Linking ecological processes with spatial and non-spatial patterns in plant communities

Calum Brown1*, Richard Law2, Janine B. Illian1 and David F. R. P. Burslem3

1CREEM, School of Mathematics and Statistics, University of St Andrews, The Observatory, Buchanan Gardens, St Andrews KY16 9LZ, UK; 2Biology Department, University of York, PO Box 373, York YO1 5YW, UK; and 3Institute of Biological and Environmental Sciences, University of Aberdeen, Cruickshank Building, St Machar Drive, Aberdeen AB24 3UU, UK

Summary

1. Attempts to infer underlying ecological process from observed patterns in ecology have been widespread, but have generally relied on first-order (non-spatial) community characteristics such as the species abundance distribution (SAD). This measure has become an important test of several theories of species coexistence, but has proved unsuccessful in distinguishing between them.

2. Spatially explicit data are increasingly available for a range of ecological communities, and analysis methods for these data are well developed. However, relatively little work has investigated the potential of these data for similar inference about mechanisms of coexistence.

3. In this study, we systematically investigate the spatial and non-spatial signals of simulated ecological processes. We include neutral, niche, lottery, Janzen–Connell and heteromyopia models, deriving and comparing first- and second-order measures for the patterns they generate.

4. We find that the SAD is unable to distinguish reliably between underlying models, with random variation in its shape concealing any systematic differences.

5. A new second-order summary measure of spatial structure derived in this paper, in contrast, proves highly sensitive to the type of ecological interaction being modelled, and is robust to random variation.

6. Synthesis. A simple summary measure of the spatial structure of plant communities is presented and found to be a more powerful indicator of underlying process in simulated data than a widely used first-order measure, the SAD. The potential for answering important ecological questions using spatial statistics, particularly concerning mechanisms of coexistence in diverse communities, appears to be great.

Key-words: coexistence mechanisms, determinants of plant community diversity and structure, heteromyopia, Janzen–Connell, lottery, neutral, niche, pair correlation function, spatial point pattern, species abundance distribution

Introduction

A great deal of research in ecology tries to infer ecological processes from patterns observed in nature. In community ecology, the species abundance distribution (SAD) has received particular attention (McGill et al. 2007). A SAD describes the absolute or relative abundances of species in a community and is found to conform to a near-universal ‘hollow curve’ shape, comprising a small number of common species and a large number of rare ones. There is no obvious a priori reason to expect this shape, and the detailed features of SADs have therefore been used to discriminate between underlying processes, such as those involved in species coexistence. This work began with theories of niche assembly (e.g. Motomura 1932; MacArthur 1957; Tokeshi 1990), and more recently comparing SADs has become a key tool for validating the neutral theory against observed data from ecological communities (Hubbell 1979, 1997, 2001).

Unfortunately, a single ecological process can produce rather variable SADs (Williamson & Gaston 2005; Magurran 2005; Volkov et al. 2005), making the detection of processes from empirically derived SADs difficult (McGill et al. 2007). This is not surprising: a SAD is, after all, just a description of species’ relative abundances averaged over space. Processes affecting coexistence rely on spatial proximity of individuals, especially in sessile organisms, and SADs convey...
no information on spatial structure. In the context of a spatial analysis of communities, a SAD would be said to be a first-order measure (Iliian et al. 2008).

In principle, spatial correlations ought to provide a more sensitive indicator of ecological interactions among plant species because of the importance of interactions as drivers of spatial pattern in plant communities (Bolker & Pacala 1997; Murrell & Law 2003; Wiegand et al. 2007). There is a long history in plant ecology of using spatial patterns to gain insight into ecological processes (e.g. Watt 1947; Clark & Evans 1954; Sterner, Ribic & Schatz 1986), and indeed this was one motivation for the development of spatial point process methods (Matérn 1960; Ripley 1977; Stoyan & Penttinen 2000). It would be unrealistic to expect a unique mapping from a spatio-temporal process to a spatial pattern because of the array of biotic and abiotic factors at play (e.g. Baddeley & Silverman 1984; Leps 1990), but it is reasonable to ask whether an analysis that makes use of spatial structure is a better discriminator among ecological mechanisms than one based on SADs that ignores this information.

Ecologists do often have far more information at their disposal than just that needed to construct SADs. For example, several complete spatial censuses exist for tropical rainforest trees on the 50-ha plot at Barro Colorado Island (BCI) in Panama (Hubbell, Condit & Foster 2005) and numerous other sites (Losos & Leigh 2004). In the search for evidence about underlying processes on these plots, it should be possible to go beyond SADs (e.g. Hubbell 2001; Volkov et al. 2003; Etienne & Oliff 2005; He 2005) to second-order measures such as spatial correlations that make use of this spatial information. The potential of these has been recognized in studies of the roles of seed dispersal and habitat heterogeneity (Condit et al. 2000; John et al. 2007), the aggregations produced by neighbourhood recruitment and mortality (Hubbell et al. 2001; Uriarte et al. 2005), and spatial patterns in diversity (Wiegand, Gunatilleke & Gunatilleke 2007). Elsewhere, temporal patterns have been used to discriminate between neutral and non-neutral mechanisms for maintaining the structure and diversity of communities (e.g. Clark & McLachlan 2003; McGill, Hadly & Maurer 2005), as have comparisons between different spatial scales (Gilbert & Lechowicz 2004; Dornelas, Connolly & Hughes 2006; McGill, Maurer & Weiser 2006). However, a systematic analysis of the spatial signatures generated by different kinds of species interaction has not been attempted.

