LETTER

Linking dispersal, immigration and scale in the neutral theory of biodiversity

Ryan A. Chisholm* and Jeremy W. Lichstein

Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08544, USA

*Correspondence: E-mail: chisholm@princeton.edu, jwl@princeton.edu

Abstract

In the classic spatially implicit formulation of Hubbell's neutral theory of biodiversity a local community receives immigrants from a metacommunity operating on a relatively slow timescale, and dispersal into the local community is governed by an immigration parameter m. A current problem with neutral theory is that m lacks a clear biological interpretation. Here, we derive analytical expressions that relate the immigration parameter m to the geometry of the plot defining the local community and the parameters of a dispersal kernel. Our results facilitate more rigorous and extensive tests of the neutral theory: we conduct a test of neutral theory by comparing estimates of m derived from fits to empirical species abundance distributions to those derived from dispersal kernels and find acceptable correspondence; and we generate a new prediction of neutral theory by investigating how the shapes of species abundance distributions change theoretically as the spatial scale of observation changes. We also discuss how our main analytical results can be used to assess the error in the mean-field approximations associated with spatially implicit formulations of neutral theory.

Keywords

Dispersal kernel, edge effects, Hubbell, immigration, scale, seed dispersal, spatially implicit and explicit neutral models, unified neutral theory of biodiversity and biogeography.

Ecology Letters (2009) **12**: 1385–1393

INTRODUCTION

Hubbell's neutral theory of biodiversity seeks to explain observed patterns of species abundance and diversity with a simple model that assumes that all species are equivalent on a per-capita basis (Hubbell 1979, 2001; Bell 2001). The neutral model is a first-order model that ignores many processes that are known to operate in real ecosystems, such as niche specialization and interspecific competition. Nevertheless, it provides remarkable fits with very few parameters to biodiversity data from ecosystems including rainforests (Volkov *et al.* 2003, 2005), coral reefs (Volkov *et al.* 2007), and freshwater fish communities (Muneepeerakul *et al.* 2008). This suggests that an understanding of the emergent macroscopic properties of ecosystems can be attained without understanding the fine-scaled processes that generate them. Since Hubbell's seminal work introducing the neutral theory (Hubbell 1979, 2001), subsequent research has focused on estimating the model's parameters (Munoz *et al.* 2007; Jabot & Chave 2009), fitting the neutral model to data (Latimer *et al.* 2005), comparing the theory's

predictions to those of non-neutral theories (Purves & Pacala 2005), simulating the theory's dynamics (Rosindell *et al.* 2008), and investigating non-equilibrium dynamics (Adler 2004; Chisholm & Burgman 2004; Vanpeteghem *et al.* 2008). Reviews of research into neutral theory can be found in Chave (2004) and Leigh (2007).

A current problem with neutral theory is the lack of a clear biological interpretation of the immigration parameter m—a key parameter in Hubbell's (2001) original spatially implicit formulation, in which a semi-isolated local community receives immigrants from a much larger metacommunity that operates on a slower timescale (henceforth we refer to Hubbell's local community/metacommunity model as the 'classic' spatially implicit neutral model). The immigration parameter m is the probability that a death in the local community is replaced by the offspring of an individual from outside the local community. Previous researchers have estimated m by fitting theoretical neutral models to observed species abundance distributions (SADs) from ecosystems including tropical forests, coral reefs and South African Fynbos (Latimer *et al.* 2005; Volkov *et al.* 2007). However, it

has been previously difficult to compare estimates of m from different ecosystems (and hence to conjecture about the ecological basis for observed differences in SADs between ecosystems) or from studies with different sampling designs (e.g., plot sizes) because general relationships describing the dependence of m on patterns of propagule dispersal and sampling design have not been available.

Another, closely related, limitation of neutral theory is the current lack of a direct link between spatially implicit and the spatially explicit models that they approximate (Etienne 2005, 2007). Spatially explicit neutral models consist (typically) of two-dimensional landscapes on which species' dispersal is governed by a dispersal kernel—a probability density function that describes the pattern of propagule scatter away from a parent individual (Ribbens *et al.* 1994; Clark *et al.* 1999; Nathan & Muller-Landau 2000; Jones & Muller-Landau 2008). Spatially explicit neutral models have yielded predictions about beta diversity (Chave & Leigh 2002; Condit *et al.* 2002) and species-area distributions (Rosindell *et al.* 2008). In cases where spatially explicit neutral models become analytically intractable or computationally infeasible, spatially implicit models, such as Hubbell's classic model, have been used as mean-field approximations instead (Hubbell 2001; Volkov *et al.* 2007). Spatially implicit neutral models have yielded predictions about alpha diversity and SADs that have, in many cases, found good agreement with real biodiversity data sets (Hubbell 2001; Volkov *et al.* 2003, 2005, 2007; Etienne & Olff 2004). However, without a direct link between the immigration parameters of spatially implicit models and the dispersal parameters of spatially explicit models, it is not possible to assess the degree of error associated with the mean-field approximations.

