Methods in Ecology and Evolution

British Ecological Society

Methods in Ecology and Evolution 2015, **6,** 324–332 doi: 10.1111/2041-210X.12319

Towards a unified descriptive theory for spatial ecology: predicting biodiversity patterns across spatial scales

Sandro Azaele[1]*, Amos Maritan[2], Stephen J. Cornell[3], Samir Suweis[2], Jayanth R. Banavar[4], Doreen Gabriel[5,6] and William E. Kunin[6]

[1]*Department of Applied Mathematics, School of Mathematics, University of Leeds, Leeds LS2 9JT, UK;* [2]*Dipartimento di Fisica 'G. Galilei' & CNISM, INFN, Università di Padova, Via Marzolo 8, 35131 Padova, Italy;* [3]*Institute of Integrative Biology, University of Liverpool, Liverpool L69 7ZB, UK;* [4]*Department of Physics, University of Maryland, College Park, MD 20742, USA;* [5]*Julius Kühn-Institut - Federal Research Centre for Cultivated Plants, Bundesallee, 50, D-38116 Braunschweig, Germany; and* [6]*School of Biology, University of Leeds, Leeds LS2 9JT, UK*

Summary

1. A key challenge for both ecological researchers and biodiversity managers is the measurement and prediction of species richness across spatial scales. Typically, biodiversity is assessed at fine scales (e.g. in quadrats or transects) for practical reasons, but often we are interested in coarser-scale (field, regional, global) diversity issues. Moreover, the pressures affecting biodiversity patterns are often scale specific, making multiscale assessment a crucial methodological priority. As species richness is not additive, it is difficult to translate from the scale of measurement to the scale(s) of interest. A number of methods have been proposed to tackle this problem, but most are too model specific or too rigid to allow general application. Here, we present a general framework (and a specific implementation of it) that allows such scale translations to be performed.

2. Building on the intrinsic relationships among patterns of species richness, abundance and spatial turnover, we introduce a framework that links and predicts the profile of the species-area relationship and the species-abundance distributions across scales when a limited number of fine-scale scattered samples are available. Using the correlation in species' abundances between pairs of samples as a function of the distance between them, we are able to link the effects of aggregation, similarity decay, species richness and species abundances across scales.

3. Our approach allows one to draw inferences about biodiversity scaling under very general assumptions pertaining to the nature of interactions, the geographical distributions of individuals and ecological processes.

4. We demonstrate the accuracy of our predictions using data from two well-studied forest stands and also demonstrate the potential value of such methods by examining the effects of management on farmland insects across scales. The framework has important applications to biodiversity research and conservation practice.

Key-words: beta diversity, biodiversity patterns, pair correlation function, species-abundance distributions, species-area relationships, spatial ecology, upscaling biodiversity patterns

Introduction

Virtually everything important in ecology is deeply tied up in issues of scale. Physical conditions vary in complex, multi-scaled ways in space and time, and different organisms perceive, move and respond to conditions (behaviourally, physiologically and demographically) at different scales (Haskell, Ritchie & Olff 2002). These different populations may interact in scale-specific ways, further complicating biodiversity patterns across scales. Different drivers of ecological change (natural or anthropogenic) tend to act over different ranges of scales (Moorcroft, Hurtt & Pacala 2001), and as a result, biodiversity change may be different in strength or even in sign at different scales (Keil *et al.* 2011; Carvalheiro *et al.* 2013). Our concerns about natural systems are also scale specific; conservation goals may concern global, national or regional diversity, whereas ecosystem service arguments largely concern local or landscape-scale community properties (Hein *et al.* 2006). Finally, our efforts to monitor, understand and manage ecological communities are also intrinsically scale-bound.

We tend to manage and measure things at scales convenient to ourselves, but often we are concerned about the properties of a community at rather different scales. Thus for example, a particular agri-environmental scheme may be implemented at farm or field scale, but we will tend to monitor its effects on diversity at quadrat, trap or transect scale, and meanwhile our interests may concern its landscape or national scale impacts (Hein *et al.* 2006). Similar issues hold for many other topics of ecological research: alien species for instance may lower species richness at fine scales but increase it at regional to continental scales (Rosenzweig 2001; Powell, Chase & Knight 2013); whatever scale sample we use to monitor their effects can only reflect part of that complexity. We cannot afford to tile a whole landscape or continent in

*Corresponding author. E-mail: S.Azaele@leeds.ac.uk

quadrats, so we typically take a standardized sample of some sort, but that will tell us only of diversity change at the scale of the sample (Chase & Knight 2013), and we cannot simply sum up the species counts of our samples to get the coarser-scale estimate, because species richness is not additive. The increase at coarse scale depends in part on the turnover in species (β diversity at different scales) between different locations across a focal landscape or region. Thus, we need to incorporate such turnover to translate biodiversity information across scales.

To make progress in this respect requires bringing together multiple spatial ecological patterns into a single overarching framework. Spatial ecologists have amassed a wide array of classic biodiversity descriptors, including, inter alia, species-area relationships (SAR) (Arrhenius 1921; Rosenzweig 1995), species-abundance distributions (SAD) (McGill *et al.* 2007) and the distance decay of similarity (Tilman & Kareiva 1997; Morlon *et al.* 2008). A great deal of attention has been devoted to debating the precise functional forms of these curves; SAR functions, for example, have varied widely, including accelerating, decelerating and triphasic functions (Connor, McCoy & Cosby 1983; Tjørve 2003). There is a growing appreciation that the various spatial biodiversity descriptors are intrinsically inter-related, and substantial efforts have been devoted to understanding the links between them (Hanski & Gyllenberg 1997; Plotkin *et al.* 2000; Storch, Marquet & Brown 2007; McGill 2010). However, there is not yet any general method for linking them across scales, and one of the central challenges we face as a discipline is to bring together these various patterns into a general unified theoretical framework. At least two recent models have successfully predicted many or all of the above patterns. Spatially explicit forms of the neutral theory of biodiversity and biogeography (Hubbell 2001; Rosindell & Cornell 2007) can predict specific SAD, SAR and distance-decay functions from simple mechanistic community dynamics. Similarly, the maximum entropy theory of ecology (METE) (Harte 2011) predicts SAD and SAR patterns from simple biological constraints and statistical considerations and has tacit assumptions about distance-decay functions as well. These two approaches predict rather different patterns; for example, the spatial neutral model predicts a triphasic SAR (Rosindell & Cornell 2007), whereas the METE model predicts a function that decelerates towards an asymptote (Harte 2011). Both of these approaches have aspired to explore wide ranges of ecological patterns using only a few basic parameters, and each had substantial success (Volkov *et al.* 2003; Alonso, Etienne & McKane 2006; Azaele *et al.* 2006; Harte *et al.* 2008; Phillips & Dudik 2008; Azaele *et al.* 2010; Suweis *et al.* 2012). However, the simplicity of these approaches is also a weakness: each presents a fairly rigid and idiosyncratic structure based on modelling idealized natural communities at stationarity. This makes them poorly suited to assessing or monitoring natural or anthropogenic communities in flux, or for assessing biodiversity changes in response to management (Drakare, Lennon & Hillebrand 2006). A more robust and general approach to unifying spatial ecological pattern is needed (Fig. 1).