This study evaluates the effectiveness of SADs and measures of spatial correlations to discriminate between multispecies spatial patterns that make different underlying assumptions about ecological interactions. Our baseline was the neutral model, with its assumption of per capita ecological equivalence between species (Hubbell 2001). Two niche models were included: a conventional niche model in which species favour specific environmental conditions that are defined spatially (e.g. Grinnell 1917; Hutchinson 1958; Zilio & Condit 2007), and a lottery model in which temporal environmental variance favours different species at different times (Sale 1977; Chesson & Warner 1981; Chesson & Hurlst 1988). The Janzen–Connell hypothesis, according to which young plants suffer increased mortality in the neighbourhood of their parents, was also implemented (Janzen 1970; Connell 1970), as was a purely spatial heteromyopia model in which interspecific competition occurs over shorter distances than intraspecific competition (Murrell & Law 2003).

We generated multispecies spatial patterns through realisations of spatio-temporal stochastic processes [stochastic individual-based models (IBMs)] using the different underlying models of ecological interactions. At first order, we computed SADs on these spatial patterns. At second order, we computed a new community-level measure of species segregation, built from spatial pair-correlation functions, referred to as the cross-pair overlap distribution (xPOD).

First-order signals of the modelled ecological interactions were expected to be limited, given the inherent variability of SADs. Some differences in community diversity and evenness were anticipated, but were difficult to predict because the relative strength of each form of interaction in promoting coexistence has not previously been assessed. Second-order spatial signals were expected to be substantially stronger, and to take a more predictable form for each model. In particular, the spatial niche model was expected to increase the segregation among species, while the Janzen–Connell model was predicted to constrain conspecific clumping and so have the opposite effect.

Materials and methods

STOCHASTIC PROCESS FOR MULTISPECIES SPATIAL PATTERNS

Multispecies spatial patterns were obtained from realisations of a stochastic IBM of a plant community based on a method in Law & Dieckmann (2000). In this setting, individuals occur at discrete points \( x = (x_1, x_2) \), \( x_1, x_2 \in [0, 1] \) in a continuous two-dimensional space. The space comprises an arena of unit area with periodic boundaries, so forming a torus and preventing the inward propagation of edge effects. The spatial pattern \( p(x,t) \) of a species \( s \) at time \( t \), \( t \in \mathbb{R}^+ \), comprises the locations of all individuals of species \( s \), and the multispecies pattern \( p(x,t) \) is the union of all these single-species patterns. Birth and death events take place in continuous time, together with occasional arrival of new species, so the multispecies spatial pattern changes at every event. The effect of particular assumptions about ecological interactions on the multispecies spatial pattern is investigated after a large number of birth and death events have taken place.

For simplicity, the birth process is common to all species, independent of location in the arena, and comprises an intrinsic probability per unit time \( b \) of producing an offspring, and a function \( n(x - x') \) giving the probability of the offspring being located at \( x' \) for a parent at \( x \). Thus, the probability per unit time \( B(x, x') \) of a parent at \( x \) producing an offspring at \( x' \) is

\[
B(x, x') = b n(x - x')
\]  
eqn 1

In a community of finite size, there would be a gradual erosion of species diversity as extinctions take place. To counter this, a low constant probability per unit time of immigration of a new species is assumed, and the location of the new individual is chosen uniformly at random in the arena.
Here, \( d_i \) is an intrinsic probability per unit time of death of species \( i \). The term inside the integral takes a neighbour of species \( j \) located at \( x' \), and attaches weight \( w_j(x' - x) \), which depends on the displacement \( x' - x \) of the neighbour from the target individual of species \( i \) located at \( x \), allowing the distance over which individuals interact to depend on the species identity of the target and neighbour. The integral adds up the effect of all neighbours of species \( j \) on the target, and is given a weight \( d'_{ij} \) so that interaction strength can also depend on the species identities. The summation adds the effects over all neighbour species \( j \). In the case where \( j = i \), the neighbours are conspecifics and the product \( \delta_i \delta_j(x') \), a Kronecker delta and Dirac delta function, is needed so that the target individual is not counted as a neighbour of itself.

Spatial structure in the community comes ultimately from local dispersal and local competition, assumed to be bivariate Gaussian functions

\[
m(x - x') = \frac{1}{2\pi \sigma_x \sigma_y} \exp \left( -\frac{|x - x'|^2}{2 \sigma_x \sigma_y} \right), \tag{3}
\]

\[
w_j(x' - x) = \frac{1}{2\pi \sigma_{dx} \sigma_{dy}} \exp \left( -\frac{|x - x'|^2}{2 \sigma_{dx} \sigma_{dy}} \right), \tag{4}
\]

normalized to make the volume = 1. The parameters \( \sigma_x \) and \( \sigma_y \) give the width of the distributions; small values indicate, respectively, that offspring are dispersed close to their parent and that competition occurs with close neighbours.

**ECOLOGICAL MODELS FOR SPECIES COEXISTENCE**

**Neutral model**

This is the simplest model and provides a standard against which to test the others. It was implemented by making the parameters \( b, d, \sigma_b \) and \( \sigma_{d_i} \) constant and identical for every individual irrespective of species (Table 1).

