-emptive competition is assumed such that one microsite can only be occupied by only one individual. Death selection process is conducted according to the survival rate of an individual affected by the matching between the two genetic-based traits (phenotypes) of the individual and the two micro-environment values in the microsites. Then, the propagules are born in the asexual or sexual reproduction process, followed by the dispersal process of the propagules with the assumption of an exponential decay. The empty microsites in the landscape can be re-occupied by the propagules randomly sampled from the local offspring or the newly-arrivals by dispersal. Finally, a disturbance process of a patch occurs with a probability that all the individuals and propagules are dead in the patch. The model is run for 5000 time-steps to reach an equilibrium state.



**Fig. 3.** The biodiversity indicated by the inverse Simpson's index at habitat (red), patch (blue) and globe (green) scales in the metacommunity. The plotting is in relation to the rate of dispersal among patches (in rows panels), the rate of dispersal within a patch (in columns panels) and the four examples of neutral model, niche model, slow evolution and rapid evolution. Habitat or patch diversity were plotted on the y-axis in the left-hand side and global diversity was plotted on the y-axis in the right-hand side. The biodiversity in the plotting was averaged over habitats, patches and replicate runs.

increase in the strength of the dispersal within a patch, global diversity and patchy diversity decrease, while habitat diversity increase. The variation in the diversity (i.e., the error bars in the plotting) in the neutral scenario was far higher than that in the other scenarios, while the niche model predicted the lowest variation in diversity at a local scale. The variation in biodiversity decreases from patches and habitats scales to the global scale, especially in the cases where the strength of dispersal within a patch is relatively low. With an increase in the dispersal among patches rate, higher variation in biodiversity occurs.