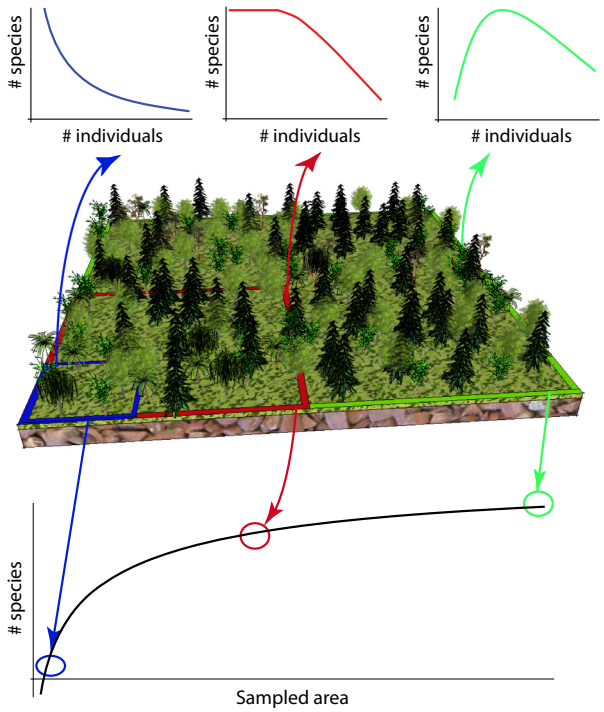

Fig. 1. Ecological patterns are interlinked and scale dependent. The profile of the species-abundance distribution (upper plots) changes from fine to coarse scales and the species-area relationship (lower plot) increases as a function of the sampled area with a scale-dependent slope. Both patterns are inextricably connected to spatial aggregation of conspecific individuals and similarity decay in species composition.

Such a unified spatial framework could have important practical value. Despite centuries of study, the biological diversity of our planet is not well described; in some regions and habitats (e.g. tropical forests, deep seas), only a small fraction of the species present are known to science (Mora & Sale 2011), while some taxonomic groups (e.g. nematodes, mites) remain poorly explored in any region (Groombridge 1992). Even in well-studied taxa and regions, the difficulties entailed in comprehensive surveys make broad-scale biodiversity change expensive or impractical to monitor. If we had a robust framework for translating biodiversity information across scales, it would allow fine-scale sampling to be upscaled to produce coarser-scale biodiversity estimates, and repeated rounds of such surveys would allow multiscale assessment of biotic change (Keil *et al.* 2011; Carvalheiro *et al.* 2013). Environmental change or management can result in biotic homogenization or altered spatial patterning, potentially shifting the shape of SARs or other spatial biodiversity patterns (Fraterrigo & Rusak 2008; Loreau 2010); such changes would be invisible to existing (idealized) unified model approaches (Drakare, Lennon & Hillebrand 2006), making such methods unsuitable for monitoring biotic change. What is required, therefore, is a general methodology for predicting and linking spatial biodiversity patterns across scales, that is sufficiently robust and flexible to allow its application to a range of natural or managed systems. Here, we propose such an approach.

At the centre of our approach is a method for describing the spatial structure of species' abundances. Most species' populations are spatially aggregated (Plotkin *et al.* 2000; Plotkin & Muller-Landau 2002; Storch, Marquet & Brown 2007; Harte, Smith & Storch 2009), and this results in a pattern of increasing species turnover with distance (Plotkin & Muller-Landau 2002; Storch, Marquet & Brown 2007; Morlon *et al.* 2008), the distance decay of similarity (Morlon *et al.* 2008). This pattern can be captured by the spatial pair correlation function (PCF), which describes the correlation in species' abundances between pairs of samples as a function of the distance between them (Chave & Leigh 2002; Zillio *et al.* 2005). If populations were randomly distributed in space, distinct communities would share the same fraction of species on average regardless of their spatial separation, and therefore, the PCF would not depend on distance; whereas in highly aggregated communities correlations in abundance would fall off steeply with increasing distance. The PCF measures not only the rate of turnover in species composition, it also reflects the variation of population clustering across scales, because the variance in species abundances at any particular scale can be calculated directly from the PCF. Therefore, the PCF is related to the spatial species-abundance distribution (sSAD), that is the number of species with a given number of individuals when we focus on a particular area. A region with a relatively high population variance leads naturally to a broad SAD with a wide spread in the abundances of species; conversely, a low population variance leads to a relatively narrow SAD. The PCF is thus able to link the effects of aggregation, similarity decay, species richness and species abundances across scales. However, the PCF does not uniquely determine the sSAD, because many different distributions of species abundances would have the same variance (i.e. a generic distribution is not specified completely by its mean and second moment). However, in the following we will show how our framework is able to predict the sSAD from the PCF with minimal additional assumptions.

Materials and methods

GENERAL FORMULATION OF THE METHODOLOGY

To link and predict the SAR and the sSAD patterns at different scales, we first compute the empirical PCF at a set of distances r and fit a function, $g(r)$, to interpolate between these distances. The specific form of the PCF is not important, so long as it gives a good fit to the empirical data; even a nonparametric form could be used. In this study, we used the modified Bessel function of the second kind (Lebedev & Silverman 1972), which provided a good description of the data and is also amenable to analytical calculations (see Implementation of the framework).

Secondly, we select a suitable family of SAD curves. A wide range of functional forms have been used in this context (McGill *et al.* 2007), and any of them can be adapted to our framework. We have opted here to use a Gamma distribution, as it is widely used, mathematically flexible and fits empirical SADs well (Dennis & Patil 1984; Engen & Lande 1996; Azaele *et al.* 2006).

Thirdly, we use the spatial information encapsulated in the PCF to define a suitable sSAD, which gives the number of species with n individuals within a circular area of radius R. Since we used the Gamma distribution for the SAD and that depends on two parameters, the predicted SAD at any particular scale is then unambiguously determined by two constraints: the mean density of individuals per species at some focal scale (which is readily obtained from the data) and the variance in abundance at that scale. By integrating twice $g(r)-1$ with respect to a given area in space, one obtains a quantity that is proportional to the variance in abundance over that area as shown in the Supporting Information. We can therefore link the PCF to the sSAD analytically. We also need to know the total number of species at some reference spatial scale, to fix the overall normalization of the SAD.