In this paper, we provide a quantitative biological interpretation for the immigration parameter m in terms of the parameters of a dispersal kernel and the geometry of plot (or plots) defining the local community. The only known previous attempt to quantify this dependence produced numerical results but did not yield simple analytical expressions (Etienne 2005; Appendix S1). Our results allow us to estimate the immigration parameter m from dispersal data, thereby reducing the number of free parameters in the neutral model and facilitating more rigorous tests of its predictions against biodiversity data. We explore two applications of our main analytical results: we conduct a test of neutral theory by comparing estimates of m based on dispersal kernel data to those from previous fits to SADs; and we generate a new testable prediction of neutral theory by investigating how the shapes of SADs change theoretically as the spatial scale of observation (the size of the local community) is increased. Our results are also a step towards the analytical unification of spatially implicit neutral models and the spatially explicit models that they approximate.

INTEGRATION OF DISPERSAL KERNELS INTO THE SPATIALLY IMPLICIT NEUTRAL THEORY

We consider a spatially homogeneous infinite two-dimensional landscape in which sessile individual organisms (such as trees) exist at density ρ per unit area. Each individual releases propagules according to a radially symmetric dispersal kernel. For convenience, we assume that all individuals have the same dispersal kernel and produce the same number of propagules per unit time, and that each propagule that arrives to an open site has the same probability of capturing the site (following the full mathematical proof of our main result in Appendix S1, we show in Appendix S2 that the main result holds independently of these assumptions). When an individual dies, it is replaced by the progeny of a parent individual randomly selected from a dispersal kernel centred at the dying individual. In reality, of course, the dispersal kernels are centred at the parent individuals, but because of the symmetry of the problem in space and time, the two approaches are equivalent. This model ignores many important biological processes, but it is consistent with other spatially explicit neutral models (e.g., Chave & Leigh 2002; Condit *et al.* 2002).

To conceptualize the relationship between our spatially explicit model and the spatially implicit model of Hubbell (2001), imagine a quadrat of area A thrown down somewhere on the infinite spatially homogeneous landscape. The $J = \rho A$ individuals inside the quadrat comprise the 'local community'. It is important to note that in defining the local community in this way, we do not posit any mechanisms governing diversity above those described in the previous paragraph—the local community is simply a convenient scale of observation.

When an individual at location (x, y) in the local community dies, the replacement individual may, by virtue of the random dispersal and recruitment processes, be from within the local community (i.e., within the quadrat) or from outside the local community (i.e., from outside the quadrat). Define $m_{x,y}$ as the probability that the replacement individual at location (x, y) is drawn from outside the local community. This parameter will be highest for individuals on the edges of the quadrat and smallest for individuals at the centre of the quadrat, where $m_{x,y} \approx 0$ for large A. We define m as the average value of $m_{x,y}$ across the whole of the local community as follows:

$$m = \frac{1}{A} \int \int_{\mathbf{A}} m_{x,y} \, dx \, dy \qquad (1)$$

This definition of m is consistent with that of Hubbell (2001) and subsequent researchers. It arises as a simple physical consequence of throwing a quadrat down on the spatially explicit neutral landscape. Although the parameter

m is central to spatially implicit models, its definition is independent of them.

In this section of the manuscript, we first use (1) to derive an analytical approximation for m that is independent of the shape of the local community and the functional form of the dispersal kernel. We then derive an exact expression for m for the special case of a square plot and a bivariate Gaussian dispersal kernel, and we verify that this expression tends to the general approximation as the plot becomes large relative to the mean dispersal distance. We compare the analytical results to the results of numerical simulations. In the Supplementary Information, we present the mathematical details of the derivations and examine the robustness of our results to violations of the neutral assumption that all species have the same dispersal kernel (Appendices S1–S3). We also derive analogous results for the simpler one-dimensional case (Appendix S4), which may provide a reasonable model of ecosystems such as riparian zones and coastlines.

The general case

From eqn 1, we derived a general approximation that relates the immigration parameter m to the mean dispersal distance d and the perimeter P and area A of a plot (see Appendix S1 for the derivation):

$$m \approx \frac{Pd}{\pi A} \tag{2}$$

Expression (2) is the central result of this paper. The approximation assumes only that the dispersal kernel is symmetric with a finite mean dispersal distance, and that the geometry of the plot is defined on a scale somewhat larger that that of the mean dispersal distance (see below). The approximation is independent of the functional form of the dispersal kernel and independent of the shape of the plot. Because the approximation depends only on the mean dispersal distance d and not the functional form of the dispersal kernel, it is in fact robust to violations of the strict neutral assumptions that all species have the same dispersal kernel and exert the same propagule pressure (see Appendix S2).