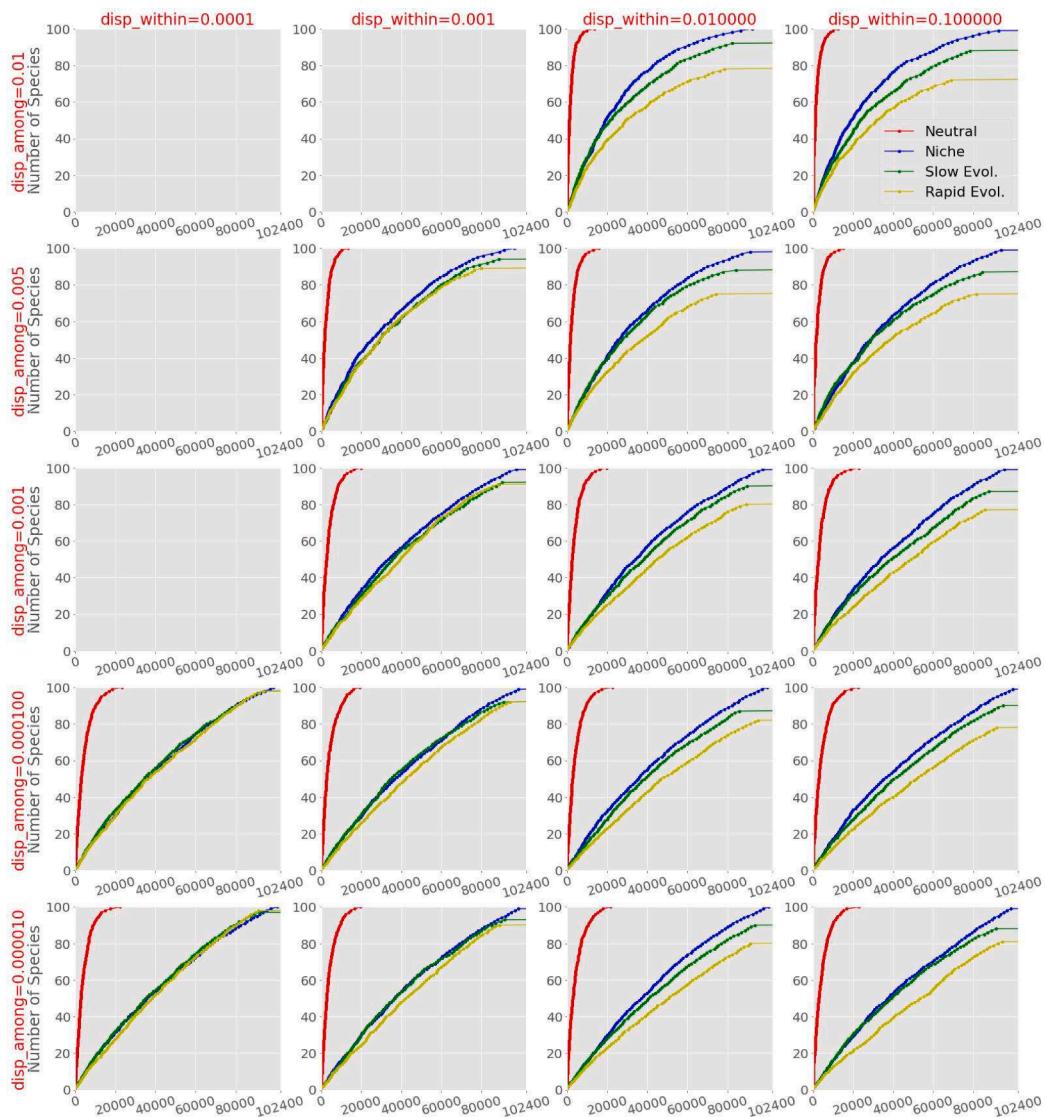
### 3.2. Species-area cumulative curves

There are 102,400 microsites in the metacommunity and each microsite is labelled with a unique ID. By counting the new species in an order of the microsites' ID, we plotted the species-area cumulative curves in relation to the dispersal rate and the four examples respectively (Fig. 4). All the curves exhibited an upward trend, but the rise of the curves varied with the niche, slow evolution and rapid evolution models from high to low, corresponding to the global diversity patterns from high to low in an order of niche, slow evolution and rapid evolution

model. However, in the neutral model where the natural selection process was not considered, the species-area cumulative curves rised up much faster than those in the others scenarios. Despite a higher global diversity in the niche model than that in the neutral one, the mixture patterns of the species distribution in the metacommunity contributed to a faster rising up of the curves for the neutral scenarios. Moreover, dispersal can influence the curve shape. A stronger dispersal among patches can lead to a rapid rising up of the curves for all the scenarios, while a sever exchange of the individuals within a patch would results in a strong difference of the curves among niche, slow evolution and rapid evolution.

### 3.3. Assembly mechanisms and their relative contributions

The relative contribution of the individual mechanisms (including species sorting, mass effect, habitat monopolisation and patch monopolisation effect) was quantified in the community assembly across the metacommunity (Fig. 5), based on the biodiversity index at each spatial scales and the decision-making table in Appendix 7. In the simulation with neutral hypothesis, the mass effect was the main mechanisms in the