Assumptions of the method

This methodology does not depend on the particular analytical forms of the PCF or SAD, so long as the function is flexible enough to fit the data across scales (with scale-specific parameters); in principle, the method could accommodate any function for the PCF or SAD. The advantage of this approach is that one can explore features of species' spatial distributions without resorting to any underlying model. However, the approach relies on some important assumptions: (i) the functional form of the SAD is the same at different spatial scales and therefore, the effect of space is introduced only through the parameters, which are functions of the spatial scale (we have borrowed this hypothesis from the phenomenological renormalization group (Plischke & Bergersen 2006; section 6·5); (ii) the empirical SAD of the region on which one applies the method should be well approximated by a smooth curve at least at some focal scale and, at finer spatial scales, should not strongly depend on the position where it was calculated. Therefore, the region should have moderate levels of random environmental variability and contain a sufficiently large number of species; (iii) at the scale of the study region the inhomogeneities and anisotropies (e.g. strong gradients, important topological differences of the landscape (Muneepeerakul *et al.* 2011), large variations on habitat quality, etc.) should not be very strong. In the presence of major abrupt environmental shifts, in fact, the species turnover between two spatial points x and y can depend on the position x and y. Instead, the method requires that the PCF depends only on the distance between the two points; (iv) the fine-scale samples scattered at random should be sufficiently representative of the environmental conditions and biodiversity richness of the entire study region; (v) The spatial distribution of individuals should not be so patchy and irregular that the total number of individuals does not scale linearly with area. Although the framework could accommodate more general relations, the current formulation assumes a linear relation between the total number of individuals and the area where they live in.

IMPLEMENTATION OF THE FRAMEWORK

In this section, we provide details for implementing the proposed general framework with specific functions for the PCF and SAD, thus obtaining a link between the PCF, sSAD and SAR patterns.

The PCF is the fundamental spatial pattern linking all the other ones in our framework. For reasons that we clarify in the following, we have used the following definition

$$g(r) := \frac{\left\langle n_{\underline{x}} n_{\underline{y}} \right\rangle}{\langle n \rangle^2}, \quad \text{where} \quad |\underline{x} - \underline{y}| = r \qquad \text{eqn 1}$$

$n_{\underline{x}}$ is the density of individuals per species at $\underline{x}$ and $\langle n \rangle = N_0/(S_0 A_0)$ is the mean population density per species. N_0 is the total number of

individuals, S_0 the total number of species and A_0 the total area of the study region.

Let us assume that $N_\mu(\underline{x}, a)$ is the number of individuals of species μ present within the sample located at $\underline{x}$ with area a. To define empirically the numerator of eqn (1), we selected two non-overlapping regions of the same area at distance r, then multiplied the number of individuals belonging to the same species present in the corresponding regions and, finally, averaged across species. Therefore, we have calculated

$$\frac{1}{S_0}\sum_{\mu=1}^{S_0} N_\mu(\underline{x}, a)N_\mu(\underline{y}, a). \qquad \text{eqn 2}$$

This correlation measure does depend on the sizes (resolution), a, of the areas of the samples. However, we can disentangle the effects of the size of different sampled areas from those due to their spatial separation by dividing the expression in eqn (2) by

$$\left(\frac{1}{S_0}\sum_{\mu=1}^{S_0} N_\mu(\underline{x}, a)\right)\left(\frac{1}{S_0}\sum_{\mu=1}^{S_0} N_\mu(\underline{y}, a)\right). \qquad \text{eqn 3}$$

The resulting ratio of the expressions in (2) and (3) is roughly independent of a at least for $|\underline{x} - \underline{y}| = r \gg \sqrt{a}$, because both are approximately proportional to a^2. We found that these considerations are in very good agreement with the empirical data as shown in Fig. S1. Therefore, the empirical PCF has been defined as

$$g^{emp}(r) = \frac{1}{K_r}\sum_{\underline{x},\underline{y}} \frac{\frac{1}{S_0}\sum_{\mu=1}^{S_0} N_\mu(\underline{x},a)N_\mu(\underline{y},a)}{\left(\frac{1}{S_0}\sum_{\mu=1}^{S_0} N_\mu(\underline{x},a)\right)\left(\frac{1}{S_0}\sum_{\mu=1}^{S_0} N_\mu(\underline{y},a)\right)}\delta_{|\underline{x}-\underline{y}|,r}, \quad \text{eqn 4}$$

where the first sum is meant to be over the samples at locations $\underline{x}$ and $\underline{y}$, $\delta_{a,b}$ equals one if $a = b$ and is zero otherwise and $K_r = \sum_{\underline{x},\underline{y}} \delta_{|\underline{x}-\underline{y}|,r}$ gives the number of sample pairs at distance r. Assuming a homogeneous and isotropic distribution of individuals, this expression is equivalent to eqn (1) and approximately independent of a.

As a theoretical curve, we have used

$$g(r) = 1 + \frac{1}{2\pi}\left(\frac{\rho}{\lambda}\right)^2 K_0\left(\frac{r}{\lambda}\right), \qquad \text{eqn 5}$$

because it provides good fits to the empirical PCF and is amenable to analytical calculations. $K_0(x)$ is the modified Bessel function of the second kind (Lebedev & Silverman 1972). This function was already derived in the context of neutral models of biodiversity (Chave & Leigh 2002; Condit *et al.* 2002; Zillio *et al.* 2005). In our framework, ρ and λ have the dimensions of a length: when $r \gg \lambda$, the system becomes effectively uncorrelated.

The variance of the population of different species living in circular areas of radius R is $\sigma^2(R) = \langle N(R)^2\rangle - \langle N(R)\rangle^2$, being $\langle N(R)\rangle = \langle n\rangle \pi R^2$. In the Supporting Information, we show how to calculate $\sigma^2(R)$ analytically from the PCF. Once we have a formula for the variance of the population at different scales, we can bring the spatial information encapsulated within the PCF into the sSAD.

As noted above, we have chosen here to use a Gamma distribution (van Kampen 1992), i.e.

$$q(n|\alpha, \beta) = \frac{1}{\beta}\frac{(n/\beta)^{\alpha-1}}{\Gamma(\alpha)}e^{-n/\beta}, \qquad \text{eqn 6}$$

as the parametric function of the SAD, although other functions, such as a log-normal, with sufficiently flexible shapes could be used in its place. This distribution approximates the population with a continuous variable, which is useful in the calculations. However, one could have also used its discrete counterpart, that is the negative binomial (van Kampen 1992). The Gamma distribution at a specific scale depends on the two free parameters α and β; we can make the SAD spatially dependent by substituting α and β with two appropriate functions, $\alpha(R)$ and $\beta(R)$, that depend on the radius, R, of the circular area we focus on, so that $q(n|\alpha(R),\beta(R))$ is our sSAD.

