



# pycoalescence and rcoalescence: Packages for simulating spatially explicit neutral models of biodiversity

Samuel E. D. Thompson<sup>1,2</sup> | Ryan A. Chisholm<sup>1</sup> | James Rosindell<sup>2</sup>

<sup>1</sup>Department of Life Sciences, Silwood Park Campus, Imperial College London, Ascot, UK

<sup>2</sup>Faculty of Science, Department of Biological Sciences, National University of Singapore, Singapore

## Correspondence

Ryan A. Chisholm  
Email: ryan.chis@gmail.com

## Funding information

Joint Imperial-NUS PhD Scholarship; Natural Environment Research Council, Grant/Award Number: NE/I021179 and NE/L011611/1; James S. McDonnell Foundation, Grant/Award Number: 220020470; Singapore Ministry of Education, Grant/Award Number: WBS R-154-000-A12-114

Handling Editor: Timothée Poisot

## Abstract

1. Neutral theory proposes that some macroscopic biodiversity patterns can be explained in terms of drift, speciation and immigration, without invoking niches. There are many different varieties of neutral model, all assuming that the fitness of an individual is unrelated to its species identity. Variants that are spatially explicit provide a means for making quantitative predictions about spatial biodiversity patterns.
2. We present software packages that make spatially explicit neutral simulations straightforward and efficient. The packages allow the user to customize both dispersal and landscape structure in a wide variety of ways. We provide a Python package *pycoalescence* and a functionally equivalent R package *rcoalescence*. In both packages, the core routines are written in C++ and make use of coalescence methods to optimize performance.
3. We explain the technical details of the packages and give examples for their application, with a particular focus on two scenarios of ecological and evolutionary interest—a landscape with habitat fragmentation, and an archipelago of islands.
4. Spatially explicit neutral models represent an important tool in ecology for understanding the processes of biodiversity generation and predicting outcomes at large scales. The effort required to implement these complex spatially explicit simulations efficiently has thus far been a barrier to entry. Our packages increase the accessibility of these models and encourage further investigation of the primary mechanisms underpinning biodiversity.

## KEYWORDS

coalescence, dispersal, ecological drift, modelling, neutral theory, spatially explicit, speciation

## 1 | INTRODUCTION

A priority for community ecology is to understand the mechanisms determining species composition (Agrawal et al., 2007; Vellend, 2010) and unravel the interplay between ecology, biogeography and evolution. One theoretical approach is to generate biodiversity patterns using theoretical models and compare the patterns to empirical data. The models incorporate processes deemed important from real-world observations or prior theoretical studies.

This interplay of theoretical and empirical work helps reveal the elusive links between ecological patterns and their mechanisms. Theoretical models can be studied mathematically or via computer simulations. In this paper, we focus on the latter.

Neutral theory represents one notable branch of ecological theory (Hubbell, 2001; Rosindell, Hubbell, & Etienne, 2011). It captures three key mechanisms (dispersal, drift and speciation; Vellend, 2010) in a species-agnostic manner, meaning that species identities do not influence the dynamics. Numerous neutral models have been used

to predict biodiversity, in some cases producing surprisingly good fits to several biodiversity patterns in real systems (e.g. Halley & Iwasa, 2011; Muneeppeerakul et al., 2008; Volkov et al., 2003; Volkov et al., 2007). Even in scenarios where the neutrality assumption is inappropriate, neutral theory may be useful as either a null model or a model where the system exists under 'idealized' conditions (Alonso, Etienne, & McKane, 2006; Rosindell, Hubbell, & Etienne, 2011). In this way, neutral theory can act as a baseline for expected biodiversity patterns whose explanation requires non-neutral processes.

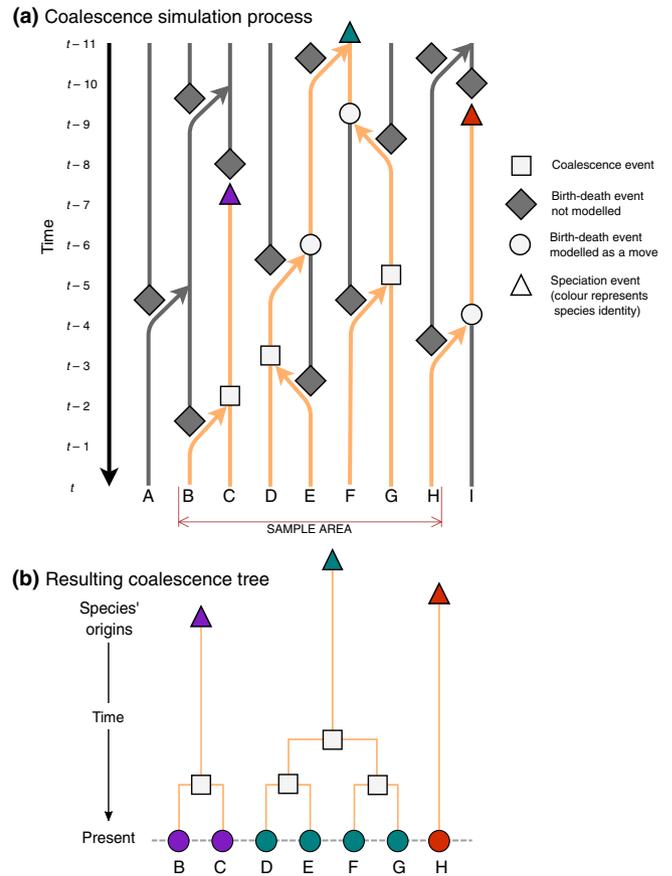
A well-studied ecological neutral model is the spatially implicit variant (Hubbell, 2001), with only a very abstract notion of space: a well-mixed (non-spatial) local community is connected to a larger well-mixed (non-spatial) metacommunity by immigration. The metacommunity undergoes drift and occasional renewal of diversity via speciation, or is assumed simply to be unchanging over relevant time-scales. The local community undergoes drift and renewal of diversity via immigration from the metacommunity. Many variants of the basic spatially implicit neutral model have been studied (Chave, 2004; Hubbell, 2001; Volkov et al., 2003), with many analytical results discovered (Alonso & McKane, 2004; Etienne, 2005; Green & Plotkin, 2007; Vallade & Houchmandzadeh, 2003), and extensive investigations conducted via simulation (Alonso et al., 2006; Azaele, Pigolotti, Banavar, & Maritan, 2006; Gilbert, Laurance, Leigh, & Nascimento, 2006).

