`MultiSPA`: simulating and inferring multispecies spatial patterns with given species associations

Chuliang Song[1,2], Marie-Josée Fortin[2], Andrew Gonzalez[1]

[1] Department of Biology, Quebec Centre for Biodiversity Science,
McGill University, Montreal, Canada H3A 1B1

[2] Department of Ecology and Evolutionary Biology,
University of Toronto, Toronto, Canada M5S 3B2

Abstract

Species distributions are highly associated (). Previous models are mostly focus. However, most null models are either spatially implicit or generate only one type of spatial association. We propose a simple spatially-explicit model that generates arbitraty intraspecific and interspecific associations.

Here, we propose a simple multispecies point process. The computational complexity of our method is linear with respect to the number of species or the number of points.

The availability of this tool allows rigorous test of ad hoc methods. For example, we show that t

Thus, our method is mathematically rigorous, ecologically intuitive, easily extendable, and computationally feasible.

1 Introduction

Species are not randomly distributed in a landscape. Instead, species are highly spatially associated with intraspecific/interspecific attraction/repulsion. For example, interspecific repulsion, also known as "checkboard distribution", occurs when ecologically similar species tend to geographically intersperse and mutually exclude (MacArthur *et al.*, 1972; Diamond, 1975).

To understand these spatial patterns, ecologists have presented a suite of null models. We can classify these into three broad classes: species distribution model, co-occurence model, and spatial point processes. The species distribution model usually focuses on the inference problem rather than generate a null model (Pichler & Hartig, 2021). The co-occurence model is not spatially explicit.

Thanks to the rise of remote sensing technology (Turner, 2014), mapping the locations of individual species has become much easier than tranditional census.

However, most models of the multivariate point processes are direct extension of single-variate point process. This includes. Thus, we lack a simple

mechanistic models to model the processes (Brown *et al.*, 2011). We take a phenomenological approach. T

On the one hand, we have a great list of computational tools on empirical .

The null models for spatial patterns of a *single* species are well-developed. Apart from the foundational Poisson process, some notable examples include Cox (Cox, 1955), Gibbs (), area-interaction point process (Baddeley & Van Lieshout, 1995). However, quite surprisingly, we lack a library of null models for *multiple* species. To fairly compare species, we need the ability

Apart from spatial point process,

An exception is (Marsh & Ewers, 2013). However, their method only considers the repulsion and only focus on the nearest neighboring discrete cells. Thus, we currently lack a null model to generate multiple spatial pattern constrained by species associations.

Mathematically, this corresponds to on two-dimensional $\mathbb{R}^2$.

There are two ways. Multivariate spatial process and Marked point process (e.g., Markov random field) (Besag, 1974; Rue & Held, 2005).

To fill this gap, here we develop a mathematically rigorous, ecologically intuitive, and computationally feasible null model. To implement the model, we have developed a new R package `MultiSPA`, standing for **Multi**species **S**patial **P**attern constrained by species **A**ssociations.

2 Methods

2.1 A model of multivariate spatial process

Here, we build a null model to generate spatial patterns given arbitrary constraints of pairwise association (attractive, null, or repulsive) for multiple species. The idea is to sample one individual at one time (a MCMC process). The probability of where the individual appears is uneven across the landscape and is determined by the neighboring individuals (similar to the Thomas process).

Focusing on a species pair, we use an association measure η to quantify the level of spatial relationship between two species. The association measure η ranges between -1 and 1: If $\eta \in (0, 1)$, then the two species are attractive to each other (more positive value denotes stronger attraction); If $\eta \in (-1, 0)$, then the two species are repulsive to each other (more negative value denotes stronger repulsion); If $\eta = 0$, then the two species are spatially unrelated. For the whole landscape with S species, we summarize all the pairwise information into an association matrix $\boldsymbol{\eta}$. $\boldsymbol{\eta}$ has dimension $S \times S$, and its element η_{ij} denotes the spatial association between species i and j. Note that η_{ij} is not necessarily equal to η_{ji}.

To make the association into a mathematically meaningful likelihood, we transform η to $\gamma(\eta)$:

$$\gamma(\eta) = \begin{cases} \tan\left(\frac{\pi\eta}{2}\right) + 1 & \eta \in [0, 1), \\ \frac{1}{1-\tan\left(\frac{\pi\eta}{2}\right)} & 0 \leq x \leq (-1, 0) \end{cases} \tag{1}$$

Figure S1 plots this transformation function.