The first two moments of the sSAD have clear interpretations: the first is simply the mean density (number per unit area) of individuals per species, the second is related to the spatial variance that depends on the PCF as we have shown before. These two conditions provide two constraints for the functions $\alpha(R)$ and $\beta(R)$ which can be used to achieve explicit expressions. It turns out that the analytic forms of $\alpha(R)$ and $\beta(R)$ are (see Supporting Information):

$$\alpha(R) \equiv \left(\frac{\langle N(R)\rangle}{\sigma(R)}\right)^2 = \frac{\pi R^2}{\rho^2}\left(1 - \frac{2\lambda}{R}\frac{K_1\left(\frac{R}{\lambda}\right)I_1\left(\frac{R}{\lambda}\right)}{I_0\left(\frac{R}{\lambda}\right)K_1\left(\frac{R}{\lambda}\right) + I_1\left(\frac{R}{\lambda}\right)K_0\left(\frac{R}{\lambda}\right)}\right)^{-1}, \qquad \text{eqn 7}$$

$$\beta(R) \equiv \frac{\sigma^2(R)}{\langle N(R)\rangle} = \langle n\rangle\rho^2\left(1 - \frac{2\lambda}{R}\frac{K_1\left(\frac{R}{\lambda}\right)I_1\left(\frac{R}{\lambda}\right)}{I_0\left(\frac{R}{\lambda}\right)K_1\left(\frac{R}{\lambda}\right) + I_1\left(\frac{R}{\lambda}\right)K_0\left(\frac{R}{\lambda}\right)}\right), \qquad \text{eqn 8}$$

where I_i and K_i are modified Bessel functions of order i of the first and second kind, respectively (Lebedev & Silverman 1972). Therefore, the final expression for the sSAD as a function of the radius $R < R_0$ is as follows:

$$sSAD(n|R, R_0) = S_{emp}(R_0)\frac{q(n|\alpha(R), \beta(R))}{\int_1^\infty q(m|\alpha(R_0), \beta(R_0))\mathrm{d}m}, \qquad \text{eqn 9}$$

where

$$q(n|\alpha(R), \beta(R)) = \frac{1}{\beta(R)}\frac{(n/\beta(R))^{\alpha(R)-1}}{\Gamma(\alpha(R))}e^{-n/\beta(R)}, \qquad \text{eqn 10}$$

$\Gamma(x)$ is the gamma function and $sSAD(n|R,R_0)$ depends on the specific scale of the whole study region which is assumed to be circular with

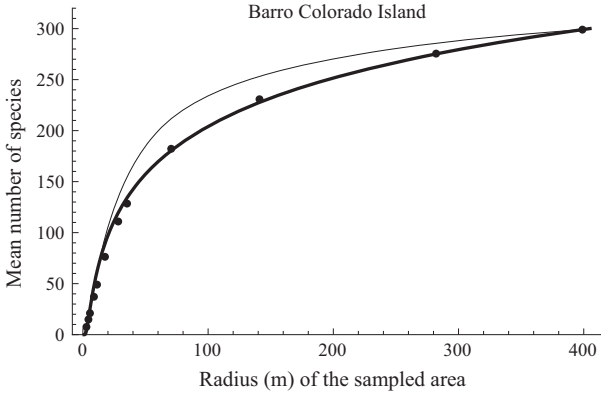

Fig. 2. The thick solid curve represents the predicted species-area relationship (SAR) as given by the pair correlation function (PCF) and the mean population density per species when we use all the available information in the study region. Black dots represent empirical data. We highlight that this curve was not best fit to the data, but rather is the SAR as predicted by the PCF and the mean population density, as outlined in the text. The thin solid line represents the SAR that would be predicted if individuals were randomly placed, that is spatial aggregation is negligible. See the Supporting Information for the fit of the PCF, the predicted species-abundance distribution and the corresponding curves for the Pasoh forest.

radius R_0. The empirical number of species found within such a study region is $S_{emp}(R_0)$.

From eqn (9), one also obtains the species-area relationship (SAR), that is the mean number of species that live in area A, which in our case is circular with radius R:

$$S(R) = S_{emp}(R_0)\frac{\int_1^\infty q(n|\alpha(R),\beta(R))\mathrm{d}n}{\int_1^\infty q(m|\alpha(R_0),\beta(R_0))\mathrm{d}m}. \qquad \text{eqn 11}$$

This formula provides us with the whole profile of the SAR and depends on the parameters λ and ρ (from the PCF) and $\langle n\rangle$ only [see Fig. S9 for the different SAR shapes obtained when using eqn (10) in eqn (11)].

EXTRAPOLATION OF THE SAR AND SAD

The methodology that we have described so far can also be used to extrapolate the profile of the SAR and SAD when a limited number of spatially scattered samples are available. Eqns (9) and (11) depend upon the total number of species in the whole region with radius R_0, which is not known explicitly from a limited set of samples but can be estimated by extrapolation. Our full procedure is described in the Supporting Information (section S3) and uses eqn (11) to upscale from our set of disconnected samples, which cover a total area $A_{samp} = \pi R_{samp}^2$ and contain a total number of species $S_{emp}(R_{samp})$. Here, we give an overview in a few steps.

Firstly, from the samples, we calculate the empirical PCF and the mean population density per species for the combined samples. Through a least-squares best fit, we then parameterize the theoretical curve defined in eqn (5) and therefore obtain values for λ and ρ. We compute α and β using Eqs. (7) and (8) with the fitted parameters λ, ρ and the empirical mean population density per species $\langle n\rangle$ as obtained from the aggregated samples. With all this information, we can use eqn (11) to extrapolate (upscale) from the scale R_{samp} to the scale of the entire region with radius R_0, that is $S_{up}(R_0)$, as explained in section S3.

Secondly, we calculate the parameters at the scale R_0: λ, ρ are already known from the samples and we assume that they do not depend on the spatial scale in a first approximation. However, the mean population density per species does depend on the scale, but we can readily calculate it as $\langle n_{up}(R_0)\rangle = S_{emp}(R_{samp})\langle n_{emp}(R_{samp})\rangle/S_{up}(R_0)$, where we have assumed that the summed density of individuals of all species combined is scale independent and $\langle n_{emp}(R_{samp})\rangle$ is the empirical mean population density per species for the aggregated samples.