## 1.1 | Coalescence methods

Computation time can quickly explode in neutral model simulations, especially as the number of individuals increases. To minimize this, coalescence methods have been imported from population genetics and have revolutionized ecological neutral simulations (Rosindell, Wong, & Etienne, 2008). In contrast to the traditional method where individuals are traced forwards through time, a coalescence approach tracks lineages backwards to their speciation origin. The two approaches are mathematically equivalent, but the coalescence approach is usually many orders of magnitude faster. A coalescence simulation essentially constructs a 'family tree' or 'coalescence tree' of all focal individuals (Figure 1b), from which the species identity of each individual and its relationship to other individuals can be inferred. The performance improvements of coalescence methods mean that simulations can be large enough to generate neutral predictions of biodiversity patterns on landscape scales. Coalescence is not only faster than traditional forwards-in-time methods, but can model scenarios that would be practically impossible under the forwards-in-time approach, including scenarios on essentially infinite landscapes (Rosindell & Cornell, 2007).

## 1.2 | Spatially explicit neutral models

Spatially explicit neutral models differ from spatially implicit models by their more realistic representation of space (Chave, Muller-Landau, & Levin, 2002; Rosindell & Cornell, 2007). Even



**FIGURE 1** Diagram of backwards-time simulations using coalescence techniques (a) and the resultant coalescence tree (b). (a) The simulation progresses backwards from the time of interest ( $t_0$ ) to time  $t_0 - 11$ . Orange lines indicate tracked lineages. Black lines describe lineages which are not tracked and are irrelevant to the simulation outcome, thus representing saved computational time from the standard forwards-in-time approach. (b) Purple, teal and red circles represent individuals in the present day belonging to each species. The respectively coloured triangles indicate the origins of each species. Grey squares represent coalescence events between individuals that correspond to a shared ancestor. Only spaces B–H lie within the focal community sample of interest, and therefore A and I are not tracked (indicated in dark grey in [a]). However, if individuals from the sample area have ancestors from outside the sample area (e.g. the individual starting at H), all events are still recorded, including coalescence and dispersal events, as these could impact the outcome

in a homogeneous environment, space can serve to isolate populations that are dispersal limited. In a spatially heterogeneous environment, space can further increase population isolation and endemism and lead to idiosyncratic biodiversity patterns (Campos, Rosas, de Oliveira, & Gomes, 2013; Gascuel, Laroche, Bonnet-Lebrun, & Rodrigues, 2016; Pereira, Borda-De-Água, & Martins, 2012). Incorporating spatial structure into ecological models is thus necessary to more accurately depict reality. Spatially explicit neutral models have seen wide use (Jordan, Barraclough, & Rosindell, 2016; O'Dwyer & Green, 2010; Rosindell & Cornell, 2007), especially in hyper-diverse ecosystems such as

coral reefs (Bode, Connolly, & Pandolfi, 2012; Dornelas, Connolly, & Hughes, 2006; Tang & Zhou, 2013) and tropical rainforests (Chave & Leigh, 2002; May, Huth, & Wiegand, 2015; Shem-Tov, Danino, & Shnerb, 2017).

### 1.3 | Neutral simulation packages

Numerous packages exist for simulating both neutral and niche dynamics of community assembly (Hankin, 2007; Laughlin, Joshi, van Bodegom, Bastow, & Fulé, 2012; Munoz et al., 2018; Novack-Gottshall, 2020; Taudiere & Violle, 2016). These tools massively reduce the barriers of entry for an ecologist wishing to model ecological communities and minimize time spent implementing code. However, to our knowledge no packages existed for Python or R prior to this paper for spatially explicit neutral models, the development of which is an immensely time-consuming task that could prohibit many ecologists from including them in their studies. In this manuscript, we introduce two packages (`pycoalescence` for Python and `rcoalescence` for R) that subsume the technical details of the coalescence methods under a user-interface layer that is accessible to any researcher. We have included a diverse feature set, including spatially and temporally varying density, sampling from subsets of the landscape and different modes of dispersal that we believe will cater for the vast majority of studies requiring spatially explicit neutral models. Published applications using these packages include exploring the effects of habitat fragmentation on extinction debt (Thompson, Chisholm, & Rosindell, 2019) and investigating the contribution of niche and neutral processes in a tropical rainforest (Bongalov et al., 2019).

## 2 | MODEL DESCRIPTION

Our model of neutral dynamics employs coalescence methods. This involves a backwards-in-time birth–death simulation to determine the shared ancestries (the coalescence tree, Figure 1b) within a focal community sample; next, each individual's species identity is determined by scanning the coalescence tree along the individual's lineage backwards in time until a speciation event is encountered. After the species identities of all individuals in the present-day community have been established in this way, any metrics of diversity which can be calculated on empirical census data, such as species richness, species abundance distributions and spatial or temporal species turnover indices, can be calculated just as easily on the simulated community using standard methods, as if the community were an empirical census.

### 2.1 | Simulation loop

Coalescence works by tracking lineages backwards through time (Rosindell, Wong, & Etienne, 2008). The backwards-in-time coalescence approach is generally possible only if the processes involved

do not depend on a species' identity, an assumption that is true by definition for neutral models. Some processes involving spatial and temporal habitat variation (e.g. fragmentation) can also be treated with the coalescence approach providing that the habitat variation affects all individuals in the same way regardless of species identity, that is, the system is still neutral.