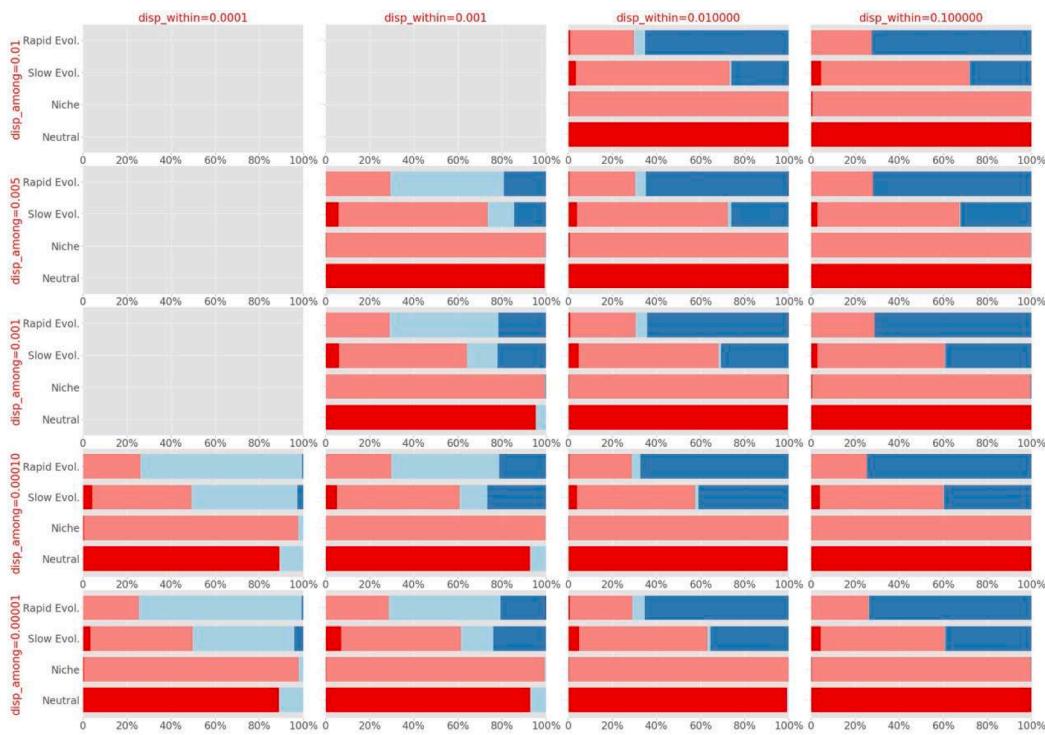
**Fig. 4.** The species-area cumulative curves for the metacommunity. There are 102,400 microsites in the metacommunity as shown in the x-axis and each microsite is assigned with a unique ID number ranging from 1 to 102,400. The species-area cumulative curves traversed the microsites in the order of the microsites' ID number to find new species. The curves shown are in relation to the rate of dispersal among patches (panels in rows) and dispersal within a patch (panels in columns) and the four examples included neutral model (red), niche model (blue), slow evolution (green) and rapid evolution (yellow).

community assembly and the priority effect in a habitat, although seldom, will occurs as well due to historical contingency in the community assembly, especially when the dispersal rate is at a low level. In the niche model, almost all the local assemblages were under the species sorting effect, indicating that nearly all the habitats were dominated by the pre-adapted species from the species pool. When evolution was taken into account, habitat and patch monopolisations were more prevalent because of the adaptation to the local environment for the monopolisation species. The evolutionary adaptation of a species in the slow evolution mode is driven by the mutation process and the micro-evolution in the rapid evolution mode is driven by a genetic recombination and mutation process. Hence, a higher proportion of the monopolisation effect contributes to the community assembly in the metacommunity for the rapid evolution scenarios. Additionally, the stronger the level of the dispersal within a patch, the greater contribution of the patch monopolisation was over the habitat monopolisation.

#### 4. Discussion

The MetaIBM was developed for users to be able to flexibly build a

wide range of modelling scenarios, and such a property is powerful to help us explore complex communities. Exploratory simulations and confirmatory experiments are the two basic research paradigms in scientific studies. When facing a new or complex research in which less knowledge or theories pre-exist, some exploratory experiments are need be conducted to generate hypotheses, gather preliminary data and gain insights that can guide further investigation, where confirmatory experiments or cases studies are needed to test or modify the hypotheses. However, traditionally it is hard to develop a unified tool to help us explore the new questions without a clear background, because these exploratory experiments are often more flexible, less structured, and may involve a broader range of variables. Building an IBM for community ecology requires the expertise in programming, and rather few unified and efficient libraries or software for IBM of community ecology have been invented. The MetaIBM is designed into various modules and developed in a way of object-oriented programming (OOP), which can help users build models solving a set of issues rather than a specific problem. More specifically, in the MetaIBM, the landscape can be defined in a highly flexible way such that the information of the landscape in spatial scales (patches or habitats) including their shapes, size



**Fig. 5.** The quantified contribution of assembly mechanisms in the metacommunity. The contributions were plotted in relation to the rate of dispersal among patches (panels in rows) and dispersal within a patch (panels in columns). The modelling scenarios included the neutral model, niche model, slow evolution and rapid evolution. The assembly mechanisms included mass effect (red), species sorting (light red), patch monopolisation (blue) and habitat monopolisation (light blue). Note that in the neutral scenarios where evolution is out of the consideration, priority effect (blue) is recognised if a habitat or patch is dominated by a single non-preadapted species.