We also need a species-specific radius R_i. This radius denotes the threshold of neighboring individuals that have an effect on the establishing probability of new individual. If the radius R_i is sufficiently large, then all individuals on the whole landscape affects the establishing probability. If the radius R_i is too small, then no existing individuals would affect the establishing probability.

Then the probability of a new individual is given by

$$p(x_i) \propto \prod_j \gamma(\eta_{ij})^{\sum_{y \in \text{species } j} \mathbb{1}\|x_i - y\| \leq R_i} \tag{2}$$

where x_i denotes a new individual of the species i, $\gamma(\eta_{ij})$ denote the base level of attraction or repulsion between species i and j, and $\sum_{y \in \text{species } j} \mathbb{1}\|x_i - y\| \leq R_i$ denotes the number of species j within the radius R_i. Figure 1 illustrates the process.

4

Finding the posterior probability is generally difficult because the normalization function is unclear. To overcome this problem, we first use a Poisson point process to sample enough points as candidates, and then we compute the establishing probability for all the candidates, and finally we choose a candidate according to the posterior probability.

2.2 Extensions of the basic method

The basic method is easily extendible. One extension is to consider the environmental heterogeneity. We can define a species-specific environmental fitness E_x^i, which denotes the environmental fitness that species i is at the location x. Then the establishing probability becomes

$$\log p(x_i) \propto E_x^i \sum_j \left(\sum_{y \in \text{species } j} \mathbb{1} \|x_i - y\| \leq R_i \right) \log(\gamma(\eta_{ij})) \tag{3}$$

Another extension we can consider is the maximum effective neighbors M_i. The introduction of this cap would help create multiple patches. Then the establishing probability becomes

$$\log p(x_i) \propto E_x^i \sum_j \left(\max\{ \sum_{y \in \text{species } j} \mathbb{1} \|x_i - y\| \leq R_i, M_i \} \right) \log(\gamma(\eta_{ij})) \tag{4}$$

To do: Zone-of-influence model and ecological field model.

2.3 Empirical application

3 Results

3.1 Simulation examples

An advantage of the method is to control for both the interspecific spatial associations and also the intraspecific ones. For two species, there are in total 9 cases:

We can easily generate arbitrarily complicated cases for multiple species. For example, Figure 3 illustrates a complicated case for 4 species.

We can also generate asymmetric spatial associations. For example, Figure **??** illustrates a two-species case where the establishment of species A is attracted to existing species B, while the establishment of species B is repulsive to existing species A.

3.2   Empirical example

4   Discussion

Our method is advantageous compared to traditional methods in the following aspects. First, our method can simultaneously model attraction and repulsion. Second, our method can consider asymmetric association (the effect of species i on j is different from the effect of species j on i). Third, our method is computationally fast (linear complexity).

Our method allows for rigorous tests of spatial inference methods. As a proof of example, we show that the multitype K Function could only detect when there are only two species, but produce spurious results when there are more than 2 species.

There are many future directions.