To calculate the SAR at finer scales (downscaling), we substitute the parameters λ, ρ, $\langle n_{up}(R_0)\rangle$ and $S_{up}(R_0)$ into eqn (11) and calculate the SAR at radii $R < R_0$. In addition, with the same parameters and eqn

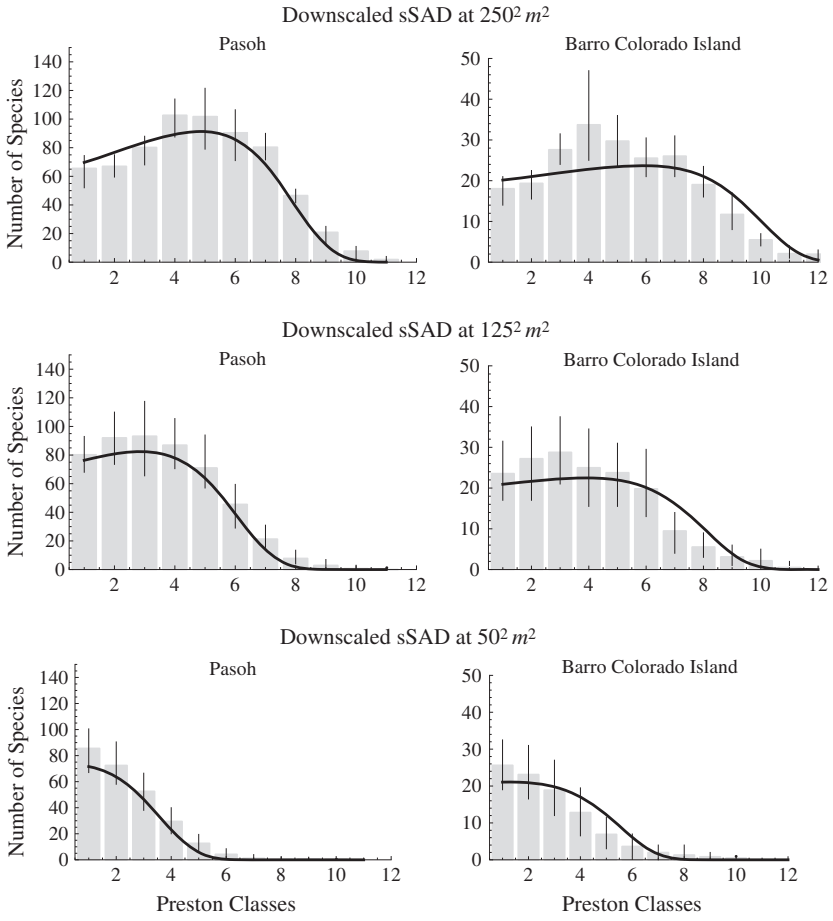

Fig. 3. Species-abundance distributions predicted by the theory at finer spatial scales. Solid curves are predictions of the spatial species-abundance distribution (sSAD) within subareas of the study region. The parameters are those used in Fig. 2 for BCI and Fig. S6 for Pasoh. Grey histograms represent empirical data. The area for which we downscaled the sSAD is indicated in the plot label. We divided the whole study region into equal non-overlapping square subareas, calculated the Preston classes of abundances as in Volkov *et al.* (2003) within each subarea and finally averaged across subareas for each corresponding bin. Errors bars in each bin represent 95% confidence intervals as obtained from the empirical data.

(9), we can also extrapolate the profile of the sSAD, thus estimating the abundances of species in regions that have not yet been sampled (see also Supporting Information for further details).

Results

To test the method, we applied it to two focal regions belonging to well-studied forest stands: Barro Colorado Island (BCI) in Panama and Pasoh Forest Reserve in Malaysia. These two plots allow us to test model predictions against known patterns, and both plots display high but greatly differing species richness. The comparisons between the predicted SAR and SAD at the level of the whole study region and the corresponding empirical data are shown in Fig. 2 and Figs S5, S6. In particular, Fig. 2 shows a comparison between the SAR predicted by a model which places individuals at random in space and our framework. The results demonstrate that spatial aggregation of individuals is important for predicting the correct shape of the SAR and is well captured by the PCF that we have introduced.

Figure 3 shows that the method correctly predicts the change of the empirical sSAD curves even at finer spatial scales. We emphasize that these profiles have not been fitted to the empirical curves, but are predictions based solely on the mean density of individuals per species as obtained from the empirical data and the rate of spatial decay in similarity as encapsulated by the PCF (see also Fig. S2). Figure 3 is also important for two additional reasons: (a) SADs at finer scales reveal that the two study regions are not homogeneous (see error bars in Fig. 3); however, the average (position-independent) SAD captures the overall behaviour of the SAD at different spatial locations with sufficient fidelity (see assumption iii) of the methodology); (b) empirical data are consistent with assumption i) of the method, and therefore, the empirical SAD at different scales can be well captured by a theoretical SAD whose functional form is the same across scales. The framework can also predict a wide range of different curves, including SARs that are accelerating, decelerating and S-shaped or 'triphasic' (see Fig. S9) as recently observed in continental or global SARs (Storch, Keil & Jetz 2012).