(grids number) and location in the X-Y coordinate system can be customised by users. In the four examples, we built a metacommunity landscape consisting of three spatial scales (habitats scale, patches scale and global metacommunity scale) and it is the spatially explicit model that enable us to study the biodiversity patterns from local to regional scales. The number of environmental parameters and their gradients along the landscape can also be defined by the users. Species number, trait number, and trait values can also be set by users. Finally, the individuals' behaviours and other processes including colonisation, dispersal, death selection, reproduction & mutation, dormancy germination and disturbance processes are parameters-controlled so that users can turn off or on and control the processes according to their needs. In the four demonstrated examples, the models were designed and simulated by simply controlling the input parameters of the sub-models. Moreover, we usually do not know which parameters or variables are matters in a whole new study without any prior knowledge, so unfortunately, we have to explore a wide range parameters space to identify some important factors for the future researches. Fortunately, owing to the modular designs, the parameters and each sub-models in the library are independent, so that users can set various parameters in each process in a more convenient way to easily explore each combination of the inputted parameters. In the four examples, we simulated the models in the parameters space of the dispersal rates by simply setting the different parameters values for the dispersal process. In the simulation of the complex eco-evolutionary dynamics in the metacommunity, several new patterns were found. For example, in the regional scale, the combination of the strong dispersal process and the evolutionary process help a local monopolisation species to become a regional monopolisation species (i.e., spill-over effect), leading to a biodiversity loss in the metacommunity. Meanwhile, in the local scale, the evolutionary process can increase the biodiversity in a patch when the dispersal levels are weak.

The optimised algorithms in the library and high-performance computing (HPC) devices guarantee that the simulation of a complex

community will run in a reasonable time. Usually, computing power is the upmost limitation in running a large simulation (e.g., a large number in the patch number or the individuals), let alone a replicated simulation in the parameters space. In the classic software for IBM, taking the *Netlogo* software as an example, to simulate the random walks of an individual in the landscape map, we usually and randomly allocate a direction and a speed parameter to the individual at each time-step. However, simulating the random dispersal of individuals, in a manner similar to the random movement of air molecules through space, is literally computationally expensive. Thus, in the MetaIBM, based on the geographical network model in graph theory, with the patches being nodes and corridors between patches being edges, Markov stochastic processes are used to describe the dispersal process in the metacommunity landscape. Calculating the short-path distance between pairwise patches by Floyd's algorithms (Floyd, 1962) and based on the exponential decay of the dispersal-distance model (Leibold et al., 2019), we can get the dispersal rates between pairwise patches, which are the elements of the Markov matrix describing the dispersal process of the individuals between patches. Subsequently, based on the dispersal rates between pairwise patches, the dispersal processes were directly executed on each single specific individual by random sampling without replacement. All the above-mentioned calculation procedures are executed automatically by calling the dispersal function in the library. Other optimised algorithms are also provided in the library as well, to reduce the consumption of computing resources. In natural communities, being prone to explosive reproduction without adversity, species always reproduce excessive propagules to compete for the limited living resources in a space and such phenomena lead to the death of tremendous propagules even before germination (Fenner, 2012). However, excessive propagules that fail to germinate would meaninglessly consume computing powers in the devices. To accelerate the computation and avoid the meaningless computation in excessive propagules, a new algorithm is provided such that the reproduction, dispersal and