**Spatial niche model**

This differs from the neutral model in that species are sensitive to a spatially defined environmental variable. We use a simple algorithm to construct a landscape based on a circular distribution divided into quadrants, producing a ‘peak’ in the centre of the arena \( x = (0.5, 0.5) \) and identical values at the periodic boundaries. The circular distribution in each quadrant is truncated at half its radius and the range within the arena scaled to the interval \([0, 1] \). The value or ‘height’ of the environment at location \( x, x = (x_1, x_2) \) is therefore:

\[
h_x = \begin{cases} 
\frac{2}{\pi} \sqrt{1 - |x_1 - 0.5|^2} \quad |x_1 - 0.5| \geq |x_2 - 0.5| \\
\frac{2}{\pi} \sqrt{1 - |x_2 - 0.5|^2} \quad \text{otherwise}
\end{cases} \tag{5}
\]

This produces an environment in which the spatial extent of each specific value, and hence the size of each potential niche, is approximately equal, ensuring that any spatial signal produced is attributable to the niche process rather than environmental availability (see Fig. 1d).

The spatial niche model requires species to respond to the environment in different ways, so the niches of species are set by environment-dependent intrinsic death rates. Each species \( i \) is assigned a uniformly distributed random number in the range \([0, 1]\) to give it a preferred niche \( h^0_i \). An individual in this optimal environment has a death probability \( d_i \) as in the neutral model. The dependence of the death rate on the environment is made explicit here by writing it as \( d_i(x) \) for an individual of species \( i \) located at a point \( x \) where the environment has a ‘height’ \( h_x \). The value of \( d_i(x) \) is assumed to have an inverted Gaussian shape around the birth rate \( b \), determined by the deviation \( h_x - h^0_i \) from the species’ preferred niche, so that \( d_i(x) \) is minimized at \( h^0_i \):

\[
d_i(x) = b - \frac{(b - d_i)}{\sqrt{2\pi \sigma_b^2}} \exp \left( -\frac{(h_x - h^0_i)^2}{2\sigma_b^2} \right) \tag{6}
\]

This means that, as the environment departs further from the preferred niche, the death rate of species \( i \) increases, eventually becoming the same as its birth rate and preventing it from becoming established. The parameter \( \sigma_b \) is a niche-width parameter common to all species. Thus, as \( \sigma_b \) becomes large, species do not perceive the landscape and the dynamics are as in the neutral model. Decreasing \( \sigma_b \) makes the success of each species increasingly dependent upon it being within its preferred niche \( h^0_i \).

**Table 1.** Parameters for ecological models; boldface values indicate differences from the neutral model. J–C denotes the Janzen–Connell model

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameter</th>
<th>Parameter</th>
<th>Parameter</th>
<th>Parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( b )</td>
<td>( d_i )</td>
<td>( d''_{ii} )</td>
<td>( d''_{ij} )</td>
</tr>
<tr>
<td>Neutral</td>
<td>0.4</td>
<td>0.2</td>
<td>( 4 \times 10^{-6} )</td>
<td>( 4 \times 10^{-6} )</td>
</tr>
<tr>
<td>Spatial niche</td>
<td>0.4</td>
<td>(0.2–0.4)*</td>
<td>( 4 \times 10^{-6} )</td>
<td>( 4 \times 10^{-6} )</td>
</tr>
<tr>
<td>Lottery</td>
<td>0.4</td>
<td>( U(0.1, 0.3) )</td>
<td>( 4 \times 10^{-6} )</td>
<td>( 4 \times 10^{-6} )</td>
</tr>
<tr>
<td>J–C</td>
<td>0.4</td>
<td>0.2</td>
<td>( 1.6 \times 10^{-5} )</td>
<td>( 4 \times 10^{-6} )</td>
</tr>
<tr>
<td>Heteromorphia</td>
<td>0.4</td>
<td>0.2</td>
<td>( 4 \times 10^{-6} )</td>
<td>( 4 \times 10^{-6} )</td>
</tr>
</tbody>
</table>

Parameters: \( b \), intrinsic birth probability per unit time; \( d_i \), intrinsic death probability per unit time; \( d''_{ii} \), weighting for within-species density-dependent deaths; \( d''_{ij} \), weighting for between-species density-dependent deaths; \( \sigma_b \), dispersal kernel standard deviation; \( \sigma_{d_i} \), within-species density-dependent death kernel standard deviation; \( \sigma_{d_{ij}} \), between-species density-dependent death kernel standard deviation. *Environment \( h_x \) defined in the text; preferred niche for species \( i \): \( h^0_i \in [0, 1] \); niche width: \( \sigma_b = 0.2 \).

Lottery (temporal niche) model

This model differs from the neutral model in that the intrinsic death rates $d_i$ of species are drawn from a uniform distribution centred on that of the neutral model. The environment has a low, constant probability per unit time of changing, set at $5 \times 10^{-3}$. When the environment changes, new values of $d_i$ for each species are drawn from the uniform distribution. As the range of the uniform distribution tends to zero, the behaviour tends to the neutral model; increasing the range makes the lottery effect stronger.

Janzen–Connell model

This model requires local dispersal of propagules, applied here when the parameter of the dispersal function is sufficiently small. It also requires a higher death rate in the presence of conspecific neighbours than in the presence of heterospecific ones, because of host-specific enemies. We do not model host-specific enemies, and so the Janzen–Connell mechanism is introduced by making the neighbourhood conspecific death term $d_i$, larger than the heterospecific one $d'$. In the natural environment, a distinction also would be made between the parent and the offspring, but we do not do this because age and size of individuals are not specified under the simple conditions of the stochastic process.

Heteromyopia model

This is a spatial mechanism for coexistence which requires shorter distances for interactions between species than within species. We introduce this mechanism by making the width parameter $a$ for interaction distances with conspecifics larger than the width parameter $b$ for interactions with heterospecifies.