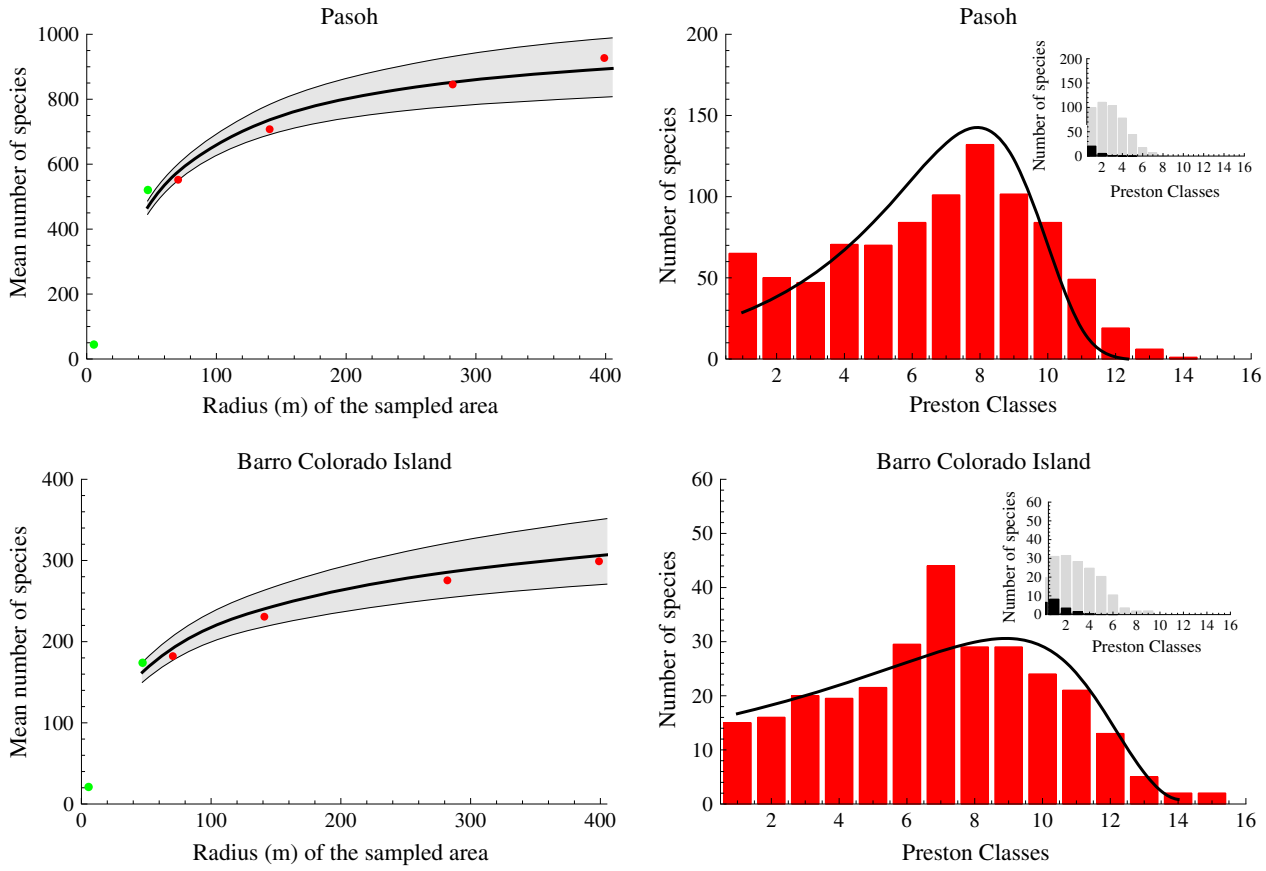

Fig. 4. Predicted species-area relationship (SAR) and species-abundance distribution (SAD) when sampling about 1% (60 samples) of each study plot (see also Supporting Information). The PCF and the density were calculated from these samples only. Left panels: for each study plot, the solid curve is the median at each scale of the predicted SAR starting from the scale of the sampled area; the grey strip encloses 95% of the predictions (out of 1,000 replicates); the red dots represent the empirical SAR, and the green dots the mean number of species at the scale of a single sample and all the selected samples, respectively. Note that the green dot at the larger scale does not belong to the empirical SAR, because samples were scattered across the entire study regions. Right panels: for each study plot, the solid curve is the prediction of the SAD for the whole study region (median within each bin). Preston classes were calculated as in Volkov *et al.* (2003). The red histograms are empirical data. The insets show the empirical SAD at the scale of one sample (black histograms) and all the selected samples (grey histograms), i.e. those which were used for the predictions.

One of the most important advances of our method is that one can exploit it to extrapolate the profile of the SAR and SAD when a limited number of spatially scattered samples are available. Indeed, we can reconstruct the SAR and SAD from fine-scale samples over a small fraction of the region. Conventional methods would require a comprehensive survey to estimate these curves, which would be prohibitively expensive and impractical for many taxa.

The forests in BCI and Pasoh are comprehensively surveyed, but we can simulate an incomplete survey using only a subset of the data to calculate the mean density of individuals per species and fit our $g(r)$ function. We then use these to estimate the empirical SAR and SAD over the whole study plot by implementing the procedure outlined in the previous sections. We randomly sampled just over 1% (60 samples) of each study plot with $10m \times 10m$ non-overlapping samples and repeated the sampling scheme 1000 times to assess the robustness of the results: Fig. 4 shows that the protocol was able to predict the correct profile of the SAR and SAD within the 95% confidence intervals. Thus, our methodology makes it possible to predict the species richness of a region from a small subsample and to infer the likely abundances of the unobserved species.

If we can upscale species richness from scattered samples, it allows us for the first time to make valid inferences about patterns at coarse scales using incomplete fine-scale data. This in turn allows the cross-scale influence of management or other interventions to be estimated. To demonstrate this, we analysed hoverfly data collected in cereal fields on paired conventional and organic farms across England [for details, see Gabriel *et al.* (2010)]. Unlike most other sets of organisms surveyed, hoverflies had been found to be commoner and more species-rich in conventional fields than in organic fields at the α-scale of an individual trap. However, despite the large fluctuations of the spatial correlation (as reflected in the standard deviation of the PCF in Fig. 5), the PCF of hoverflies was lower on average in organic farms than in conventional ones (upper panels, Fig. 5), due in large part to the role of widespread aphidophagous species (e.g. *Episyrphus balteatus*) that dominate conventional farm samples. This implies greater turnover in species across space in organic farms than in their conventional neighbours, resulting in more rapid species accumulation across scales. Consequently, the predicted SAR for organic farms crosses that of conventional ones, predicting higher diversity for hoverflies on organic farms at a regional scale, despite the lower diversity seen at fine scales (lower panel, Fig. 5).

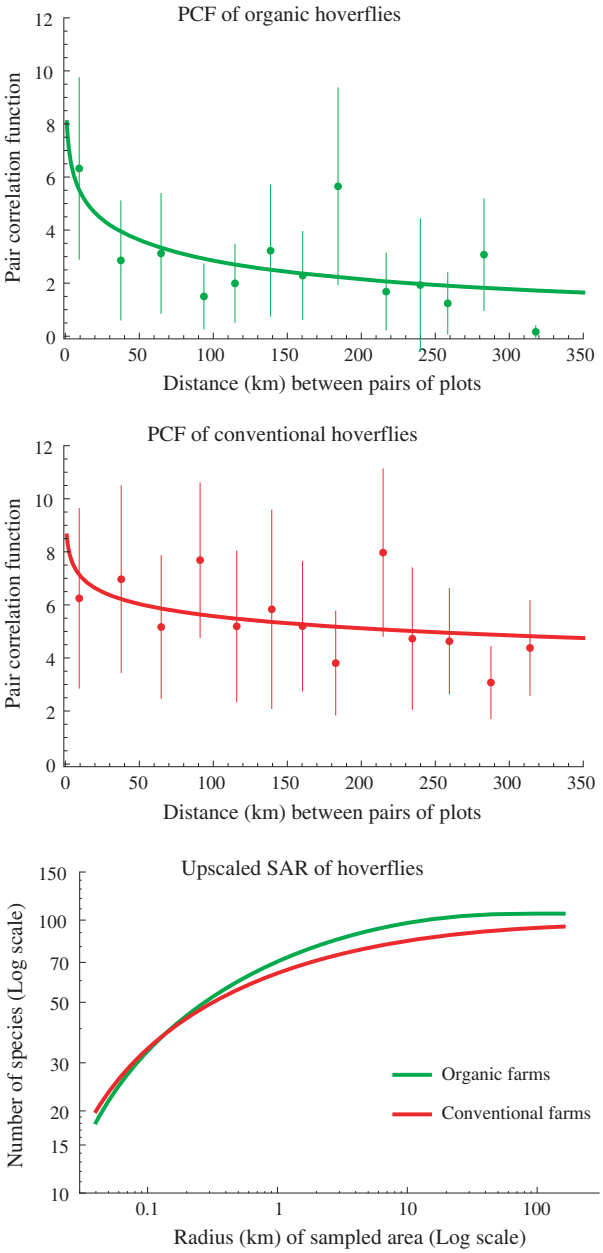

Fig. 5. Pair correlation functions (PCF, upper panels) and upscaled species-area relationship (SAR, lower panel) of hoverflies in 2008 [for further details, see Gabriel *et al.* (2010)]. Green (red) dots represent the empirical PCF found for hoverflies within organic (conventional) farmlands, and solid curves are the corresponding best-fitted curves ($R^2_{adj,green} = 0{\cdot}81$, $R^2_{adj,red} = 0{\cdot}94$) with eqn 5. Each dot shows the average PCF calculated within a spatial interval of 25 km. Every error bar is one standard deviations wide and represents the variability within each bin. Solid curves in the lower panel show the upscaled SAR for hoverflies found within organic (green) and conventional (red) farmlands.