The Similar to species interaction matrix in community ecology.
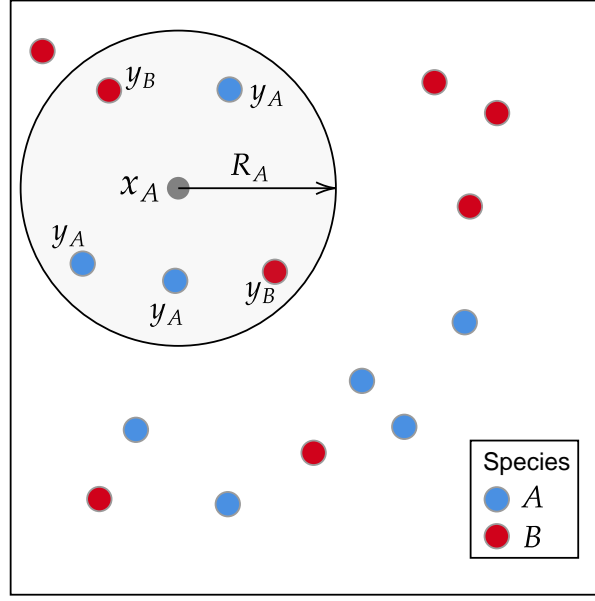
Figure 1: Illustration of proposed multispecies model. Species A is colored in red while species B is colored in blue. In the illustrated case, $P(x_A) \propto \gamma(\eta_{AA})^2 \gamma(\eta_{AB})^3$.
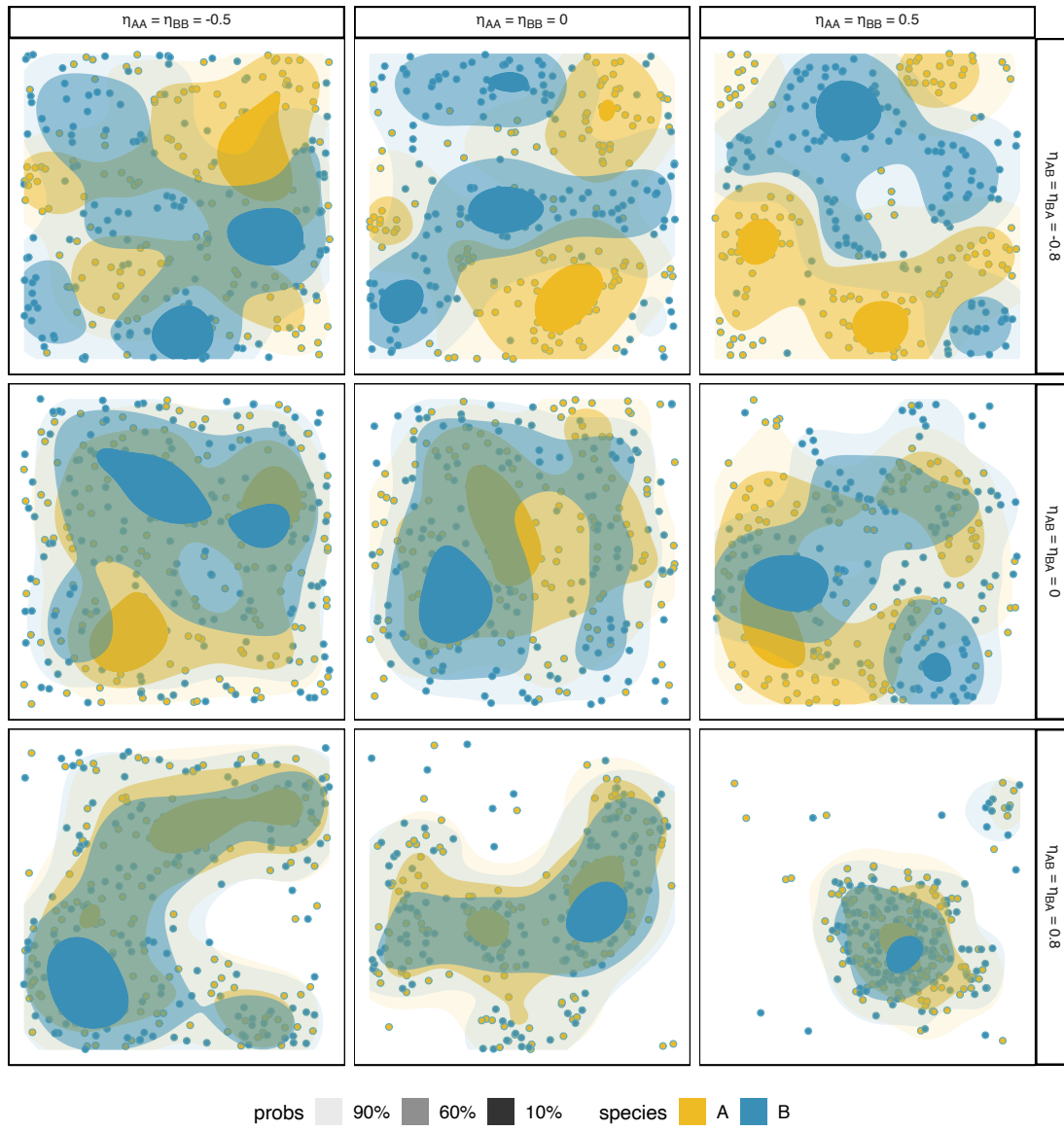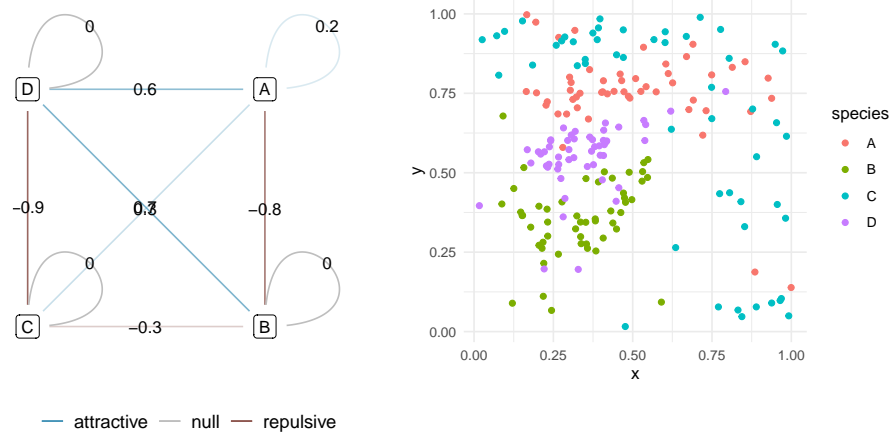
Figure 2: Illustration for 2 species.