In a forwards-in-time model, the simulation progresses through repeated death, dispersal and birth events. A backwards-in-time coalescence model is the reverse: each time step involves choosing an individual to be 'un-born' and finding its parent via an 'un-dispersal' event (reproduction is modelled asexually so that each individual has only one parent). For  $n$  individuals in a community,  $n$  lineages are initially tracked backwards in time. At each time step, there is a small probability,  $\nu$  (the 'speciation rate' for the baseline simulation) that a lineage will undergo speciation (technically, 'un-speciation'). At this point a species identity is assigned to the lineage and tracking is halted as it will not affect the ecological composition of the  $n$  sampled individuals. The number of lineages being tracked can also decrease via 'coalescence events', that is, when two or more lineages share the same ancestor at a particular point in time. After a coalescence event, the coalesced lineages form a single lineage continuing further back in time. Once there is a single lineage left, this lineage is assigned a unique species identity, the simulation is completed and the entire coalescence tree is stored. Algorithmically, the process of coalescence within a spatial model is to

1. Initialize the algorithm to track the lineages of  $n$  individuals in a focal community sample. These are the individuals whose species identities we want to determine.
2. Choose a random individual from the current set of lineages being tracked. Then one of two events occurs:
  - a. Either, with some small probability  $\nu$ , the lineage is deemed to have arisen by speciation at this point in time, a speciation event is stored in the coalescence tree, and the lineage does not need to be tracked further (the number of tracked lineages decreases by one). Go back to step 2.
  - b. Otherwise, the individual must be 'un-born'. Go to step 3.
3. Calculate the dispersal event:
  - a. Randomly generate a dispersal distance (using inverse dispersal events) and a direction.
  - b. Check that the end location is habitat. If there is no habitat present, go back to 3a. Otherwise, this is the new location for the parent. Go to step 4.
4. Check the parent's destination for tracked individuals.
  - a. If the location of the parent contains another lineage that is being tracked then merge the two lineages (the number of tracked lineages decreases by one), record a coalescence event in the coalescence tree and continue tracking the merged lineage.
  - b. Otherwise, simply move the lineage to the parent's location.
5. Go to step 2.

The process continues until only one individual remains.

## 2.2 | Package implementation

We provide three broad scenarios for neutral models (Table 1). Our implementation contains many advances over the basic coalescence algorithm; we summarize these features in Table 2 (see also Appendix C). Many of these features have been previously

**TABLE 1** Key scenarios of neutral theory available within the packages

Scenario	Approach
Non-spatial	Well-mixed community of $n$ individuals. Can be made spatially implicit by adding immigration from a metacommunity (see Table 2)
Spatially explicit landscape	Landscape within which individuals have defined spatial positions and dispersal governed by a specified kernel
Archipelago (partly spatial)	System of islands, each a well-mixed non-spatial community, connected via prescribed dispersal probabilities

implemented (e.g. spatial dispersal processes in infinite landscapes, Rosindell & Cornell, 2007) but this is the first time combinations of many of these features can be used together. An example of such would be spatially explicit landscapes with habitat change over time and space, whilst sampling individuals only from certain regions. For the user, the advanced functionality is presented as a set of straightforward options during the simulation set-up.

Packages for both Python (pycoalescence) and R (rcoalescence) interface with the same C++ simulator. A number of functions for each package setup simulation variables and generate simple biodiversity metrics, retaining equivalent core functionality but providing an interface customized to the requirements of each language. Installation instructions are provided in Appendices A and B.

### 2.2.1 | Spatial landscapes and dispersal

Dispersal is the key component required to make a simulation spatially explicit. Individuals have a fixed location in space and

**TABLE 2** Package components and their applicability within each scenario (see Table 1 for scenarios)

Model component	Options	Scenario		
		Non-spatial	Spatially explicit	Archipelago
Density of individuals	Density of individuals is uniform over space (i.e. contiguous)	X	✓	✓
	Density varies across space using maps of individual densities	X	✓	✓
	Density varies across space and time using density maps for each time slice	X	✓	✓
Landscape boundaries	Landscape has hard boundaries for the simulated world	✓	✓	✓
	Landscape has an infinite contiguous landscape outside the focal region defined by maps	X	✓	X
	Landscape has infinitely tiling repeats of the density maps in all directions	X	✓	X
Dispersal of individuals	Dispersal process uses a dispersal kernel (e.g. normal, fat-tailed or uniform)	X	✓	X
	Dispersal process uses a dispersal map defining dispersal probabilities between every pair of cells	X	✓	✓
Demographic rates (reproduction and/or death)	Demographic rates are equal for all individuals	✓	✓	✓
	Maps define spatially varying demographic rates. These rates do not affect densities, but do affect the relative turnover rates of different areas	X	✓	✓
Sampling	All simulated individuals are sampled at a single point in time	✓	✓	✓
	All simulated individuals are sampled at multiple points in time	✓	✓	✓
	A defined proportion of simulated individuals is sampled globally	✓	✓	✓
	Maps define sampling effort across space	X	✓	✓
Species generation	Speciation occurs randomly at a single point in time	✓	✓	✓
	As above, but a minimum number of generations is required before speciation (i.e. protracted speciation)	✓	✓	✓
	Immigration from a metacommunity defines the assemblage of potential species	✓	✓	✓

propagules disperse across the landscape according to the location of their parent and a function which controls the dispersal distance. Two methods of dispersal exist for individual-based simulations—either the new location is chosen based on a two-dimensional dispersal kernel, or a dispersal map is used, whereby dispersal probabilities are set by the user between all combinations of cells in a landscape.

### 2.3 | Dispersal kernels

Dispersal kernels refer to the process of choosing the parent of an individual by randomly drawing distances. The probability distribution of distances in our implementation can be normal, fat-tailed or uniform, presented as a selection of options for the user. A normal distribution is chosen by default, because this has been the most common selection in the literature and should be suitable for many modelling processes. The parent cell is selected by drawing from the probability distribution and selecting a random direction, describing a dispersal from the original cell to a new location (the parent cell). To ensure that density is correctly accounted for, the cell is then rejection-sampled based on its density relative to the maximum density from across the landscape. For example, a cell that has a density of 10 individuals in a landscape that has a maximal density of 20 individuals will be rejected half of the time. If rejection occurs, a distance and a direction are again drawn to find another potential parent cell and the process is repeated until a dispersal event is accepted. Rejection sampling represents the increased probability of offspring to originate from more populated cells, and allows density to vary over space.

### 2.4 | Dispersal maps

A dispersal map is an alternative method of selecting destinations by providing a map file specifying the probability of dispersal from each cell to every other cell in the landscape. Although this is only possible for finite landscapes, it allows the user to exactly control connectivity between cells. This option is more appropriate for systems of connected habitat islands where each patch is well mixed, such as in various previous studies of archipelagos (Chisholm, Fung, Chimalakonda, & O'Dwyer, 2016; Gascuel et al., 2016; Rosindell & Harmon, 2013; Van Der Gast, Lilley, Ager, & Thompson, 2005).