Discussion

In the sections above, we have outlined and tested a new analytical approach to dealing with issues of scale in ecological data. It not only provides a method for translating information across scales (e.g. for downscaling or upscaling species richness between scales of measurement and scales of interest), it also links multiple ecological patterns together, without the constraints of a particular mechanistic community model.

Other approaches have tended to be insufficiently scale-specific to allow the full range of scales to be inferred (Shen & He 2008) or else to fit a specific clustering process, which requires more data-hungry methods such as Ripley's K-statistic (Morlon *et al.* 2008; McGill 2011). Our approach, in contrast, is designed to be flexible and suitable for sparsely sampled data. It allows a wide range of different SAR shapes, and

further flexibility could be incorporated by the use of different functions within the framework. As demonstrated above, this property should make it possible to conduct biodiversity assessment and monitoring over substantial areas with only a fraction of the effort that would be required for comprehensive surveying. It also allows us to explore some of the broad-scale issues we care about using the sort of fine-scale data we are capable of collecting. If further tests confirm the robustness and predictive power of the method, it should provide a powerful tool for biodiversity research and management.

Our method can be extended to data sets whose similarity-distance decay cannot be captured by simple analytical functions or require a more flexible family of SADs. In fact, provided higher spatial moments can be estimated from the data, additional information can be introduced into an appropriate sSAD and finer details captured. For example, the third moment of the sSAD is related to the triplet correlation function, which can be obtained from point samples in the same way as the PCF. Thus, the specific method presented above is only one example of a broader family of methods using different functional forms or potentially nonparametric approaches. The degree to which we can further improve the predictive abilities of these methods through such refinements remains an open question for research. We also expect that it will be possible to extend the approach to calculate other patterns such as the endemic area relationship, to downscale biodiversity and abundance patterns (Azaele, Cornell & Kunin 2012; Barwell *et al*. 2014) to finer scales, and to allow the use of presence/absence or proportional cover information, in addition to the abundance count data used here. Such extensions would widen the set of potential applications of our analytical approach.

In conclusion, this general mathematical framework provides a common language to link different spatial patterns, thus providing a promising route for important applications to biodiversity research and conservation practice.

Acknowledgements

We thank two anonymous referees and the associate editor of the journal for their helpful comments. We are also grateful to the FRIM Pasoh Research Committee (M.N.M. Yusoff, R. Kassim) and the Center for Tropical Research Science (R. Condit, S. Hubbell, R. Foster) for providing the empirical data of the Pasoh and BCI forests, respectively. This work was supported by the EU FP7 SCALES project ('Securing the Conservation of biodiversity across Administrative Levels and spatial, temporal and Ecological Scales'; project No. 26852). A.M. and S.S. acknowledge Cariparo foundation for financial support. Original data are available at http://www.ctfs.si.edu.

Data accessibility

MATHEMATICA code: uploaded as online supporting information.

References

Alonso, D., Etienne, R. & McKane, A. (2006) The merits of neutral theory. *Trends in Ecology & Evolution*, **21**, 451–457.

Arrhenius, O. (1921) Species and area. *African Journal of Ecology*, **9**, 95–99.

Azaele, S., Pigolotti, S., Banavar, J.R. & Maritan, A. (2006) Dynamical evolution of ecosystems. *Nature*, **444**, 926–928.

Azaele, S., Muneepeerakul, R., Rinaldo, A. & Rodriguez-Iturbe, I. (2010) Inferring plant ecosystem organization from species occurrences. *Journal of Theoretical Biology*, **262**, 323–329.

Azaele, S., Cornell, S.J. & Kunin, W.E. (2012) Downscaling species occupancy from coarse spatial scales. *Ecological Applications*, **22**, 1004–1014.

Barwell, L.J., Azaele, S., Kunin, W.E. & Isaac, N.J. (2014) Can coarse-grain patterns in insect atlas data predict local occupancy? *Diversity and Distributions*, **20**, 895-907.

Carvalheiro, L.G., Kunin, W.E., Keil, P., Aguirre-Gutiérrez, J., Ellis, W.N., Fox, R. *et al*. (2013) Species richness declines and biotic homogenisation have slowed down for nw-european pollinators and plants. *Ecology Letters*, **16**, 870–878.

Chase, J.M. & Knight, T.M. (2013) Scale-dependent effect sizes of ecological drivers on biodiversity: why standardised sampling is not enough. *Ecology Letters*, **16**, 17–26.

Chave, J. & Leigh, E. (2002) A spatially explicit neutral model of beta-diversity in tropical forests. *Theoretical Population Biology*, **62**, 153–168.

Condit, R., Pitman, N., Leigh, E., Chave, J., Terborgh, J., Foster, R. *et al*. (2002) Beta-diversity in tropical forest trees. *Science*, **295**, 666.

Connor, E.F., McCoy, E.D. & Cosby, B. (1983) Model discrimination and expected slope values in species-area studies. *American Naturalist*, **122**, 789–796.

Dennis, B. & Patil, G. (1984) The gamma distribution and weighted multimodal gamma distributions as models of population abundance. *Mathematical Biosciences*, **68**, 187–212.

Drakare, S., Lennon, J.J. & Hillebrand, H. (2006) The imprint of the geographical, evolutionary and ecological context on species-area relationships. *Ecology Letters*, **9**, 215–227.

Engen, S. & Lande, R. (1996) Population dynamic models generating the lognormal species abundance distribution. *Mathematical Biosciences*, **132**, 169–183.

Fraterrigo, J.M. & Rusak, J.A. (2008) Disturbance-driven changes in the variability of ecological patterns and processes. *Ecology Letters*, **11**, 756–770.

Gabriel, D., Sait, S.M., Hodgson, J.A., Schmutz, U., Kunin, W.E. & Benton, T.G. (2010) Scale matters: the impact of organic farming on biodiversity at different spatial scales. *Ecology Letters*, **13**, 858–869.

Groombridge, B. (ed) (1992) *Global Biodiversity: Status of the Earth's Living Resources*. Chapman & Hall, ISBN 0412472406.

Hanski, I. & Gyllenberg, M. (1997) Uniting two general patterns in the distribution of species. *Science*, **275**, 397–400.

Harte, J. (2011) *Maximum Entropy and Ecology: A Theory of Abundance, Distribution, and Energetics*. Oxford University Press, Oxford.