Figure 3: Illustration for 4 species.

References

Baddeley, A.J. & Van Lieshout, M. (1995) Area-interaction point processes. *Annals of the Institute of Statistical Mathematics*, **47**, 601–619.

Besag, J. (1974) Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B (Methodological)*, **36**, 192–225.

Brown, C., Law, R., Illian, J.B. & Burslem, D.F. (2011) Linking ecological processes with spatial and non-spatial patterns in plant communities. *Journal of ecology*, **99**, 1402–1414.

Cox, D.R. (1955) Some statistical methods connected with series of events. *Journal of the Royal Statistical Society: Series B (Methodological)*, **17**, 129–157.

Diamond, J.M. (1975) Assembly of species communities. *Ecology and evolution of communities*, pp. 342–444.

MacArthur, R.H., Diamond, J.M. & Karr, J.R. (1972) Density compensation in island faunas. *Ecology*, **53**, 330–342.

Marsh, C.J. & Ewers, R.M. (2013) A fractal-based sampling design for ecological surveys quantifying β-diversity. *Methods in Ecology and Evolution*, **4**, 63–72.

Pichler, M. & Hartig, F. (2021) A new joint species distribution model for faster and more accurate inference of species associations from big community data. *Methods in ecology and evolution*, **12**, 2159–2173.

Rue, H. & Held, L. (2005) *Gaussian Markov random fields: theory and applications*. Chapman and Hall/CRC.

Turner, W. (2014) Sensing biodiversity. *Science*, **346**, 301–302.

Supplementary Material for

SpatMultiSpec: multispecies spatial patterns constrained by spatial associations

Contents

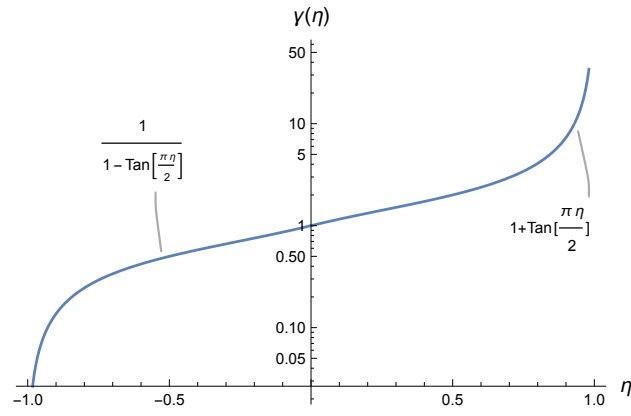# A   Transformation function

Figure S1: Function of $\gamma(\eta)$.