We implemented the dispersal map as an unraveled map where dispersal occurs from  $i$ th row to the  $j$ th column, with the  $i$  and  $j$  indices denoting a unique location,  $l$ , according to  $l = x + y \times x_{\text{dim}}$ ;  $x$ ,  $y$  are the coordinates of a cell on the landscape and  $x_{\text{dim}}$  is the  $x$  dimension of the landscape. Thus the dispersal map should be square with dimensions  $x_{\text{dim}} \times y_{\text{dim}}$  by  $x_{\text{dim}} \times y_{\text{dim}}$ . With a dispersal map, spatially varying densities and reproductive rates (see Appendix C) are incorporated into the probability distributions.

## 3 | EXAMPLE APPLICATIONS

The following two sections provide templates for using the packages in a habitat loss model and an island archipelago model. For each, we outline the scenario, describe the questions the modelling is aiming to resolve and present the results. Full vignettes for both scenarios, including simulation, analysis and visualization code, are available from [https://github.com/thompsonsd/rcoalescence\\_examples](https://github.com/thompsonsd/rcoalescence_examples) and provided in Appendices E and F.

### 3.1 | Landscape model

A common scenario for neutral modelling is simulating a fully spatially explicit landscape, possibly with changes to the structure and proportion of habitat over time (e.g. Bongalov et al., 2019; Chisholm et al., 2018; Thompson, Chisholm, & Rosindell, 2019). As an example, consider a scenario where an area outside a hypothetical national park will be deforested. Our goal is to estimate the species richness and spatial distribution of biodiversity in the long term following deforestation. These landscapes are shown in Figure 2a.

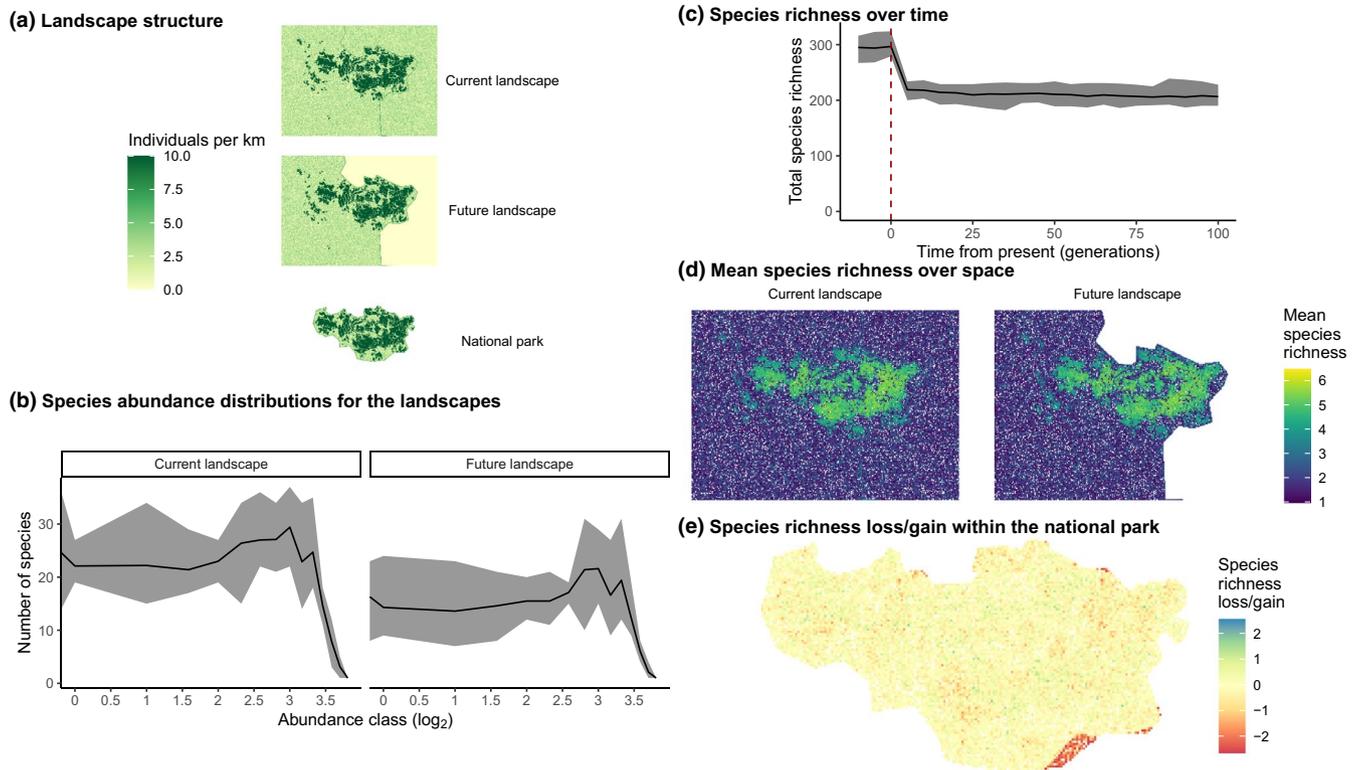
We make several assumptions for this scenario. In pristine forest (mostly within the national park), the density of individuals is 10 per  $\text{km}^2$  (i.e. per cell on the map). Areas with lower quality habitat have scaled density accordingly. We assume that individuals disperse according to a normal dispersal kernel with  $\sigma = 1$  km and that the per-capita speciation rate of the system is  $nu = 0.0001$ . To average over the stochasticity in simulations, we run 10 simulations with different random number seeds and aggregate the results.

The scenario aims to answer three questions:

1. How many species are lost when the habitat is lost?
2. What are the species abundance distributions before and after the habitat loss?
3. How does the spatial distribution of species richness change after the habitat is lost?