We used numerical techniques to investigate the rate of convergence of the approximation (2) to the true theoretical value of m (Figs 1–3; Table 1). Given an estimate of the mean dispersal distance d for a given ecological community, our results can be used to determine whether the approximation (2) is valid for circular, square or rectangular plots of a particular size. As a rule of thumb, the edge length of a square plot should be roughly five times greater than the mean dispersal distance for the error in the approximation to be < 10%, and roughly ten times greater than the mean dispersal distance for the error to be < 5% (Fig. 1;
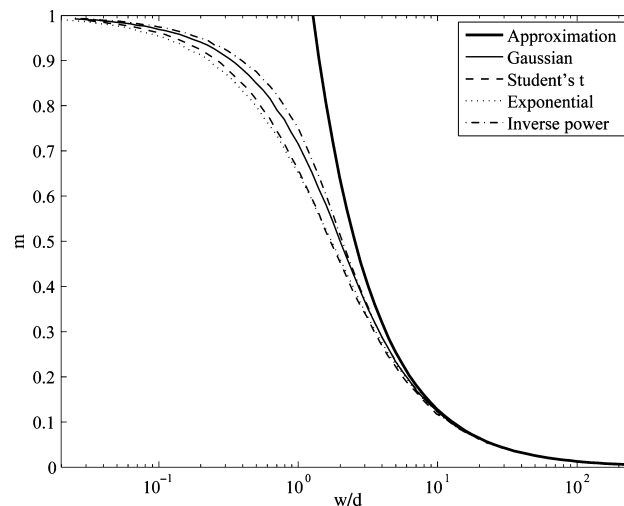

Figure 1 The relationship between the immigration rate m and the ratio of plot edge length w to mean dispersal distance d for a square plot. The approximation (thick unbroken line) given by (2) in the text converges to the values obtained from numerical simulations for Gaussian, Student's t (with $v = 3$ degrees of freedom), exponential and inverse power (with power $a = 5$) dispersal kernels.
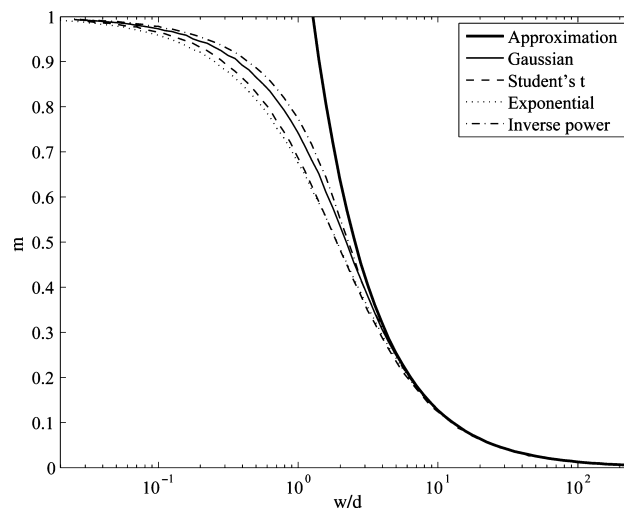

Figure 2 The relationship between the immigration rate m and the ratio of plot diameter w to mean dispersal distance d for a circular plot. The approximation (thick unbroken line) given by (2) in the text converges to the values obtained from numerical simulations for Gaussian, Student's t (with $v = 3$ degrees of freedom), exponential and inverse power (with power $a = 5$) dispersal kernels. Convergence is somewhat faster than for the square plot (Fig. 1).

Table 1). For a circular plot, the diameter should be roughly three times greater than the mean dispersal distance for the error in the approximation to be < 10%, and roughly four
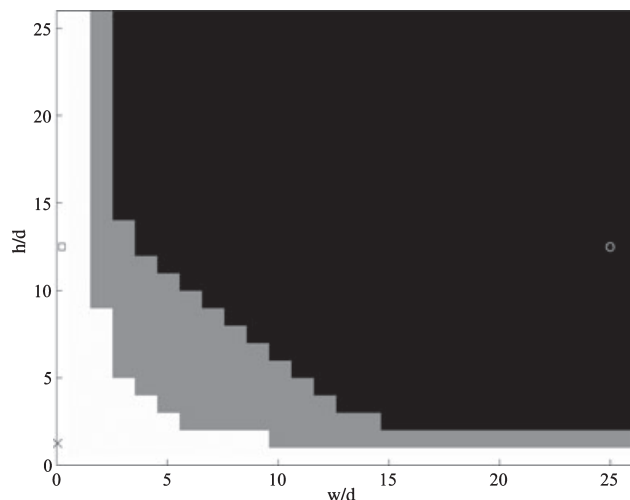
Figure 3 Error associated with estimates of the immigration rate m from the approximation (2) for rectangular plots of different dimensions w by h and for a Gaussian dispersal kernel with mean dispersal distance d. Black areas indicate error < 5%. Grey areas indicate error between 5 and 10%. White areas indicate error > 10%. The circle represents the BCI plot ($w = 1000$ m, $h = 500$ m, $d \approx 40$ m). The square represents a hypothetical long thin forest transect in the same forest ($w \times h = 5$ m $\times$ 500 m; $d \approx 40$ m). The cross represents a short transect in the same forest ($w \times h = 2$ m $\times$ 50 m; $d \approx 40$ m).

Table 1 Minimum ratio of plot dimension w (= edge length for square or diameter for circle) to mean dispersal distance d required for the error in the estimate of m given by (2) in the text to be < 10% or < 5%. Ratios of w/d are estimated from numerical simulations

		Square plot		Circular plot	
	Error threshold	10%	5%	10%	5%
Dispersal kernel	Gaussian	4.5	9.1	2.7	3.6
	Rayleigh	3.6	7.2	2.1	2.7
	Student's t*	6.8	13.7	4.3	6.6
	Exponential	5.5	11.0	3.6	5.0
	Inverse power†	4.1	7.9	2.5	3.5

*With $v = 3$ degrees of freedom.
†With power $a = 5$.

times greater than the mean dispersal distance for the error to be < 5% (Fig. 2; Table 1). Note that these rules of thumb mask considerable variability between dispersal kernels: in particular, larger plots are required for the approximation to be accurate if the dispersal kernel is fat-tailed (e.g., Student's t distribution; Table 1). For rectangular plots, the accuracy of the approximation depends both on the width and height of the plot (Fig. 3).