germination processes were coded into the same sub-model, and are executed synchronously, while the population demographics in the processes are controlled in a mathematical way to ensure the veracity in a simulation. In MetaIBM, the lottery model is assumed such that the occupation of the empty microsites from the excessive propagules in the germination process is just a probabilistic game, which is also conducted by random sampling without replacement from immigrants and local propagules in the habitats in the programming. To be more specific, in a time-step, we first calculate the expected number of immigrants, emigrants and local propagules in the habitats demographically (with birth rate and dispersal rate) and then based on the number of the empty microsites after the selection process (controlled by the survival rate), we further calculate how much immigrants, emigrants and local propagules is really needed to reoccupy the empty microsites in the meta-community, if the expected number of propagules is far excessive to the empty microsite number; finally, the reproduction and germination processes are executed at the same time. All the procedures mentioned above in the calculation are executed automatically by calling the *reproduction-dispersal-germination* function in the library. Notably, if the excessive propagules are also important in some scenarios, for example the propagules that fail to germinate can get into dormancy, users can also simulate the reproduction, germination and dormancy processes separately and the excessive propagules will keep being calculated in the modelling.

The MetaIBM can simulate metacommunity dynamics with a capacity up to millions of individuals, whereas a similar library of IBM can only run a simulation with hundreds of thousands of individuals in the past (Landguth and Cushman, 2010; Landguth et al., 2017). Such an individual-based method creates a simulation similar to the real natural communities by directly executing the processes on individuals. Taking the stochastic process as an example, compared to the traditional modelling methods, when describing a biological community by IBM, the model can consider the stochasticity events on every individual, rather than by describing the stochasticity in terms of probability and ignoring the stochasticity on individuals into the population concepts. However, these features make the computation difficult. Conventionally, computational complexity rises exponentially with an increase of the number of the individuals in an IBM system, as the number of the interactions among the individuals and between individuals and the environment, which the computer has to tackle with, surges (Grimm and Railsback, 2005; Railsback and Grimm, 2019). Cheerfully, because of the optimised algorithms offered in the MetaIBM, a linear increasement in the simulation time has accomplished, with an enlargement of the metacommunity landscapes (i.e., the grids number, population size or the individuals' number in the modelling system). Moreover, the HPC devices make it easier to simulate the large model with a large population size repeatedly and synchronously. Many ecological phenomena are dependent on the population size, such as the Allee-effect, ecological drift, gene drift, the genetic variation in a population (Lande, 1976; Stephens et al., 1999; Hubbell, 2001; Barrett and Schlüter, 2008). The more individuals and their details the MetaIBM can simulate, the more the emerging results of a modelling system are in tune with the natural communities. With a higher carrying capacity in a simulation, a large number of species can be simultaneously simulated in the model as well and users may discover some patterns of the rare species, which otherwise may not be easily found in a small community. Additionally, the genetic diversity in a population is also referred to the population size. When rapid evolution occurs, the standing genetic variation within a population can sufficiently affect the resulting patterns. Such an information regarding the genetic properties of each unique individual in a population cannot be ignored (Thompson, 1998; Hairston et al., 2005; Barrett and Schlüter, 2008). The MetaIBM can also save the information data of every single individual in the simulation, even if a large population size is designed in the model. In the four examples demonstrated, we used these detailed and large data on every single individual to plot the species-area cumulative curves. We attempt to develop a tool that

could model the real world as far as possible. In reality, take a high-plant community as an example, the trees community dataset of a 50-hectares plot in Barro Colorado Island, Panama records 21,457 trees and 255 distinct species with greater than 10 cm diameter at breast height (Condit et al., 2002). Then, the MetaIBM library can easily deal with the complexity of simulating such a plant community, at least in the terms of population size.

The complexity of the natural communities can be described and explored by disentangling it into a number of executable processes applying IBM. To offer a unified tool that is suitable for exploring some general theories in community ecology theoretically, we developed the library including various processes coded primitively and controlled by parameters to help users to explore community ecology using IBM conveniently. Now, we intend to release the MetaIBM framework coded as a Python-based library in the first version. The library framework will be continuously maintained and improved to provide more functions in the library in the future.

## CRediT authorship contribution statement

**Jian-Hao Lin:** Writing – review & editing, Writing – original draft, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Yu-Juan Quan:** Writing – review & editing, Software. **Bo-Ping Han:** Writing – review & editing, Writing – original draft, Supervision, Funding acquisition, Conceptualization.