Numerical values for simulations

Each realization starts with 50 species of 100 individuals (5000 individuals in total), distributed independently and uniformly at random in the arena. The spatial pattern is updated by birth and death events using the Gillespie algorithm (Gillespie 1977), following the rules of the stochastic process defined above with parameter values as in Table 1. The ‘Mersenne-twister’ pseudo-random number generator (Matsumoto & Nishimura 1998) is used throughout. A small amount of large-scale dispersal is introduced by placing newborn individuals of existing species uniformly at random in the arena with a low probability of 0.01. In addition to more accurately describing the variable dispersal mechanisms employed by rainforest trees, this element of random dispersal enables species to colonize distant areas of preferred environment in the spatial niche model.

Parameter values in the different models are controlled to generate simple and clearly distinguishable departures from neutrality. Many of the parameters do not affect spatial structure and are used to set the spatial and temporal scale of the simulations; these include the birth and death rates $b$ and $d$, and the density-dependent death rates $d_i$ and $d'$. Parameters controlling dispersal and density-dependence kernels are set to allow fine-scale behaviour but also some spatial mixing on the scale of the arena. These are common to all species under neutrality and are varied only where necessary in other models (Table 1).

A low probability per unit time of an immigrant birth, 0.001, is allowed to counter extinctions of species that happen naturally as a result of the birth–death process in the finite arena. Immigrant species are drawn from an infinite pool, and consequently immigrations by existing species do not occur. In the lottery and niche models, immigrant species are randomly assigned death rates or preferred niches, as above; in the other models, their characteristics match those of existing species, incorporating Janzen–Connell or heteromyopia effects where appropriate. After $10^6$ events, the realization is stopped and the emergent multispecies spatial pattern is used to test for differences between ecological models of interactions. We checked to make sure that this allows sufficient time for the realization to reach a stationary distribution in terms of both the total number of individuals and the number of species (e.g. Fig. 1a,b). The characteristics of the spatial patterns were also found to be stable by this point, with no systematic temporal changes occurring. For each model, 10 independent spatial patterns are generated, and are referred to as replicates below.

Cross-pair overlap distribution

We define a new community-level measure of the spatial overlap of species, based on the distribution of spatial overlaps of pairs of species. A measure aggregated to the community level is needed as it would be unmanageable to work with all species pairs separately in a multispecies community. The measure is built from the cross pair correlation function $g(r)$ (Diggle 2003; Illian et al. 2008; Appendix S1 in Supporting Information) of each species pair $i, j$, which is reduced to a scalar quantity, the area $A_{ij}$ obtained by integrating the log of the function. This describes the average overlap of species $i$ and $j$ up to a fixed distance $R$.

$$A_{ij} = \int_{0}^{R} \log(g(r))dr$$  eqn 7

An $A_{ij}$ close to zero implies that the two species are close to independent on average, up to $R$. A positive $A_{ij}$ implies that they tend to occur together, and a negative one that they are separated in space. Taking $A_{ij}$ of all non-self combinations of $i$ and $j$ gives a xPOD for the community as a whole. (We do not consider the self-pair overlap distribution, i.e. the case $i = j$, because it is the between-species spatial structure that is of concern here.) At the community level, an xPOD with predominantly positive values of $A_{ij}$ implies an overall tendency for species to occur together, and one with predominantly negative values implies a tendency for species to be separated.

Estimation of the cross-pair correlation functions $g(r)$ is based on the method given in Law et al. (2009). Following Baddeley & Turner (2005), an upper limit for estimation of $g(r)$ is set at $R = 0.25$, so that spatial behaviour of interest is not averaged out over larger areas. We chose this upper limit as being substantially greater than the spatial scale of the dispersal and competition kernels, but substantially smaller than the spatial scale of the arena. Tests with an upper limit of $R = 0.15$ gave similar results, but variation in observed behaviour with spatial scale could occur and warrants further investigation.

Estimation of $g(r)$ and $A_{ij}$ requires discretization of $r$, for which we use 10 bins of equal width. Comparison with results generated by up to 100 bins suggested that 10 are sufficient to capture the spatial behaviour without being unduly influenced by small-scale noise. The xPODs are constructed for species pairs in which both species have at least 500 individuals, so that random noise in spatial pattern from small samples does not mask the signal for the ecological interaction.

Community evenness and diversity

In addition to first- and second-order summaries of community structure, we derive mean indices of community evenness and diversity for
each model. The indices used are Pielou’s evenness index (Pielou 1966) and Shannon’s diversity index (Shannon & Weaver 1949). The larger the positive values of these, the greater the evenness and diversity of the community, respectively.

**STATISTICAL INFERENCE**

To compare SADs and xPODs generated by each simulation, we use the Kolmogorov-Smirnov (K-S) goodness-of-fit test statistic (Massey 1951), a nonparametric and distribution-free measure of the similarity of distributions. The comparison is made separately for the two measures so that their capacity to distinguish ecological models can be evaluated, and, in the case of SADs, at abundance thresholds of both 1 and 500 to ensure fair comparison with the xPOD. The K-S test statistic gives a measure of similarity of pairs of distributions; its value is calculated for every pair of SADs and xPODs in the study, giving $45 \times 5$ values within ecological models and $100 \times 10$ values between ecological models. At this broad level, the distribution of K-S values shows the relative magnitude of random differences within ecological models and systematic differences between ecological models. Comparisons within and between particular ecological models are also made by disaggregating the K-S values down to single ecological models. This is not intended as a method of identifying the underlying model, however, but simply as an expression of the magnitude and consistency of visible differences between distributions.