A special case: square plot and bivariate Gaussian dispersal kernel

We derived an exact analytical expression for the immigration parameter m for the special case of a square plot of edge length w and Gaussian dispersal described by the bivariate density function:

$$f(x,y) = \frac{1}{\sigma^2 \pi} \exp\left(-\frac{(x-x_0)^2 + (y-y_0)^2}{\sigma^2}\right)$$

The exact expression for m is (see Appendix S3 for derivation):

$$m = 1 - \left\{ \text{erf}\left(\frac{w\sqrt{\pi}}{2d}\right) - \left(1 - \exp\left[-\frac{\pi w^2}{4d^2}\right]\right) \frac{2d}{\pi w} \right\}^2 \quad (3)$$

where erf is the error function (Abramowitz & Stegun 1972). As the edge length of the plot w becomes large relative to the mean dispersal distance d, the value of the error function approaches one and the value of the exponential approaches zero, so we have the following approximation:

$$m \approx \frac{4d}{\pi w} = \frac{Pd}{\pi A}$$

which matches the approximation for the general case given by (2), as expected.

Given that values of m for several different biologically plausible dispersal kernels are consistent to within about 10% (Figs 1 and 2), eqn 3 may provide suitable estimates of m for small plots even when the functional form of the dispersal kernel is unknown (the simpler eqn 2 makes no assumptions about the functional form of the dispersal kernel but can only be used for large plots).

APPLICATIONS

Estimating the immigration parameter from dispersal data

To date, the parameter m of the neutral theory has almost always been treated as a free parameter and estimated indirectly from fits to SADs. With our new analytical results we are now able to estimate m directly from dispersal data (Etienne (2005) performed such an exercise using numerical results; see Appendix S5 for a discussion of his fundamental dispersal number in the context of our results). If the SAD in a given community is the result of a neutral process, then the estimate of m from dispersal data should be similar to the estimate of m fitted to the SAD, thereby providing a test of neutral theory.

On Barro Colorado Island (BCI) the mean dispersal distance d, based on seed-trap data for 81 tree species in the plot, is 39.5 m and the 95% confidence interval for d is (32.8, 46.7) (confidence interval estimated by bootstrap-resampling raw data in Muller-Landau 2001; see also Condit et al. 2002).

The plot sizes at which the approximation (2) becomes useful are thus fairly modest on BCI: it is valid to within about 10% in square plots of area 4 ha, and to within about 5% on square plots of area 16 ha (calculations based on Table 1). The approximation is better in circular plots than in square plots: it is valid to within about 10% in circular plots of area 1 ha, and to within 5% in circular plots of area 2 ha (calculations based on Table 1). The 50 ha permanent BCI forest plot is rectangular with dimensions 1000 m × 500 m, which is well above the size for which the approximation is valid (Fig. 3). Plugging A = 500 000 m^2 (50 ha), perimeter P = 3000 m and d in the range (32.8, 46.7) into approximation (2) gives a mean estimate of $m \approx 0.075$ and a 95% confidence interval for m of (0.063, 0.089). This interval does not quite overlap with estimates of m = 0.1 (Hubbell 2001), and m = 0.098–0.133 (Etienne 2005) based on species abundance data from the same plot, and almost overlaps with another estimate m = 0.09 (Volkov et al. 2007).

We also applied our approximation (2) to estimate m from dispersal data for three Peruvian tropical floodplain species (Clark et al. 1999). We used their estimated mean dispersal parameters from fits of a bivariate Student's t distribution to estimate the values of m that would correspond to 50 ha local communities comprised of three Peruvian tropical floodplain species (Table 2). The estimates of m (0.0058–0.0109) were smaller than those estimated from other tropical forests (Volkov et al. 2007), and this is directly attributable to the lower estimates of dispersal distance for the Peruvian trees. Thus, we predict that SADs from this ecosystem should have a stronger interior mode characteristic of dispersal limitation.

Although approximation (2) performs well for large forest plots, it is not useful for long, thin forest transects (Fig. 3), such as the 500 m × 5 m Amazonian forest transects used by Tuomisto et al. (2003) or the 50 m × 2 m transects used in the Gentry method (Gentry 1982, 1988). In the Supporting Information we include a MATLAB program that calculates m numerically for plots of arbitrary size and shape and for arbitrary dispersal kernels. We used this program to calculate m = 0.888 for a hypothetical 500 m × 5 m forest transect and m = 0.956 for a hypothetical 50 m × 2 m forest transect on BCI with a Gaussian dispersal kernel and mean dispersal distance d = 40 m (the approximation (2) generates non-sensical values m > 1 for these plots). Note that values for m for the long thin plots are much larger than $m \approx 0.075$ for the 50 ha plot—this is attributable to smaller plot sizes (relative to the dispersal distance) and greater edge effects.