Just a few lines of code are required to run the simulation, which was repeated 10 times with different random number seeds. In Python, the minimal code is

```
sim = Simulation()
s.set_simulation_parameters(
    seed=1,
    task=1,
    output_directory="output",
    min_speciation_rate = 0.0001,
    sigma=2,
    deme=10,
    landscape_type="tiled_file",
    times=[x for x in range(0, 111, 5)])
```



**FIGURE 2** Results from a hypothetical habitat loss scenario. (a) The landscapes being modelled, including the location of a national park and the change in habitat structure following deforestation. (b) The species abundance distributions (Preston plots) for the current and future landscapes (Preston, 1960). (c) The species richness over time, taken from across simulations. For (b) and (c), solid lines represent the mean and grey shaded areas represent the minima and maxima, taken from across simulation runs. (d) The mean species richness for each 1 km<sup>2</sup> cell in the landscape. (e) The species richness loss and gain for the each 1 km<sup>2</sup> cell in the national park

```
s.set_map_files("null", "future_landscape.tif")
s.add_historical_map(
  fine_file="current_landscape.tif",
  coarse_file="none",
  time=99.9)
s.run()
```

and equivalently, in R

```
sim <- SpatialTreeSimulation$new()
sim$setSimulationParameters(
  task = 1,
  seed = 1,
  output_directory = "output",
  min_speciation_rate = 0.0001,
  times_list = seq(0, 110, 5),
  fine_map_file = "future_landscape.tif",
  deme = 10,
  historical_fine_map = "current_landscape.tif",
  gen_since_historical = 99.9,
  habitat_change_rate = 0.0,
  sigma = 2,
  landscape_type = "tiled_fine"
```

```
)
sim$runSimulation()
```

We used the package functions to extract the mean species abundance distribution and mean species richness across simulations (Figure 2b,c) to answer questions 1 and 2. In terms of species richness, the models estimate a roughly 29% decrease following habitat loss in both the short and long term. However, within the park, the decrease is only around 3.7% (from an average of 111.7 before the habitat loss to an average of 107.6 afterwards). The pattern is reflected in the species abundance distributions (Figure 2b); the curve is generally lower as the number of species overall has decreased.

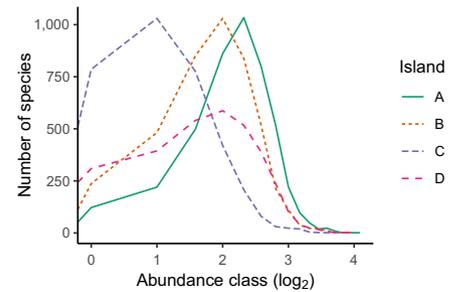
We calculated the species richness within each 1 km<sup>2</sup> cell in the landscape for each time point, exploring the spatial distribution of biodiversity in order to answer question 3. For the current and future landscapes, the local species richness roughly follows the density of individuals (Figure 2a,c), except that the edges of high-quality habitat are degraded by influence from the surrounding area. Within the park, there is no significant species loss except for along the southeastern edge, where the deforestation has occurred within the park boundaries. The conclusion is therefore that the deforestation has little effect outside the area which experienced the habitat loss for these model parameters.

**FIGURE 3** Results from an example island archipelago scenario. (a) The matrix that defines the dispersal probabilities between each island. Dispersal rates are given as relative probabilities from row indices to column indices. (b) The species abundance distributions (Preston plots) for each island (Preston, 1960). (c) The species richness for each island. (d) The endemic richness for each island. For (c) and (d), the weights of the lines represent the relative dispersal rates and the size of the islands represents the number of individuals within each island

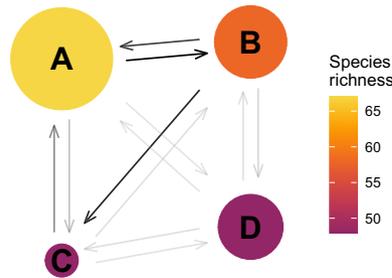
**(a) Table of dispersal probabilities from row indices to column indices**

island	A	B	C	D
A	0.614	0.355	0.052	0.001
B	0.251	0.630	0.297	0.020
C	0.134	0.009	0.641	0.005
D	0.001	0.006	0.010	0.974

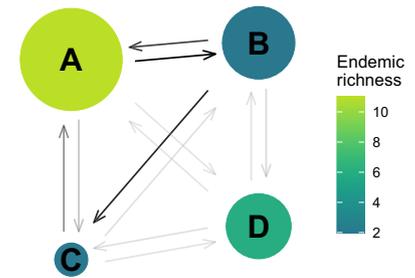
**(b) Island species abundance distributions**



**(c) Island species richness**



**(d) Island endemic richness**



### 3.2 | Archipelago model

The biodiversity of islands is a field of primary interest to many ecologists (MacArthur & Wilson, 1967). Island–mainland studies can be represented under neutral theory (e.g. Burbrink, McKelvy, Alexander Pyron, & Myers, 2015; Chisholm et al., 2016; Rosindell & Harmon, 2013), as can more generalized models for any number of interconnected islands. In these scenarios, dispersal occurs from the mainland to the islands and between all islands in the system (Figure 4a, Appendix C). In our simulation packages, these dispersal probabilities can be represented by a matrix and stored as a map file (Figure 4b, Appendix C).

In our island scenario, we want to estimate biodiversity across each of the islands based on known biodiversity from a small sample on the largest island. In this example, we assume that the dispersal probabilities are known (Figure 3a). The islands are well connected, except for island D, which has a relatively low probability of individuals dispersing to and from other islands.

Species abundance distributions generated from these islands are shown in Figure 3b as Preston (1960) plots. The differences between the islands in Figure 3b follows the differences in species richness shown in Figure 3c. In particular, islands C and D have very similar species richness values, but very different Preston plots. Island C is a small, but well-connected island and consequently has a large number of rare species immigrating from the other, larger islands. In contrast, island D has a larger proportion of the biodiversity originating from within the island itself (Figure 3d).

Although this example included just four islands, this approach could be used to model many other systems involving patches of interconnected habitat, including bacterial tree-hollow communities, coral reefs or laboratory experimental systems. The only restriction is that the RAM footprint of the simulation increases with the square

of the landscape size, which, in our experience, limits the approach to scenarios with <250,000 habitat cells.

## 4 | DISCUSSION

Neutral theory provides a scalable, mechanistic model of biodiversity that incorporates dispersal, drift and speciation dynamics at the level of the individual. Models employing neutral theory have already been applied to a variety of systems, including river systems (Munepeperakul et al., 2008), coral reefs (Dornelas et al., 2006; Volkov et al., 2007), birds (Graves & Rahbek, 2005; Halley & Iwasa, 2011; Ostling, 2005) and tropical trees (Chave et al., 2002; Condit et al., 2002). These disparate ecosystems can often be conceptualized within one of the two general scenarios we outline in this manuscript—spatially heterogeneous landscapes (including habitat loss and fragmentation) and island archipelagos.