B Examples of the extended model

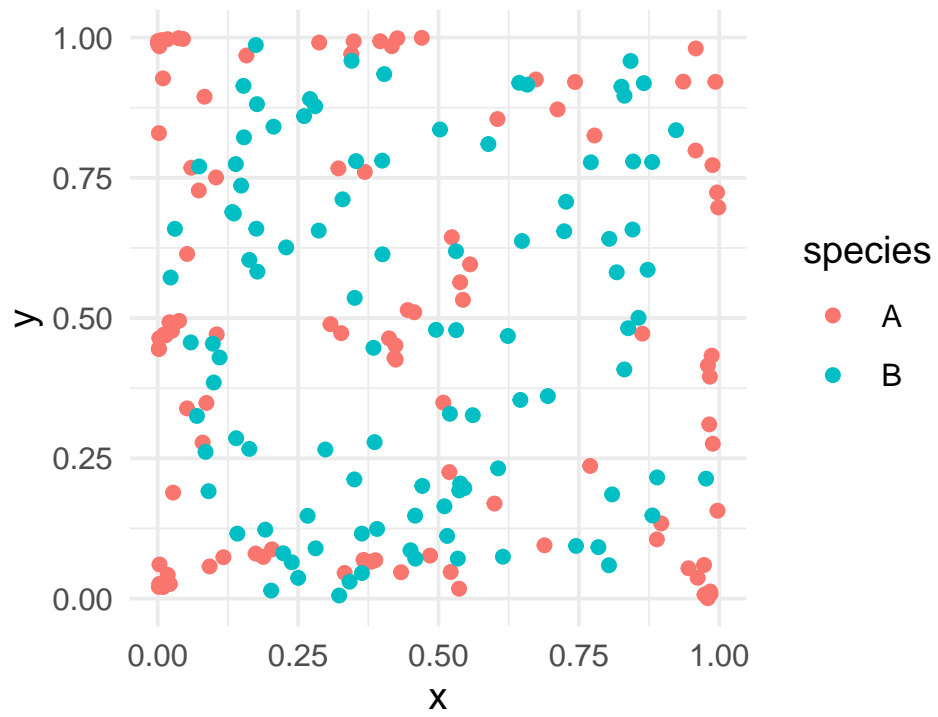
Figure S2: Illustration for 2 species with asymmetric association.

C Computational performance
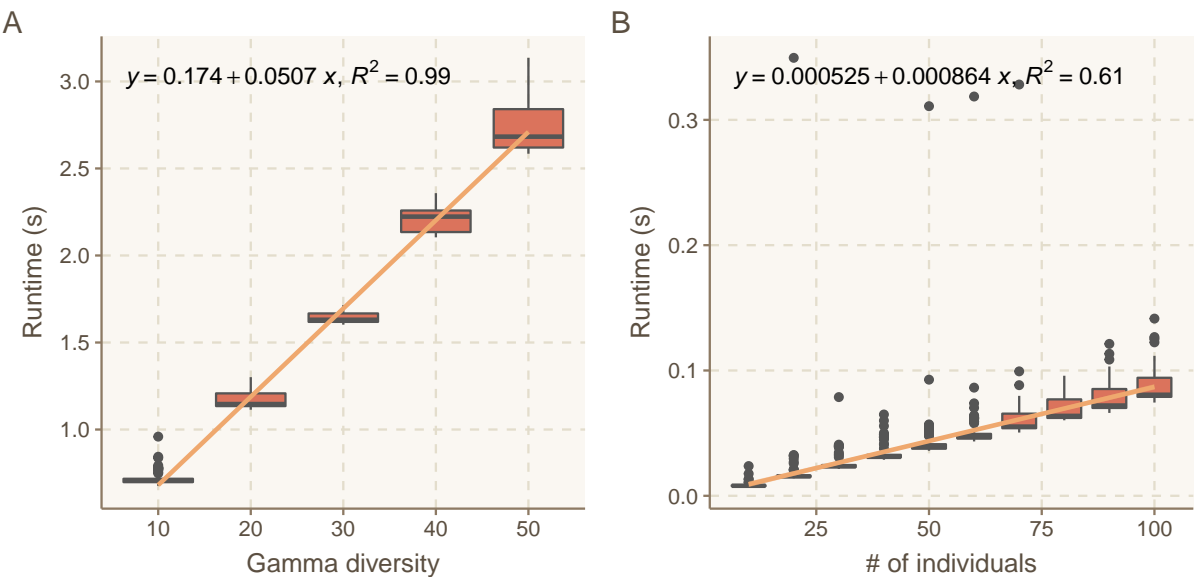
The computational complexity $\sim \mathcal{O}(N)$

A

$y = 0.174 + 0.0507\, x, R^2 = 0.99$

Runtime (s): 3.0, 2.5, 2.0, 1.5, 1.0

Gamma diversity: 10, 20, 30, 40, 50

B

$y = 0.000525 + 0.000864\, x, R^2 = 0.61$

Runtime (s): 0.3, 0.2, 0.1, 0.0

# of individuals: 25, 50, 75, 100

Figure S3: xx.

References