Spatial scale dependence of species abundance distributions

Many tests of neutral theory have focussed on fitting theoretical SADs to ecological data (e.g., Hubbell 2001; McGill 2003; Volkov et al. 2003, 2005, 2007). Our results presented here allow us to investigate, for the first time, how the theoretical shape of an SAD changes as the spatial scale of observation is increased (i.e., as the size of a plot, or local community, is increased within a homogeneous metacommunity; see Appendices S6 and S7 for the mathematical details). Note that there is a separate well studied issue of how the shape of an SAD changes as sample size increases on a fixed spatial scale (i.e., for a fixed community size) (Preston 1948; Chisholm 2007; Green & Plotkin 2007; Volkov et al. 2007). Here, we ignore the sampling issue and assume that the SADs represent exhaustively sampled plots. There is another separate, but less well studied, issue of how the theoretical shape of an SAD changes when the data set comprises multiple spatially separated subplots (Latimer et al. 2005; Etienne 2007). We investigate this in the context of our main results in Appendix S8.

An expression for the SAD (the expected number of species with abundance n) in a local community with J individuals and immigration rate m is given by:

$$<\varphi_n> = \theta \frac{x^n}{n!} \int_0^\infty \frac{\Gamma(n+y)}{\Gamma(1+y)} e^{-\omega y} dy \qquad (4)$$

where θ is the fundamental biodiversity number, $\omega = \theta(1-m)/(Jm) - \ln(1-x)$, and x is the ratio of the per-capita birth rate to the per-capita death rate and can be determined from the constraint $J = \sum_{n=1}^{\infty} n<\varphi_n>$ (Volkov

Table 2 t-Distributed dispersal kernel parameters and corresponding estimates of m for three Peruvian tropical floodplain tree species

	u (m^2)	P	σ (m)*	d (m)†	A (ha)‡	P (m)‡	ρ (m^{-2})‡	m
Calycophyllum spruceanum	195	2.94	6.3	4.8	50	3000	0.043	0.0092
Hyeronima laxiflora	17	0.82	5.2	3.0	50	3000	0.043	0.0058
Virola sebifera	163	1.82	7.9	5.7	50	3000	0.043	0.0109

*This is given by $u = \sigma^2(2p-1)$ as described in Appendix S4.
†This is given by eqn (S8) in Appendix S4.
‡These parameters are set to the same values as the BCI plot.

et al. 2007). Combining (4) with our main result (2) allows us to study the spatial scaling behaviour of the local community SAD (Fig. 4; Appendix S6). Previously, this was not possible because the quantitative dependence of *m* on *J* was unknown. For small *J* the SAD is a left-skewed humped distribution, and as *J* increases the left-skewness decreases and the height of the hump gradually increases (Fig. 4). In the limit as *J* becomes large the local community SAD approaches a log-series distribution:

$$<\varphi_n> \rightarrow \psi \frac{x^n}{n}$$

where $\psi = \theta / \ln(1/(1-x))$ (see Appendix S7 for derivation), consistent with other theoretical and empirical observations of log-series distributions at large scales (Hubbell 2001; Volkov *et al.* 2007).

The success of neutral models in fitting observed SADs does not constitute a robust test of neutral theory, because neutral and niche-structured assembly rules can result in similar SADs (Purves & Pacala 2005). Spatial scaling of SADs may provide a more robust test of neutral theory: the predicted changes in SADs (Fig. 4; Appendix S6) could be tested with different-sized plots from a given ecosystem.

Determining the robustness of this test (i.e., the novelty of the predictions relative to niche-structured models) and developing it in detail (e.g., test statistics for nested vs. non-nested plots) is beyond the scope of this paper.

DISCUSSION

We have quantified how the immigration parameter *m* in the classic spatially implicit formulation of neutral theory depends on the parameters of a dispersal kernel and the geometry of the plot that defines the local community. Our results facilitate more rigorous confrontation of neutral theory with biodiversity data, because we can now estimate the parameter *m* from dispersal data (it was previously fit as a free parameter). Furthermore, our results allow us to predict how neutral SADs should vary across spatial scales (i.e., local community or plot sizes).