Habitat loss represents one of the biggest threats to biodiversity globally (Wilson et al., 2016). In investigating such scenarios, spatial structure of habitat is critical, requiring, by definition, a spatially explicit model. Real habitats vary across time and space in their demographic rates and densities; the core features of our packages were developed to meet these requirements. Our example habitat loss scenario (Figure 2) demonstrates how a simulation with spatially and temporally varying maps can be performed.

Within island biogeography (MacArthur & Wilson, 1967), the spatial component in neutral models takes the form of dispersal rates defined explicitly between pairs of locations (Gascuel et al., 2016; Kopp, 2010; Pinheiro et al., 2017; Rosindell & Phillimore, 2011). Our representation of this allows the user to specify an arbitrary matrix of dispersal probabilities, allowing, for example, a very large number

of islands to be efficiently modelled without copious amounts of manual set-up. For all but the simplest of models, simulations remain the easiest way to investigate biodiversity outcomes. By providing an easy way to test realistic island structures, we hope to enhance theoretical and practical understanding of how diversity is generated on islands.

Spatially explicit neutral models have been used in several systems already (Chave et al., 2002; Muneeppeerakul et al., 2008; Rosindell & Cornell, 2007), but their broader application across ecology has been hampered in part by a lack of convenient tools for running such models. Previously, performing a spatially explicit neutral simulation required the arduous process of implementing a simulator from scratch in a high performance language, including coalescence techniques and fully spatial dispersal. We believe our packages should allow anyone with limited programming experience to run spatially explicit neutral models, exposing the potential of these powerful models to a new audience.

## ACKNOWLEDGEMENTS

We would like to thank Rogier Hintzen, Lucas Fernandes, Tak Fung and Meryl Theng for their providing user feedback on earlier versions of the presented packages. S.E.D.T. was supported by the Joint Imperial-NUS PhD Scholarship. R.A.C. was supported by grants from the James S. McDonnell Foundation (#220020470) and the Singapore Ministry of Education (WBS R-154-000-A12-114). J.R. was funded by fellowships from the Natural Environment Research Council (NERC; NE/I021179, NE/L011611/1). This study is a contribution to Imperial College's Grand Challenges in Ecosystems and the Environment initiative. All simulations were run on high throughput compute systems at Imperial College London.

## AUTHORS' CONTRIBUTIONS

All authors conceived of features and implementation; J.R. provided the initial coalescence algorithm; S.E.D.T. designed and authored the packages and wrote the initial manuscript draft with supervisory input throughout from J.R. and R.A.C. All authors contributed critically to drafts and gave final approval for publication.

## PEER REVIEW

The peer review history for this article is available at <https://publons.com/publon/10.1111/2041-210X.13451>.

## DATA AVAILABILITY STATEMENT

pycoalescence source code can be found at <https://doi.org/10.5281/zenodo.3924587> (Thompson et al., 2020a). pycoalescence can be installed via PyPI (<https://pypi.org/project/pycoalescence>) or conda (<https://anaconda.org/conda-forge/pycoalescence>) with a development version available through GitHub (<https://github.com/thompsons/pycoalescence>). rcoalescence source code can be found at <https://doi.org/10.5281/zenodo.3924585> (Thompson et al., 2020b). rcoalescence can be installed directly from GitHub (<https://github.com/thompsons/rcoalescence>). Code and data for generating all figures

can be found at <https://doi.org/10.5281/zenodo.3924564> (Thompson et al., 2020c) and <https://doi.org/10.5281/zenodo.3924567> (Thompson et al., 2020d) or on GitHub ([https://github.com/thompsons/pycoalescence\\_examples](https://github.com/thompsons/pycoalescence_examples) and [https://github.com/thompsons/rcoalescence\\_examples](https://github.com/thompsons/rcoalescence_examples)).

## ORCID

Samuel E. D. Thompson  <https://orcid.org/0000-0003-4486-1390>

Ryan A. Chisholm  <https://orcid.org/0000-0002-9847-1710>

James Rosindell  <https://orcid.org/0000-0002-5060-9346>

## REFERENCES

- Agrawal, A. A., Ackerly, D. D., Adler, F., Arnold, A. E., Cáceres, C., Doak, D. F., ... Power, M. (2007). Filling key gaps in population and community ecology. *Frontiers in Ecology and the Environment*, 5(3), 145–152. [https://doi.org/10.1890/1540-9295\(2007\)5\[145:FKGIPAJ\]2.0.CO;2](https://doi.org/10.1890/1540-9295(2007)5[145:FKGIPAJ]2.0.CO;2)
- Alonso, D., Etienne, R. S., & McKane, A. J. (2006). The merits of neutral theory. *Trends in Ecology & Evolution*, 21(8), 451–457. <https://doi.org/10.1016/j.tree.2006.03.019>
- Alonso, D., & McKane, A. J. (2004). Sampling Hubbell's neutral theory of biodiversity. *Ecology Letters*, 7(10), 901–910. <https://doi.org/10.1111/j.1461-0248.2004.00640.x>
- Azalee, S., Pigolotti, S., Banavar, J. R., & Maritan, A. (2006). Dynamical evolution of ecosystems. *Nature*, 444(7121), 926–928. <https://doi.org/10.1038/nature05320>
- Bode, M., Connolly, S. R., & Pandolfi, J. M. (2012). Species differences drive nonneutral structure in pleistocene coral communities. *The American Naturalist*, 180(5), 577–588. <https://doi.org/10.1086/667892>
- Bongalov, B., Burslem, D. F. R. P., Jucker, T., Thompson, S. E. D., Rosindell, J., Swinfield, T., ... Coomes, D. A. (2019). Reconciling the contribution of environmental and stochastic structuring of tropical forest diversity through the lens of imaging spectroscopy. *Ecology Letters*, 22(10), 1608–1619. <https://doi.org/10.1111/ele.13357>
- Burbrink, F. T., McKelvy, A. D., Alexander Pyron, R., & Myers, E. A. (2015). Predicting community structure in snakes on Eastern Nearctic islands using ecological neutral theory and phylogenetic methods. *Proceedings of the Royal Society B: Biological Sciences*, 282(1819). <https://doi.org/10.1098/rspb.2015.1700>
- Campos, P. R. A., Rosas, A., de Oliveira, V. M., & Gomes, M. A. F. (2013). Effect of landscape structure on species diversity. *PLoS ONE*, 8(6), e66495. <https://doi.org/10.1371/journal.pone.0066495>
- Chave, J. (2004). Neutral theory and community ecology. *Ecology Letters*, 7(3), 241–253. <https://doi.org/10.1111/j.1461-0248.2003.00566.x>
- Chave, J., & Leigh, E. G. (2002). A spatially explicit neutral model of  $\beta$ -diversity in tropical forests. *Theoretical Population Biology*, 62(2), 153–168. <https://doi.org/10.1006/tpbi.2002.1597>
- Chave, J., Muller-Landau, H. C., & Levin, S. A. (2002). Comparing classical community models: Theoretical consequences for patterns of diversity. *The American Naturalist*, 159(1), 1–23. <https://doi.org/10.1086/324112>
- Chisholm, R. A., Fung, T., Chimalakonda, D., & O'Dwyer, J. P. (2016). Maintenance of biodiversity on islands. *Proceedings of the Royal Society B: Biological Sciences*, 283(1829), 20160102. <https://doi.org/10.1098/rspb.2016.0102>
- Chisholm, R. A., Lim, F., Shuen Yeoh, Y., Wei Seah, W., Condit, R., & Rosindell, J. (2018). Species–area relationships and biodiversity loss in fragmented landscapes. *Ecology Letters*, 21(6), 804–813. <https://doi.org/10.1111/ele.12943>
- Condit, R., Pitman, N., Leigh, E. G., Chave, J., Terborgh, J., Foster, R. B., ... Muller-Landau, H. C. (2002). Beta-diversity in tropical forest