## Declaration of competing interest

The authors declare no conflict of interest in relation to this paper.

## Data availability

The source code and the user's manual of the library has been deposited in Zenodo, <https://doi.org/10.5281/zenodo.10662687>. The simulation code and data in the examples and its analysing codes has been deposited in Zenodo, <https://doi.org/10.5281/zenodo.10662708>. The MetaIBM library is released on GitHub in an open-source project and the project will keep being maintained and updated in the future. <https://github.com/katherine999/MetaIBM>.

## Acknowledgement

This research was funded by the National NSF of China (No. 32171538 and No. 31670460). The National Supercomputer Centre in Guangzhou and the High-Performance Computing Platform of Jinan University are appreciated for their support with extensive simulations.

## Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.ecolmodel.2024.110730](https://doi.org/10.1016/j.ecolmodel.2024.110730).

## References

- Alonso, D., Etienne, R.S., McKane, A.J., 2006. The merits of neutral theory. *Trends Ecol. Evol. (Amst.)* 21 (8), 451–457.
- Aumann, C.A., 2007. A methodology for developing simulation models of complex systems. *Ecol. Modell.* 202 (3–4), 385–396.
- Barrett, R.D., Schlüter, D., 2008. Adaptation from standing genetic variation. *Trends Ecol. Evol. (Amst.)* 23 (1), 38–44.
- Bell, G., 2001. Neutral macroecology. *Science* 293 (5539), 2413–2418.
- Bergandini, D., Blandin, P., 1998. Holism vs. reductionism: do ecosystem ecology and landscape ecology clarify the debate? *Acta Biotheor.* 46 (3), 185–206.
- Berryman, A.A., 2002. Population: a central concept for ecology? *Oikos* 97 (3), 439–442.
- Boyd, R.J., Sibly, R., Hyder, K., Walker, N., Thorpe, R., Roy, S., 2020a. Simulating the summer feeding distribution of Northeast Atlantic mackerel with a mechanistic individual-based model. *Prog. Oceanogr.* 183, 102299.