We additionally quantify random variability within models using prediction intervals for SADs and xPODs. These intervals provide, with a confidence of 95%, the range within which future observations will be found, given the observations already made. We chose prediction intervals instead of confidence intervals because the range of possible results is of more interest than the potential value of a mean.

**Results**

**ILLUSTRATIVE EXAMPLE**

To help understand the results that follow, we start by showing the statistics estimated from a single multispecies spatial pattern. To do this, we use the stochastic IBM with a spatial niche model. By the end of the simulation, when the spatial pattern was extracted for analysis, both the number of individuals and the number of species were close to steady state, as in all other realizations (Fig. 1a,b). This behaviour over time is typical of all the simulations of the IBM carried out. The SAD calculated from the emergent multispecies spatial pattern had the shape typically observed in these distributions, i.e. a relatively small number of common species and many rare ones (Fig. 1c).

To construct the xPOD from the multispecies spatial pattern, the spatial patterns of species were extracted in pairs. For the sake of illustration, the pair in Fig. 1d is chosen as one that exhibits clear spatial segregation. The correlation function $g(r)$ of this pair thus has values less than one at small radii, because it tends to lack pairs at short distances, and rises to close to one as radius increases and the density of pairs approaches the spatial average. Corresponding to this, $\log(g(r))$ is negative at short distances and increases to around zero as radius increases (Fig. 1e). The integral over $r$, $A_g$, sums the shaded areas in Fig. 1e, keeping track of the sign; this is also negative, and provides one datum for constructing the xPOD. Repeating this calculation for all pairs of species above the threshold abundance gives a distribution of $A_g$ – the xPOD of the spatial pattern (Fig. 1f). Although the xPOD has a centre near to zero, it has a tail of negative values because the spatial niche model leads to segregation of species with dissimilar niche requirements.

**SPECIES ABUNDANCE DISTRIBUTIONS**

Irrespective of the ecological model, SADs had the general property of a small number of abundant species and a large number of rare ones (Fig. 2). The most frequent abundance class was for species with just one individual, and these mainly comprised recent immigrants that would subsequently have become extinct.

Importantly, there were limited differences between the SADs arising from different ecological models (Fig. 2). Some small differences were apparent in community evenness, with the lottery model producing the least even and diverse communities on average, and the Janzen-Connell model producing the most even and diverse (Table 2). However, these differences were not consistently detected by the indices used and were small in comparison with the random variation in shape produced within each model, which was especially large in the lottery model.

This impression is confirmed by the distributions of K-S test statistics (Fig. 3). These indicate that variation between SADs was similar irrespective of whether the pairs were drawn from replicate realisations within ecological models or drawn from different ecological models. Disaggregating down to particular ecological models still led to mean values of the K-S test statistic for pairs of SADs from replicates within single models as large as those for pairs of SADs taken from different models (Table 3).

Species of low abundance were one cause of variation between SADs, and the mean values of the K-S test statistic for pairs of SADs within some models were reduced when restricted to species with at least 500 individuals; the same abundance threshold applied to the xPODs (Table 4). Although variation among SADs was greater between than within ecological models when using this threshold, the magnitude of the difference was small. Hence, distributions generated using the same model were not consistently more similar than those generated by different models. The use of this abundance threshold did not therefore affect our ability to distinguish underlying models using SADs.

**CROSS-PAIR OVERLAP DISTRIBUTIONS**

In contrast to the SADs, xPODs produced by different ecological models were consistently distinct, and their prediction intervals showed systematic differences between models that far outweighed random variation between replicates within models (Fig. 4). The degree of overlap between species at the community level was ordered: spatial
niche < lottery < heteromyopia/neutral < Janzen–Connell. This can be seen in the first three moments of the xPODs showing an increasing (though always negative) mean value, a declining standard deviation, and a broadly declining skewness (Table 5). The only exceptions were the neutral and heteromyopia xPODs which were almost entirely coincident.

Goodness-of-fit test statistics showed that xPODs generated by the same model were far more alike than the equivalent SADs, while those produced by different models were far more dissimilar (Fig. 5). Disaggregating down to particular ecological models, the neutral, niche and lottery models produced distinct results, while the results of the neutral and heteromyopia xPODs were very similar.

Fig. 1. Realization of a spatial multispecies birth–death process, showing the statistics calculated on the multispecies spatial pattern at the end. (a) Total number of individuals with time; (b) total number of species with time; (c) species abundance distribution; (d) the spatial pattern of two species extracted from the multispecies pattern on a contour plot of the underlying environment; (e) logarithm of the pair correlation function of the two species in (d); \( A_{ij} \), the sum of the shaded areas; (f) cross pair overlap distribution, the histogram of \( A_{ij} \).
models were the most similar. Variability in lottery model xPODs occurred where the number of abundant species was low and the distribution correspondingly sparse.

Discussion

Attempts to infer underlying process from pattern in ecology have tended to rely on first-order community characteristics.
such as the SAD, which is the chief empirical test of many theories of species coexistence and the focus of a great deal of theoretical study in its own right (Fisher, Corbet & Williams 1943; Hubbell 2001; Volkov et al. 2003; McGill et al. 2007). Relatively little work has investigated the potential of second-order spatial information for similar inference, and we are aware of none that has systematically assessed the first- and second-order signals of modelled ecological processes.