- trees. *Science*, 295(5555), 666–669. <https://doi.org/10.1126/science.1066854>
- Dornelas, M., Connolly, S. R., & Hughes, T. P. (2006). Coral reef diversity refutes the neutral theory of biodiversity. *Nature*, 440(7080), 80–82. <https://doi.org/10.1038/nature04534>
- Etienne, R. S. (2005). A new sampling formula for neutral biodiversity. *Ecology Letters*, 8(3), 253–260. <https://doi.org/10.1111/j.1461-0248.2004.00717.x>
- Gascuel, F., Laroche, F., Bonnet-Lebrun, A. S., & Rodrigues, A. S. L. (2016). The effects of archipelago spatial structure on island diversity and endemism: Predictions from a spatially-structured neutral model. *Evolution*, 70(11), 2657–2666. <https://doi.org/10.1111/evo.13067>
- Gilbert, B., Laurance, W. F., Leigh, E. G., & Nascimento, H. E. M. (2006). Can neutral theory predict the responses of Amazonian tree communities to forest fragmentation? *The American Naturalist*, 168(3), 304–317. <https://doi.org/10.1086/506969>
- Graves, G. R., & Rahbek, C. (2005). Source pool geometry and the assembly of continental avifaunas. *Proceedings of the National Academy of Sciences of the United States of America*, 102(22), 7871–7876. <https://doi.org/10.1073/pnas.0500424102>
- Green, J. L., & Plotkin, J. B. (2007). A statistical theory for sampling species abundances. *Ecology Letters*, 10(11), 1037–1045. <https://doi.org/10.1111/j.1461-0248.2007.01101.x>
- Halley, J. M., & Iwasa, Y. (2011). Neutral theory as a predictor of avifaunal extinctions after habitat loss. *Proceedings of the National Academy of Sciences of the United States of America*, 108(6), 2316–2321. <https://doi.org/10.1073/pnas.1011217108>
- Hankin, R. K. S. (2007). Introducing untb, an R package for simulating ecological drift under the unified neutral theory of biodiversity. *Journal of Statistical Software*, 22(12), 1–15. <https://doi.org/10.18637/jss.v022.i12>
- Hubbell, S. P. (2001). The unified neutral theory of biodiversity and biogeography. *Monographs in Population Biology*, 17(3), 375. [https://doi.org/10.1016/S0006-3207\(02\)00228-8](https://doi.org/10.1016/S0006-3207(02)00228-8)
- Jordan, S. M. R., Barraclough, T. G., & Rosindell, J. (2016). Quantifying the effects of the break up of Pangaea on global terrestrial diversification with neutral theory. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1691). <https://doi.org/10.1098/rstb.2015.0221>
- Kirkby, M. J., MacArthur, R. H., & Wilson, E. O. (1967). *The theory of island biogeography*. Princeton University Press. <https://doi.org/10.2307/1796430>
- Kopp, M. (2010). Speciation and the neutral theory of biodiversity: Modes of speciation affect patterns of biodiversity in neutral communities. *BioEssays*, 32(7), 564–570. <https://doi.org/10.1002/bies.201000023>
- Laughlin, D. C., Joshi, C., van Bodegom, P. M., Bastow, Z. A., & Fulé, P. Z. (2012). A predictive model of community assembly that incorporates intraspecific trait variation. *Ecology Letters*, 15(11), 1291–1299. <https://doi.org/10.1111/j.1461-0248.2012.01852.x>
- May, F., Huth, A., & Wiegand, T. (2015). Moving beyond abundance distributions: neutral theory and spatial patterns in a tropical forest. *Proceedings of the Royal Society B: Biological Sciences*, 282(1802). <https://doi.org/10.1098/rspb.2014.1657>
- Muneepeerakul, R., Bertuzzo, E., Lynch, H. J., Fagan, W. F., Rinaldo, A., & Rodriguez-Iturbe, I. (2008). Neutral metacommunity models predict fish diversity patterns in Mississippi-Missouri basin. *Nature*, 453(7192), 220–222. <https://doi.org/10.1038/nature06813>
- Munoz, F., Grenié, M., Denelle, P., Taudière, A., Laroche, F., Tucker, C., & Violle, C. (2018). ecolottery: Simulating and assessing community assembly with environmental filtering and neutral dynamics in R. *Methods in Ecology and Evolution*, 9(3), 693–703. <https://doi.org/10.1111/2041-210X.12918>
- Novack-Gottshall, P. (2020). *ecospace: Simulating Community Assembly and Ecological Diversification Using Ecospace Frameworks*. R package version 1.4.2. Retrieved from <https://CRAN.R-project.org/package=ecospace>
- O'Dwyer, J. P., & Green, J. L. (2010). Field theory for biogeography: A spatially explicit model for predicting patterns of biodiversity. *Ecology Letters*, 13(1), 87–95. <https://doi.org/10.1111/j.1461-0248.2009.01404.x>
- Ostling, A. (2005). Neutral theory tested by birds. *Nature*, 436(7051), 635–636. <https://doi.org/10.1038/436635a>
- Pereira, H. M., Borda-De-Água, L., & Santos Martins, I. (2012). Geometry and scale in species-area relationships. *Nature*, 482(7386), 368–371. <https://doi.org/10.1038/nature10857>
- Pinheiro, H. T., Bernardi, G., Simon, T., Joyeux, J.-C., Macieira, R. M., Luiz Gasparini, J., ... Rocha, L. A. (2017). Island biogeography of marine organisms. *Nature*, 549(7670), 82–85. <https://doi.org/10.1038/nature23680>
- Preston, F. W. (1960). Time and space and the variation of species. *Ecology*, 41(4), 611–627. <https://doi.org/10.2307/1931793>
- Rosindell, J., & Cornell, S. J. (2007). Species-area relationships from a spatially explicit neutral model in an infinite landscape. *Ecology Letters*, 10(7), 586–595. <https://doi.org/10.1111/j.1461-0248.2007.01050.x>
- Rosindell, J., & Harmon, L. J. (2013). A unified model of species immigration, extinction and abundance on islands. *Journal of Biogeography*, 40(6), 1107–1118. <https://doi.org/10.1111/jbi.12064>
- Rosindell, J., Hubbell, S. P., & Etienne, R. S. (2011). The unified neutral theory of biodiversity and biogeography at age ten. *Trends in Ecology & Evolution*, 26(7), 340–348. <https://doi.org/10.1016/j.tree.2011.03.024>
- Rosindell, J., & Phillimore, A. B. (2011). A unified model of island biogeography sheds light on the zone of radiation. *Ecology Letters*, 14(6), 552–560. <https://doi.org/10.1111/j.1461-0248.2011.01617.x>
- Rosindell, J., Yan, W., & Etienne, R. S. (2008). A coalescence approach to spatial neutral ecology. *Ecological Informatics*, 3(3), 259–271. <https://doi.org/10.1016/j.ecoinf.2008.05.001>
- Shem-Tov, Y., Danino, M., & Shnerb, N. M. (2017). Solution of the spatial neutral model yields new bounds on the Amazonian species richness. *Scientific Reports* 7, 1–7. <https://doi.org/10.1038/srep42415>
- Tang, J., & Zhou, S. (2013). Hybrid niche-neutral models outperform an otherwise equivalent neutral model for fitting coral reef data. *Journal of Theoretical Biology*, 317, 212–218. <https://doi.org/10.1016/j.jtbi.2012.10.019>
- Taudière, A., & Violle, C. (2016). cati: An R package using functional traits to detect and quantify multi-level community assembly processes. *Ecography*, 39(7), 699–708. <https://doi.org/10.1111/ecog.01433>
- Thompson, S. E. D., Chisholm, R. A., & Rosindell, J. (2019). Characterizing extinction debt following habitat fragmentation using neutral theory. *Ecology Letters*, 22(12), 2087–2096. <https://doi.org/10.1111/ele.13398>
- Thompson, S. E. D., Chisholm, R. A., & Rosindell, J. (2020a). Source code for pycoalescence from: Pycoalescence and rcoalescence: Packages for simulating spatially explicit neutral models of biodiversity. *Zenodo*, <https://doi.org/10.5281/zenodo.3924587>
- Thompson, S. E. D., Chisholm, R. A., & Rosindell, J. (2020b). Source code for rcoalescence from: Pycoalescence and rcoalescence: Packages for simulating spatially explicit neutral models of biodiversity. *Zenodo*, <https://doi.org/10.5281/zenodo.3924585>
- Thompson, S. E. D., Chisholm, R. A., & Rosindell, J. (2020c). Example code for pycoalescence from: Pycoalescence and rcoalescence: Packages for simulating spatially explicit neutral models of biodiversity. *Zenodo*, <https://doi.org/10.5281/zenodo.3924567>
- Thompson, S. E. D., Chisholm, R. A., & Rosindell, J. (2020d). Example code for rcoalescence from: Pycoalescence and rcoalescence: Packages for simulating spatially explicit neutral models of biodiversity. *Zenodo*, <https://doi.org/10.5281/zenodo.3924564>