- Boyd, R., Walker, N., Hyder, K., Thorpe, R., Roy, S., Sibly, R., 2020b. SEASIM-NEAM: a spatially-explicit Agent-based SIMulator of north east Atlantic mackerel population dynamics. *MethodsX* 7, 101044.
- Chase, J.M., 2003. Community assembly: when should history matter? *Oecologia* 136, 489–498.
- Chave, J., 2004. Neutral theory and community ecology. *Ecol. Lett.* 7 (3), 241–253.
- Chave, J., 2013. The problem of pattern and scale in ecology: what have we learned in 20 years? *Ecol. Lett.* 16, 4–16.
- Chesson, P.L., Warner, R.R., 1981. Environmental variability promotes coexistence in lottery competitive systems. *Am. Nat.* 117 (6), 923–943.
- Clements, F.E., 1916. Plant succession: an analysis of the development of vegetation (No. 242). Carnegie institution of Washington.
- Condit, R., Pitman, N., Leigh Jr, E.G., Chave, J., Terborgh, J., Foster, R.B., Hubbell, S.P., 2002. Beta-diversity in tropical forest trees. *Science* 295 (5555), 666–669.
- Cornell, H.V., Harrison, S.P., 2014. What are species pools and when are they important? *Annu. Rev. Ecol. Evol. Syst.* 45, 45–67.
- De Meester, L., Vanoverbeke, J., Kilsdonk, L.J., Urban, M.C., 2016. Evolving perspectives on monopolization and priority effects. *Trends Ecol. Evol. (Amst.)* 31 (2), 136–146.
- DeAngelis, D.L., Grimm, V., 2014. Individual-based models in ecology after four decades. *F1000Prime Rep.* 6.
- Fenner, M.W., 2012. *Seed Ecology*. Springer Science & Business Media.
- Floyd, R.W., 1962. Algorithm 97: shortest path. *Commun. ACM* 5 (6), 345.
- Forbes, S.A., 1887. The lake as a microcosm. *Bull. Ill. State Nat. Hist. Surv.* 15, 537–550.
- Fukami, T., 2015. Historical contingency in community assembly: integrating niches, species pools, and priority effects. *Annu. Rev. Ecol. Evol. Syst.* 46, 1–23.
- Fukami, T., 2005. Integrating internal and external dispersal in metacommunity assembly: preliminary theoretical analyses. *Ecol. Res.* 20 (6), 623–631.
- Gleason, H.A., 1926. The individualistic concept of the plant association. *Bull. Torrey Bot. Club* 7–26.
- Gotelli, N.J., 1995. *A Primer of Ecology*. Sinauer Associates Incorporated.
- Grimm, V., Railsback, S.F., 2005. *Individual-based Modeling and Ecology*. Princeton University Press.
- Grimm, V., Aylón, D., Railsback, S.F., 2017. Next-generation individual-based models integrate biodiversity and ecosystems: yes we can, and yes we must. *Ecosystems* 20, 229–236.
- Grimm, V., Berger, U., Bastiansen, F., Eliassen, S., Ginot, V., Giske, J., DeAngelis, D.L., 2006. A standard protocol for describing individual-based and agent-based models. *Ecol. Model.* 198 (1–2), 115–126.
- Hairston Jr, N.G., Ellner, S.P., Geber, M.A., Yoshida, T., Fox, J.A., 2005. Rapid evolution and the convergence of ecological and evolutionary time. *Ecol. Lett.* 8 (10), 1114–1127.
- Hanski, I., 1998. Metapopulation dynamics. *Nature* 396 (6706), 41–49.
- Harary, F., 1962. The determinant of the adjacency matrix of a graph. *Siam Rev.* 4 (3), 202–210.
- Hendry, A.P., Kinnison, M.T., 1999. Perspective: the pace of modern life: measuring rates of contemporary microevolution. *Evolution (N Y)* 53 (6), 1637–1653.
- Hendry, A.P., Kinnison, M.T., 2001. An introduction to microevolution: rate, pattern, process. *Genetica* 112, 1–8.
- Hubbell, S.P., 2001. *The Unified Neutral Theory of Biodiversity and Biogeography* (Vol. 32). Princeton University Press, Princeton.
- Judson, O.P., 1994. The rise of the individual-based model in ecology. *Trends Ecol. Evol. (Amst.)* 9 (1), 9–14.
- Kingsland, S.E., 2005. *The Evolution of American Ecology*. JHU Press, pp. 1890–2000.
- Kingsolver, J.G., Paine, R.T., 1991. Conversational biology and ecological debate. Foundations of Ecology. Univ. of Chicago Press, Chicago, pp. 309–317. Edited by Leslie A. Real and James H. Brown.
- Lande, R., 1976. Natural selection and random genetic drift in phenotypic evolution. *Evolution (N Y)* 314–334.
- Landguth, E.L., Cushman, S., 2010. CDPOP: a spatially explicit cost distance population genetics program. *Mol. Ecol. Resour.* 10 (1), 156–161.