The SADs produced by the models considered here show, on visual inspection, a great deal of overlap and considerable variation within those produced by any one model. There may be certain regions of the distributions where the models are distinct from one another, particularly in terms of community evenness. The lottery model produced the least even communities and the Janzen–Connell model the most even and diverse ones. These differences are small, however, and do not prove reliably distinguishable by the statistical tests used here. Differences in overall species diversity are attributable to model design, especially in the niche model where fecundity is depressed outside species’ optimum environment.

The Kolmogorov–Smirnov test indicates that random variations in species abundances outweigh similarities engendered by ecological process, as the test statistic takes very similar values for single-model pairs and cross-model pairs. On the basis of the Kolmogorov–Smirnov test statistic comparing pairs of species abundance distributions within and between ecological models (producing 45 and 100 values respectively). Results shown are from comparisons: (a) within neutral; (b) neutral–niche; (c) neutral–lottery; (d) neutral–Janzen–Connell; (e) neutral–heteromyopia. Whiskers extend to the largest and smallest values within 1.5 times the interquartile range and outliers are not shown. Low values of the test statistic indicate similarity of distributions.

Table 2. Mean evenness and diversity of each model. Indices used are Pielou’s evenness index and Shannon’s diversity index

<table>
<thead>
<tr>
<th>Model</th>
<th>Lottery</th>
<th>Niche</th>
<th>Heteromyopia</th>
<th>Neutral</th>
<th>Janzen–Connell</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pielou</td>
<td>0.43</td>
<td>0.72</td>
<td>0.71</td>
<td>0.71</td>
<td>0.81</td>
</tr>
<tr>
<td>Shannon</td>
<td>2.59</td>
<td>3.84</td>
<td>4.41</td>
<td>4.44</td>
<td>5.27</td>
</tr>
</tbody>
</table>

Fig. 3. Boxplots of Kolmogorov–Smirnov test statistics obtained from comparing pairs of species abundance distributions within and between ecological models (producing 45 and 100 values respectively). Results shown are from comparisons: (a) within neutral; (b) neutral–niche; (c) neutral–lottery; (d) neutral–Janzen–Connell; (e) neutral–heteromyopia. Whiskers extend to the largest and smallest values within 1.5 times the interquartile range and outliers are not shown. Low values of the test statistic indicate similarity of distributions.

Table 3. Mean and 90% limit values (in brackets) of the Kolmogorov–Smirnov test statistic comparing pairs of species abundance distributions (SADs) within and between contrasting models of ecological interactions, using species of all abundances. Mean values based on 45 pairs of SADs within ecological models and 100 pairs between models. Within-model means are in bold. J–C denotes Janzen–Connell

<table>
<thead>
<tr>
<th>Model</th>
<th>Neutral</th>
<th>Niche</th>
<th>Lottery</th>
<th>J–C</th>
<th>Heteromyopia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neutral</td>
<td>0.194</td>
<td>(0.222)</td>
<td>(0.156)</td>
<td>0.212</td>
<td>(0.348)</td>
</tr>
<tr>
<td>Niche</td>
<td>0.271</td>
<td>(0.309)</td>
<td>(0.255)</td>
<td>0.259</td>
<td>(0.384)</td>
</tr>
<tr>
<td>Lottery</td>
<td>0.264</td>
<td>(0.348)</td>
<td>(0.200)</td>
<td>0.317</td>
<td>(0.408)</td>
</tr>
<tr>
<td>J–C</td>
<td>0.135</td>
<td>(0.153)</td>
<td>(0.123)</td>
<td>0.230</td>
<td>(0.294)</td>
</tr>
<tr>
<td>Heteromyopia</td>
<td></td>
<td></td>
<td></td>
<td>0.207</td>
<td>(0.175)</td>
</tr>
</tbody>
</table>

Table 4. Mean and 90% limit values (in brackets) of the Kolmogorov–Smirnov test statistic comparing pairs of species abundance distributions within and between contrasting models of ecological interactions, using an abundance threshold of 500 individuals. Within-model means are in bold. J–C denotes Janzen–Connell

<table>
<thead>
<tr>
<th>Model</th>
<th>Neutral</th>
<th>Niche</th>
<th>Lottery</th>
<th>J–C</th>
<th>Heteromyopia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neutral</td>
<td>0.175</td>
<td>(0.228)</td>
<td>(0.106)</td>
<td>0.405</td>
<td>(0.696)</td>
</tr>
<tr>
<td>Niche</td>
<td>0.198</td>
<td>(0.309)</td>
<td>(0.127)</td>
<td>0.348</td>
<td>(0.179)</td>
</tr>
<tr>
<td>Lottery</td>
<td>0.348</td>
<td>(0.137)</td>
<td>(0.15)</td>
<td>0.542</td>
<td>(0.515)</td>
</tr>
<tr>
<td>J–C</td>
<td></td>
<td></td>
<td></td>
<td>0.266</td>
<td>(0.151)</td>
</tr>
<tr>
<td>Heteromyopia</td>
<td></td>
<td></td>
<td></td>
<td>0.152</td>
<td>(0.111)</td>
</tr>
</tbody>
</table>
of the values found here, SADs produced by the same model are, at best, only slightly more similar to one another than they are to those produced by different – or even contradictory – processes. The variations we find in SADs cannot, therefore, be classified as random or systematic by magnitude alone.