We have provided several tools for relating *m* to dispersal data. Our analytical approximation for large plots states that *m* is proportional to the perimeter of the plot and the mean dispersal distance and inversely proportional to the area of the plot. The approximation is independent of the specific geometry of the plot defining the local community and
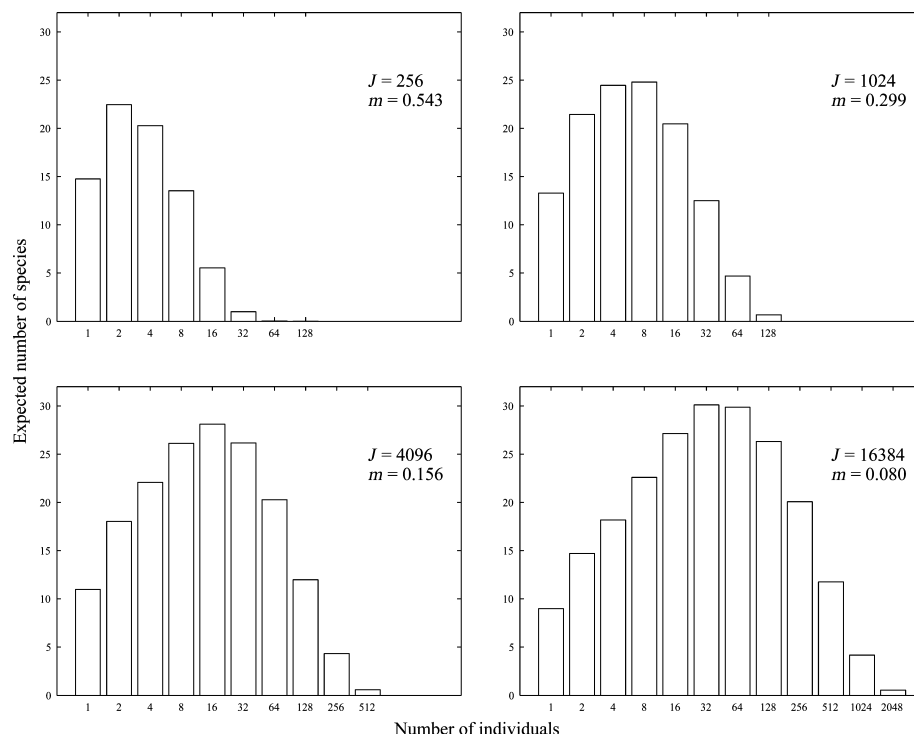

Figure 4 Theoretical species abundance distributions (SADs) for local communities of different sizes (*J*) within a homogeneous metacommunity with mean dispersal distance $d = 39.5$ m, density of individuals per unit area $\rho = 0.043$ m^{-2} and biodiversity number $\theta = 50$ (parameter values correspond roughly to tropical forest on BCI). The immigration parameter *m* is calculated from the other parameters assuming a bivariate Gaussian dispersal kernel (eqn 3). Species are binned into log$_2$ abundance categories following the method of Preston (1948).

independent of the shape of the dispersal kernel (provided that the dispersal kernel is radially symmetric). The plot size (relative to the mean dispersal distance) at which the approximation becomes accurate is smaller for circular plots than for square plots, suggesting that analytical and empirical work specifically targeted at neutral theory would benefit from the use of circular plots (although unfortunately large circular plots are more difficult to sample in the field than large rectangular plots). For cases where the approximation breaks down (small plots; long, thin plots; local communities comprising multiple disconnected plots), we have provided a computer program to estimate m numerically from dispersal parameters and plot geometry.

Although m is useful for analytical purposes, such as deriving the theoretical functional forms of SADs (e.g., Volkov *et al.* 2003), we propose that the mean dispersal distance d is a more suitable fundamental parameter because it is scale invariant, biologically meaningful and easily related to m analytically (see Appendix S5 for an investigation of the scale dependence of another proposed dispersal parameter for neutral theory, I (Etienne 2005, 2009)). In the limit as community size becomes sufficiently large (a few hectares for tropical forest plots), the theoretical spatially implicit SAD for a local community embedded in a homogeneous metacommunity is completely determined by the local community size J and two scale invariant parameters: d and the fundamental biodiversity number θ. Even for smaller community sizes, the value of m and hence the shape of the SAD is determined largely by the value of d for biologically realistic choices of the dispersal kernel (Figs 1 and 2).

The predictions of spatially implicit neutral theory appear robust to violations of the assumption that all species in a community have the same dispersal kernel: our main approximation (2) holds even when different species have different dispersal kernels and exert different propagule pressures, provided that the mean dispersal distance d is calculated as the average of the mean dispersal distances of different species (weighted by the species' relative abundances and propagule pressures; see Appendix S2). This is important, because the assumption that species have identical dispersal kernels is patently false (Ribbens *et al.* 1994; Clark *et al.* 1998; Muller-Landau *et al.* 2008). Our finding of robustness to interspecific variation in dispersal kernels provides some justification for a fundamental premise of neutral theory, which is that inter and intraspecific trait variation averages out statistically at large scales and is not important for predicting macroscopic patterns of diversity (Chave 2004, p250–251).

In this paper, we have not attempted a thorough and extensive application of the theoretical results to field estimates of dispersal kernels and neutral model parameters, but have instead limited our analysis to a few basic checks to verify that the theoretical results produce sensible values.

We found that previous estimates of m from neutral model fits to SADs on the 50 ha BCI plot (Hubbell 2001; Etienne 2005) are within about 25% of the theoretical value based on mean dispersal distances (Muller-Landau 2001; Condit *et al.* 2002). A possible explanation for the discrepancy is that the seed-trap methods used to parameterize dispersal kernels tend to underestimate the frequency of long-distance dispersal events (Jones & Muller-Landau 2008) and therefore underestimate the mean dispersal distance. Molecular evidence also suggests that mean dispersal distances of neotropical trees are greater than those estimated from seed-trap methods (Jones *et al.* 2005; Hardy *et al.* 2006). The estimates of m from species abundance data may also be unreliable, because very different combinations of the parameters m and θ (the fundamental biodiversity number) can lead to similarly good fits (Etienne 2007). Of course, if the estimates of m (or, equivalently, of d) from dispersal and species abundance data within a given ecosystem cannot ultimately be reconciled (e.g., by improved estimates of seed dispersal parameters), it may be necessary to invoke non-neutral mechanisms to describe the processes governing species abundances in this ecosystem. In other words, the estimate based on species abundances is unbiased only if the neutral model assumptions are correct, in which case the two estimates should yield the same answer (subject to sampling errors). In contrast, if the neutral model assumptions are violated, then the species abundance method is biassed, and we would not expect the two estimates to agree.