- Vallade, M., & Houchmandzadeh, B. (2003). Analytical solution of a neutral model of biodiversity. *Physical Review E*, 68(6). <https://doi.org/10.1103/PhysRevE.68.061902>
- Van Der Gast, C. J., Lilley, A. K., Ager, D., & Thompson, I. P. (2005). Island size and bacterial diversity in an archipelago of engineering machines. *Environmental Microbiology*, 7(8), 1220–1226. <https://doi.org/10.1111/j.1462-2920.2005.00802.x>
- Vellend, M. (2010). Conceptual synthesis in community ecology. *The Quarterly Review of Biology*, 85(2), 183–206. <https://doi.org/10.1086/652373>
- Volkov, I., Banavar, J. R., Hubbell, S. P., & Maritan, A. (2003). Neutral theory and relative species abundance in ecology. *Nature*, 424(6952), 1035–1037. <https://doi.org/10.1038/nature01883>
- Volkov, I., Banavar, J. R., Hubbell, S. P., & Maritan, A. (2007). Patterns of relative species abundance in rainforests and coral reefs. *Nature*, 450(7166), 45–49. <https://doi.org/10.1038/nature06197>
- Wilson, M. C., Chen, X.-Y., Corlett, R. T., Didham, R. K., Ding, P., Holt, R. D., ... Yu, M. (2016). Habitat fragmentation and biodiversity

conservation: Key findings and future challenges. *Landscape Ecology*, 31(2), 219–227. <https://doi.org/10.1007/s10980-015-0312-3>

#### SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article:** Thompson SED, Chisholm RA, Rosindell J. pycoalescence and rcoalescence: Packages for simulating spatially explicit neutral models of biodiversity. *Methods Ecol Evol.* 2020;00:1–10. <https://doi.org/10.1111/2041-210X.13451>