Truncating the SADs at an abundance of 500 slightly depresses within-model variation relative to between-model variation. Further analysis indicates that a threshold of 100 is better yet; this suggests that it may be possible to find an optimum section of the distribution in which random variation is

Fig. 4. Cross-pair overlap distributions (xPODs) obtained from different ecological models showing the shape of an example xPOD from a single realization (dashed lines) and 95% prediction intervals for xPODs based on 10 realizations (solid lines). Frequencies of areas are given relative to the total number of areas in each distribution.
els of species interactions in so far as they generate different spatial patterns. There are large differences between results from each model, with the exception of the neutral and heteromyopia models, and prediction intervals demonstrate that random within-model variation does not erode the distinction between models. This holds true even for the lottery model, in which random variation is large. At these parameter settings, the ecological processes modelled here are eminently distinguishable by their spatial signals.

The neutral and heteromyopia models produce xPODs with slightly negative means and small variances, the lottery and niche models give more negative means and skew distributions with greater variances, and the Janzen–Connell model a smaller negative mean and smaller variance. These signals broadly conform to expectations. The negative mean of the xPOD under neutrality is because of separation of species by density-dependent mortality, which is balanced within species by local dispersal. In the niche model, a greater spread of (mainly negative) values is directly attributable to the introduction of spatial niches which separate species according to their response to the underlying environment. While the spatial extent of each niche is approximately equal, the strength of niche separation itself is weak; niches are randomly assigned and so may overlap or entirely coincide. Stronger niche separation, corresponding to competition between species, would be expected to increase the spread of the xPOD further (and has been found to do so in trial results).

The lottery model lacks a mechanism for such strong separation of species, but the temporal variations in species fecundity allow dominant species to mingle with others and force them apart, so producing the observed spread of values. This effect is likely to be dependent upon the frequency of variation in fecundity (O’Malley et al. 2010). The Janzen–Connell effect counteracts the tendency of species to separate slightly under neutrality by suppressing conspecific clumping. There remains no mechanism to encourage the mingling of species, however, so the distribution produced has a smaller spread, as expected. Although the Janzen–Connell effect modelled here is not explicitly overcompensating, it does effectively prevent the survival of individuals in the immediate vicinity of their parent at equilibrium density, as required by the original theory and observed in some tropical tree species (Janzen 1970; Freckleton & Lewis 2006; Bagchi et al. 2010).

The heteromyopia model produces distributions indistinguishable from those of the neutral model, with identical means and variances. This may be because density-dependent mortality within species is spread over a larger area, so allowing increased clumping at small scales which counterbalances

### Table 5. Mean, standard deviations and skewness of cross pair overlap distributions for contrasting models of ecological interactions. J–C denotes Janzen–Connell.

<table>
<thead>
<tr>
<th>Model pair</th>
<th>Niche</th>
<th>Lottery</th>
<th>Heteromyopia</th>
<th>Neutral</th>
<th>J–C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>−0.267</td>
<td>−0.107</td>
<td>−0.055</td>
<td>−0.053</td>
<td>−0.050</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>0.319</td>
<td>0.132</td>
<td>0.043</td>
<td>0.040</td>
<td>0.026</td>
</tr>
<tr>
<td>Skewness</td>
<td>−0.556</td>
<td>−0.968</td>
<td>−0.434</td>
<td>−0.329</td>
<td>−0.262</td>
</tr>
</tbody>
</table>

**Fig. 5.** Boxplots of Kolmogorov–Smirnov test statistics obtained from comparing pairs of cross-pair overlap distributions within and between ecological models: (a) within neutral; (b) neutral–niche; (c) neutral–lottery; (d) neutral–Janzen–Connell; (e) neutral–heteromyopia. Whiskers extend to the largest and smallest values within 1.5 times the interquartile range and outliers are not shown. Low values of the test statistic indicate similarity of distributions.

minimized. Nevertheless, random and systematic differences between distributions, as captured by the Kolmogorov–Smirnov test statistic, remain of the same order in the threshold values that we have investigated (1, 100 and 500).

Previous studies of the SAD have generally concentrated either on its mathematical description (e.g. Fisher, Corbet & Williams 1943; Preston 1948; Dewdney 2003; McGill 2003; Volkov et al. 2003) or on comparing the shapes of the distribution generated by theoretical and empirical data (e.g. Motoura 1932; Tokeshi 1990; Hubbell 2001). It has also been argued that potential exists for distinguishing disturbance factors (Mouillot & Lepretre 2000), habitat complexity (Hurlbert 2004) and community stability (Thibault, White & Ernest 2004) by comparing SADs. While the SAD may be of value in studies of processes related to these, we find little justification for the assumption that the coexistence mechanisms that we modelled can be distinguished in its shape (Hubbell 2001; Marquet, Keymer & Cofre 2003; Etienne & Olf 2005).

The xPODs of each model are more reliably different from one another than the SADs. This would be expected because an xPOD is essentially an accessible summary of the spatial relationships between species and should reflect different modes of species interactions in so far as they generate different spatial patterns. There are large differences between results from each model, with the exception of the neutral and heteromyopia models, and prediction intervals demonstrate that random within-model variation does not erode the distinction between models. This holds true even for the lottery model, in which random variation is large. At these parameter settings, the ecological processes modelled here are eminently distinguishable by their spatial signals.