Harte, J., Zillio, T., Conlisk, E. & Smith, A. (2008) Maximum entropy and the state-variable approach to macroecology. *Ecology*, **89**, 2700–2711.

Harte, J., Smith, A.B. & Storch, D. (2009) Biodiversity scales from plots to biomes with a universal species-area curve. *Ecology Letters*, **12**, 789–797.

Haskell, J.P., Ritchie, M.E. & Olff, H. (2002) Fractal geometry predicts varying body size scaling relationships for mammal and bird home ranges. *Nature*, **418**, 527–530.

Hein, L., Van Koppen, K., De Groot, R.S. & Van Ierland, E.C. (2006) Spatial scales, stakeholders and the valuation of ecosystem services. *Ecological Economics*, **57**, 209–228.

Hubbell, S. (2001) *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton University Press, Princeton, New Jersey, USA; Oxford.

van Kampen, N.G. (1992) *Stochastic Processes in Physics and Chemistry*. Elsevier, North Holland, Amsterdam.

Keil, P., Biesmeijer, J.C., Barendregt, A., Reemer, M. & Kunin, W.E. (2011) Biodiversity change is scale-dependent: an example from Dutch and UK hoverflies (Diptera, Syrphidae). *Ecography*, **34**, 392–401.

Lebedev, N. & Silverman, R. (1972) *Special Functions and Their Applications*. Dover, New York, USA.

Loreau, M. (2010) *From Populations to Ecosystems: Theoretical Foundations For a New Ecological Synthesis*. Princeton University Press, Princeton, New Jersey, USA; Oxford.

McGill, B. (2010) Towards a unification of unified theories of biodiversity. *Ecology Letters*, **13**, 627–642.

McGill, B.J. (2011) Linking biodiversity patterns by autocorrelated random sampling. *American Journal of Botany*, **98**, 481–502.

McGill, B., Etienne, R., Gray, J., Alonso, D., Anderson, M., Benecha, H. *et al*. (2007) Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. *Ecology Letters*, **10**, 995–1015.

Moorcroft, P., Hurtt, G. & Pacala, S.W. (2001) A method for scaling vegetation dynamics: the ecosystem demography model (ED). *Ecological Monographs*, **71**, 557–586.

Mora, C. & Sale, P.F. (2011) Ongoing global biodiversity loss and the need to move beyond protected areas: a review of the technical and practical shortcom-

ings of protected areas on land and sea. *Marine Ecology Progress Series*, **434**, 251–266.
Morlon, H., Chuyong, G., Condit, R., Hubbell, S., Kenfack, D., Thomas, D., Valencia, R. & Green, J. (2008) A general framework for the distance decay of similarity in ecological communities. *Ecology Letters*, **11**, 904–917.
Muneepeerakul, R., Azaele, S., Levin, S.A., Rinaldo, A. & Rodriguez-Iturbe, I. (2011) Evolution of dispersal in explicitly spatial metacommunities. *Journal of Theoretical Biology*, **269**, 256–265.
Phillips, S.J. & Dudik, M. (2008) Modeling of species distributions with maxent: new extensions and a comprehensive evaluation. *Ecography*, **31**, 161–175.
Plischke, M. & Bergersen, M. (2006) *Equilibrium Statistical Physics*. World Scientific, Singapore.
Plotkin, J.B. & Muller-Landau, H.C. (2002) Sampling the species composition of a landscape. *Ecology*, **83**, 3344–3356.
Plotkin, J.B., Potts, M.D., Leslie, N., Manokaran, N., LaFrankie, J. & Ashton, P.S. (2000) Species-area curves, spatial aggregation, and habitat specialization in tropical forests. *Journal of Theoretical Biology*, **207**, 81–99.
Powell, K.I., Chase, J.M. & Knight, T.M. (2013) Invasive plants have scale-dependent effects on diversity by altering species-area relationships. *Science*, **339**, 316–318.
Rosenzweig, M.L. (1995) *Species Diversity in Space and Time*. Cambridge University Press, Cambridge.
Rosenzweig, M.L. (2001) The four questions: what does the introduction of exotic species do to diversity? *Evolutionary Ecology Research*, **3**, 361–367.
Rosindell, J. & Cornell, S. (2007) Species area relationships from a spatially explicit neutral model in an infinite landscape. *Ecology Letters*, **10**, 586–595.
Shen, T.J. & He, F. (2008) An incidence-based richness estimator for quadrats sampled without replacement. *Ecology*, **89**, 2052–2060.
Storch, D., Marquet, P.A. & Brown, J. (2007) *Scaling Biodiversity*. Cambridge University Press, Cambridge.
Storch, D., Keil, P. & Jetz, W. (2012) Universal species-area and endemics-area relationships at continental scales. *Nature*, **488**, 78–81.
Suweis, S., Bertuzzo, E., Mari, L., Rodriguez-Iturbe, I., Maritan, A. & Rinaldo, A. (2012) On species persistence-time distributions. *Journal of Theoretical Biology*, **303**, 15–24.
Tilman, D. & Kareiva, P. (1997) *Spatial Ecology: The Role of Space in Population Dynamics and Interspecific Interactions*. Princeton University Press, Princeton, New Jersey, USA; Oxford.
Tjørve, E. (2003) Shapes and functions of species-area curves: a review of possible models. *Journal of Biogeography*, **30**, 827–835.
Volkov, I., Banavar, J.R., Hubbell, S.P. & Maritan, A. (2003) Neutral theory and relative species abundance in ecology. *Nature*, **424**, 1035–1037.
Zillio, T., Volkov, I., Banavar, J.R., Hubbell, S.P. & Maritan, A. (2005) Spatial scaling in model plant communities. *Physical Review Letters*, **95**, 098101.

Received 10 June 2014; accepted 18 November 2014
Handling Editor: David Murrell

Supporting Information

Additional Supporting Information may be found in the online version of this article.

Fig. S1. Pair correlation function calculated in two different ways.

Fig. S2. Pair correlation functions for the BCI and Pasoh forests.

Fig. S3. Spatial variance, $\sigma^2(r)$, as a function of the radius of the sampled area for the BCI and Pasoh forests.

Fig. S4. Mean number of individuals per species (including all species) as a function of the radius of the sampled area.

Fig. S5. Comparison between the predicted and empirical SAR at the level of the whole study region for the BCI forest.

Fig. S6. Ecological patterns predicted by the theory at the scale of the study region for the Pasoh forest.

Fig. S7. This figure shows the SAR (black dots) as a function of the radius R of the sampled area A, where the area is meant to be connected

Fig. S8. Predicted total species richness as a function of the percentage of sampled area.

Fig. S9. Bilogarithmic plots of different SAR curves according to Eqs. 10 and 11 of the main text

Data S1. The spatial variance.

Data S2. The spatial species abundance distribution.

Data S3. Extrapolation of species richness, SAR and SAD: detailed description.