Our estimates of the immigration parameter m from dispersal data in Peruvian floodplain forests were low compared to those from BCI data, and this led us to predict that SADs from the Peruvian forests should have a stronger interior mode characteristic of dispersal limitation. Plot-scale SADs for this floodplain ecosystem could be used to test this prediction (aggregated data, but not raw plot-scale data, are available in Pitman *et al.* 1999).

The future development of neutral ecological theory requires the unification of the spatially implicit and explicit models (Etienne 2005, 2007). The spatially implicit models are mean-field approximations of the spatially explicit models, but, without a direct link between them, it has not previously been possible to assess the approximation errors. Our work provides this link. For example, our results could be used to assess errors in local community SADs in Hubbell's (2001) classic spatially implicit model by direct comparison to theoretical SADs from a large, spatially explicit landscape (Rosindell *et al.* 2008). This comparison would entail (1) fixing a mean dispersal distance d for the spatially explicit model; (2) calculating the corresponding value of m for the spatially implicit model using the methods described in this paper; and (3) calculating SADs for local communities in both the spatially implicit and explicit models (in the latter case, these local communities are 'plots'

within the simulated landscape). Assuming that the effects of stochasticity are removed from the analysis, any differences between the SADs from the two models would reflect errors in the spatially implicit approximation. Thus, our quantification of the relationship between the spatially implicit model's immigration parameter and the spatially explicit model's dispersal kernel is a step towards analytical unification of the two versions of neutral theory.

ACKNOWLEDGEMENTS

REFERENCES

Abramowitz, M. & Stegun, I.A. (1972). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover, New York.

Adler, P.B. (2004). Neutral models fail to reproduce observed species-area and species-time relationships in Kansas grasslands. *Ecology*, 85, 1265–1272.

Bell, G. (2001). Ecology – neutral macroecology. *Science*, 293, 2413–2418.

Chave, J. (2004). Neutral theory and community ecology. *Ecol. Lett.*, 7, 241–253.

Chave, J. & Leigh, G.L.J. (2002). A spatially explicit neutral model of β-diversity in tropical forests. *Theor. Popul. Biol.*, 62, 153–168.

Chisholm, R.A. (2007). Sampling species abundance distributions: resolving the veil-line debate. *J. Theor. Biol.*, 247, 600–607.

Chisholm, R.A. & Burgman, M.A. (2004). The unified neutral theory of biodiversity and biogeography: comment. *Ecology*, 85, 3172–3174.

Clark, J.S., Macklin, E. & Wood, L. (1998). Stages and spatial scales of recruitment limitation in southern Appalachian forests. *Ecol. Monogr.*, 68, 213–235.

Clark, J.S., Silman, M., Kern, R., Macklin, E. & HilleRisLambers, J. (1999). Seed dispersal near and far: patterns across temperate and tropical forests. *Ecology*, 80, 1475–1494.

Condit, R., Pitman, N., Leigh, E.G., Chave, J., Terborgh, J., Foster, R.B. *et al.* (2002). Beta-diversity in tropical forest trees. *Science*, 295, 666–669.

Etienne, R.S. (2005). A new sampling formula for neutral biodiversity. *Ecol. Lett.*, 8, 253–260.

Etienne, R.S. (2007). A neutral sampling formula for multiple samples and an 'exact' test of neutrality. *Ecol. Lett.*, 10, 608–618.

Etienne, R.S. (2009). Maximum likelihood estimation of neutral model parameters for multiple samples with different degrees of dispersal limitation. *J. Theor. Biol.*, 257, 510–514.

Etienne, R.S. & Olff, H. (2004). A novel genealogical approach to neutral biodiversity theory. *Ecol. Lett.*, 7, 170–175.

Gentry, A.H. (1982). Patterns of neotropical plant-species diversity. *Evol. Biol.*, 15, 1–85.

Gentry, A.H. (1988). Changes in plant community diversity and floristic composition on environmental and geographical gradients. *Ann. Mo. Bot. Gard.*, 75, 1–34.

Green, J.L. & Plotkin, J.B. (2007). A statistical theory for sampling species abundances. *Ecol. Lett.*, 10, 1037–1045.

Hardy, O.J., Maggia, L., Bandou, E., Breyne, P., Caron, H., Chevallier, M.H. *et al.* (2006). Fine-scale genetic structure and gene dispersal inferences in 10 Neotropical tree species. *Mol. Ecol.*, 15, 559–571.

Hubbell, S.P. (1979). Tree dispersion, abundance, and diversity in a tropical dry forest. *Science*, 203, 1299–1309.

Hubbell, S.P. (2001). *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton University Press, Princeton, NJ.