The neutral and heteromyopia models produce xPODs with slightly negative means and small variances, the lottery and niche models give more negative means and skew distributions with greater variances, and the Janzen–Connell model a smaller negative mean and smaller variance. These signals broadly conform to expectations. The negative mean of the xPOD under neutrality is because of separation of species by density-dependent mortality, which is balanced within species by local dispersal. In the niche model, a greater spread of (mainly negative) values is directly attributable to the introduction of spatial niches which separate species according to their response to the underlying environment. While the spatial extent of each niche is approximately equal, the strength of niche separation itself is weak; niches are randomly assigned and so may overlap or entirely coincide. Stronger niche separation, corresponding to competition between species, would be expected to increase the spread of the xPOD further (and has been found to do so in trial results).

The lottery model lacks a mechanism for such strong separation of species, but the temporal variations in species fecundity allow dominant species to mingle with others and force them apart, so producing the observed spread of values. This effect is likely to be dependent upon the frequency of variation in fecundity (O’Malley et al. 2010). The Janzen–Connell effect counteracts the tendency of species to separate slightly under neutrality by suppressing conspecific clumping. There remains no mechanism to encourage the mingling of species, however, so the distribution produced has a smaller spread, as expected. Although the Janzen–Connell effect modelled here is not explicitly overcompensating, it does effectively prevent the survival of individuals in the immediate vicinity of their parent at equilibrium density, as required by the original theory and observed in some tropical tree species (Janzen 1970; Freckleton & Lewis 2006; Bagchi et al. 2010).

The heteromyopia model produces distributions indistinguishable from those of the neutral model, with identical means and variances. This may be because density-dependent mortality within species is spread over a larger area, so allowing increased clumping at small scales which counterbalances

any repression of clumping at medium scales. This conformity of spatial patterns between models illustrates the absence of unique links between process and pattern.

Nevertheless, goodness-of-fit tests show that xPODs have far more success in distinguishing models than SADs. xPODs produced by different models are highly dissimilar, except those produced by the neutral and heteromyopia models, while distributions drawn from the same model are far more alike, producing smaller and more restricted values of the test statistic (Fig. 5, Table 6). The sole exception is the lottery model, in which random fluctuations in fecundities give rise to substantial variation in spatial pattern. Despite this, the difference between these and the comparable results for SADs (Fig. 3) is striking; the fingerprint of each model emerges from the background noise clearly and consistently in the case of the xPOD.

Our findings are subject to several caveats, the most important of which relates to the direct deduction of process from pattern. While we find that second-order patterns are significantly more informative than their first-order counterparts, it remains unreasonable to assume an inviolable causal link from underlying mechanisms (e.g. Baddeley & Silverman 1984; Lepš 1990). It must also be stressed that our results come from computer-based simulations, and that behaviour which we hold constant in the interests of clarity does not obey this requirement in real-world plant communities, potentially producing quite different outcomes (as similarly noted by e.g. Chave & Leigh 2002; Chave, Muller-Landau & Levin 2002; Levin et al. 2003). In particular, dispersal distances vary widely between species and have a dramatic impact on spatial structure (Levin et al. 2003). Changes in the relative scales of dispersal and environmental variation may have a similarly confounding effect (e.g. Lande, Engen & Saether 1999; Wiegand et al. 2007). Complex species-specific parameterization of dispersal kernels would be difficult to perform accurately and would undermine our attempts to identify the spatial signals of isolated processes, and so requires further research.

We also make assumptions about the environment and the expression of niche differentiation in our spatial and temporal niche models which may affect observed behaviour. While greater separation of niches is expected to strengthen the signal detected in our niche model, greater niche overlap would have the opposite effect because species remain effectively neutral within shared niches. We constrain our environment to a single circular distribution of values to ensure that differences in niche availability do not affect our results. However, the signal of niche differentiation in patchier environments would remain similar to those observed here, at similar relative scales, because of spatial separation between species with differing niche requirements. As in all models presented here, niche processes act through changes to the death term for the sake of consistency but are equivalent to the opposite adjustment to the birth term.

It is notable that the xPOD, as used here, still only makes use of a small fragment of the spatial information that is often available. For instance, marked point patterns from forests often contain information on spatial structure in relation to size, age and environment, and at different spatial scales. Moreover, there is spatial structure at third order and beyond. Nevertheless, second-order information evidently helps to distinguish underlying process, and is more informative than SADs, the well-established first-order measure. Together with the ready availability of spatially explicit data, this suggests that much of the power of spatial analysis in ecology remains to be exploited.

While bearing the above caveats in mind, the differences between the spatial signatures found here are large and intuitive, especially between the neutral model and the spatial niche model. We therefore predict, as a hypothesis for future empirical research, that plant communities in environments with more physical heterogeneity should have xPODs characterized by more negative means and larger variances than plant communities in environments with less physical heterogeneity. This hypothesis can be tested on tropical rainforest data currently available and provides a stronger test of the neutral theory than first-order measures. If the neutral model is correct, the hypothesis will not be supported.

Acknowledgements

This work required a great deal of computing time for simulations, the majority of which were run on the RNHI/BioSS Beowulf cluster at the University of Aberdeen Rowett Institute of Nutrition and Health (http://bioinformat ics.ri.sari.ac.uk). We thank Tony Travis, Kevin Hammond, Vladimir Janjic, Herbert Fruchtli, Greg Michaudson and Evan Brown for their assistance in running all simulations. We are also grateful for useful discussions with Angelika Stedney and Glenna Evans, and to two anonymous referees for their insightful and helpful comments. C.B. is funded by a Microsoft PhD Scholarship.

References


Received 12 November 2010; accepted 6 July 2011
Handling Editor: Kyle Harms

**Supporting Information**

Additional supporting information may be found in the online version of this article:

**Appendix S1.** Definition of the pair correlation function.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.