- Landguth, E.L., Bearlin, A., Day, C.C., Dunham, J., 2017. CDMetaPOP: an individual-based, eco-evolutionary model for spatially explicit simulation of landscape demogenetics. *Methods Ecol. Evol.* 8 (1), 4–11.
- Leibold, M.A., McPeek, M.A., 2006. Coexistence of the niche and neutral perspectives in community ecology. *Ecology* 87 (6), 1399–1410.
- Lawton, J.H., 1999. Are there general laws in ecology? *Oikos* 177–192.
- Leibold, M.A., Holyoak, M., Mouquet, N., Amarasekare, P., Chase, J.M., Hoopes, M.F., Gonzalez, A., 2004. The metacommunity concept: a framework for multi-scale community ecology. *Ecol. Lett.* 7 (7), 601–613.
- Leibold, M.A., Urban, M.C., De Meester, L., Klausmeier, C.A., Vanoverbeke, J., 2019. Regional neutrality evolves through local adaptive niche evolution. *Proc. Natl. Acad. Sci.* 116 (7), 2612–2617.
- Levin, S.A., 1992. The problem of pattern and scale in ecology: the Robert H. MacArthur award lecture. *Ecology* 73 (6), 1943–1967.
- Levins, R., 1969. Some demographic and genetic consequences of environmental heterogeneity for biological control. *Bull. ESA* 15 (3), 237–240.
- Levins, R., 1970. Extinction. *Lect. Notes Math.* 2, 75–107.
- Loeuille, N., Leibold, M.A., 2008. Evolution in metacommunities: on the relative importance of species sorting and monopolization in structuring communities. *Am. Nat.* 171 (6), 788–799.
- MacArthur, R.H., Wilson, E.O., 1967. *The Theory of Island Biogeography*. Princeton University Press.
- May, R., McLean, A.R., 2007. *Theoretical Ecology: Principles and Applications*. Oxford University Press.
- Mouquet, N., Loreau, M., 2002. Coexistence in metacommunities: the regional similarity hypothesis. *Am. Nat.* 159 (4), 420–426.
- Railsback, S.F., Grimm, V., 2019. *Agent-based and Individual-Based Modeling: a Practical Introduction*. Princeton university press.
- Ricklefs, R.E., 1987. Community diversity: relative roles of local and regional processes. *Science* 235 (4785), 167–171.
- Schön, I., Martens, K., Dijk, P., 2009. *Lost Sex: The Evolutionary Biology of Parthenogenesis*. Springer, Dordrecht, Netherland.
- Smart, R.M., 1995. Preemption: an important determinant of competitive success. In: *Proceedings, 29th Annual Meeting, Aquatic Plant Control Research Program: 14–17 November 1994. US Army Engineer Waterways Experiment Station, Vicksburg, Mississippi*.
- Stephens, P.A., Sutherland, W.J., Freckleton, R.P., 1999. What is the Allee effect? *Oikos* 185–190.
- Thompson, J.N., 1998. Rapid evolution as an ecological process. *Trends Ecol. Evol. (Amst.)* 13 (8), 329–332.
- Uchmanski, J., Grimm, V., 1996. Individual-based modelling in ecology: what makes the difference? *Trends Ecol. Evol. (Amst.)* 11 (10), 437–441.
- Urban, M.C., De Meester, L., 2009. Community monopolization: local adaptation enhances priority effects in an evolving metacommunity. *Proc. R. Soc. B: Biol. Sci.* 276 (1676), 4129–4138.
- Urban, M.C., Skelly, D.K., 2006. Evolving metacommunities: toward an evolutionary perspective on metacommunities. *Ecology* 87 (7), 1616–1626.
- Urban, M.C., Strauss, S.Y., Pelletier, F., Palkovacs, E.P., Leibold, M.A., Hendry ..., A.P., Giery, S.T., 2020. Evolutionary origins for ecological patterns in space. *Proc. Natl. Acad. Sci.* 117 (30), 17482–17490.
- Vanoverbeke, J., Urban, M.C., De Meester, L., 2016. Community assembly is a race between immigration and adaptation: eco-evolutionary interactions across spatial scales. *Ecography* 39 (9), 858–870.
- Vellend, M., 2010. Conceptual synthesis in community ecology. *Q. Rev. Biol.* 85, 183–206.
- Wennekes, P.L., Rosindell, J., Etienne, R.S., 2012. The neutral–niche debate: a philosophical perspective. *Acta Biotheor.* 60, 257–271.
- Whitfield, J., 2002. Ecology: neutrality versus the niche. *Nature* 417 (6888), 480–482.
- Wiegert, R.G., 1988. Holism and reductionism in ecology: hypotheses, scale and systems models. *Oikos* 267–269.
- Wilson, D.S., 1988. Holism and reductionism in evolutionary ecology. *Oikos* 269–273.
- Winsberg, E., 2019. *Science in the Age of Computer Simulation*. University of Chicago Press.
- Zobel, M., van der Maarel, E., Dupré, C., 1998. Species pool: the concept, its determination and significance for community restoration. *Appl. Veg. Sci.* 1 (1), 55–66.