Jabot, F. & Chave, J. (2009). Inferring the parameters of the neutral theory of biodiversity using phylogenetic information and implications for tropical forests. *Ecol. Lett.*, 12, 239–248.

Jones, F.A. & Muller-Landau, H.C. (2008). Measuring long-distance seed dispersal in complex natural environments: an evaluation and integration of classical and genetic methods. *J. Ecol.*, 96, 642–652.

Jones, F.A., Chen, J., Weng, G.J. & Hubbell, S.P. (2005). A genetic evaluation of seed dispersal in the neotropical tree Jacaranda copaia (Bignoniaceae). *Am. Nat.*, 166, 543–555.

Latimer, A.M., Silander, J.A. & Cowling, R.M. (2005). Neutral ecological theory reveals isolation and rapid speciation in a biodiversity hot spot. *Science*, 309, 1722–1725.

Leigh, E.G. (2007). Neutral theory: a historical perspective. *J. Evol. Biol.*, 20, 2075–2091.

McGill, B.J. (2003). A test of the unified neutral theory of biodiversity. *Nature*, 422, 881–885.

Muller-Landau, H.C., Wright, S.J., Calderon, O., Condit, R. & Hubbell, S.P. (2008). Interspecific variation in primary seed dispersal in a tropical forest. *J. Ecol.*, 96, 653–667.

Muller-Landau, H.C. (2001). Seed dispersal in a tropical forest: empirical patterns, their origins, and their consequences for community dynamics. In: *Department of Ecology and Evolutionary Biology*. Ph.D. thesis, Princeton University, Princeton, NJ.

Muneepeerakul, R., Bertuzzo, E., Lynch, H.J., Fagan, W.F., Rinaldo, A. & Rodriguez-Iturbe, I. (2008). Neutral metacommunity models predict fish diversity patterns in Mississippi-Missouri basin. *Nature*, 453, 220–U9.

Munoz, F., Couteron, P., Ramesh, B.R. & Etienne, R.S. (2007). Estimating parameters of neutral communities: from one single large to several small samples. *Ecology*, 88, 2482–2488.

Nathan, R. & Muller-Landau, H.C. (2000). Spatial patterns of seed dispersal, their determinants and consequences for recruitment. *Trends Ecol. Evol.*, 15, 278–285.

Pitman, N.C.A., Terborgh, J., Silman, M.R. & Nuez, P. (1999). Tree species distributions in an upper Amazonian forest. *Ecology*, 80, 2651–2661.

Preston, F.W. (1948). The commonness and rarity of species. *Ecology*, 29, 254–283.

Purves, D.W. & Pacala, S.W. (2005). Ecological drift in niche-structured communities: neutral pattern does not imply neutral process. In: *Biotic Interactions in the Tropics: Their Role in the Maintenance of Species Diversity* (eds Burslem, D., Pinard, M. & Hartley, S.). Cambridge University Press, Cambridge, pp. 107–138.

Ribbens, E., Silander, J.A. & Pacala, S.W. (1994). Seedling recruitment in forests – calibrating models to predict patterns of tree seedling dispersion. *Ecology*, 75, 1794–1806.

Rosindell, J., Wong, Y. & Etienne, R.S. (2008). A coalescence approach to spatial neutral ecology. *Ecol. Inform.*, 3, 259–271.

Tuomisto, H., Ruokolainen, K. & Yli-Halla, M. (2003). Dispersal, environment, and floristic variation of western Amazonian forests. *Science*, 299, 241–244.

Vanpeteghem, D., Zemb, O. & Haegeman, B. (2008). Dynamics of neutral biodiversity. *Math. Biosci.*, 212, 88–98.

Volkov, I., Banavar, J.R., Hubbell, S.P. & Maritan, A. (2003). Neutral theory and relative species abundance in ecology. *Nature*, 424, 1035–1037.

Volkov, I., Banavar, J.R., He, F.L., Hubbell, S.P. & Maritan, A. (2005). Density dependence explains tree species abundance and diversity in tropical forests. *Nature*, 438, 658–661.

Volkov, I., Banavar, J.R., Hubbell, S.P. & Maritan, A. (2007). Patterns of relative species abundance in rainforests and coral reefs. *Nature*, 450, 45.

SUPPORTING INFORMATION

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

Figure S1 A schematic diagram of a plot A, which has only one edge L, of length L, open to dispersal.
Figure S2 A schematic diagram of a plot A = $A_0 \cup A_1 \cup A_2$, which has only one edge L, of length L, open to dispersal. O marks the origin ($x = 0, y = 0$).
Figure S3 The effect of different spatial configurations of plots defining the local community (black areas in panels a and b) on the theoretical shape of the species abundance distribution (panels c and d), assuming the classic spatially implicit neutral model, in which there is a single panmictic local community.

Appendix S1 Derivation of the approximation linking dispersal and immigration.
Appendix S2 Robustness of the approximation to violations of the neutrality assumptions.
Appendix S3 Derivation of the exact expression for a bivariate Gaussian kernel on a square plot.
Appendix S4 Dispersal and immigration in one-dimensional landscapes.
Appendix S5 The relationship between the fundamental dispersal number I and scale.
Appendix S6 Spatial scaling of SADs.
Appendix S7 Limiting behaviour of the local community SAD.
Appendix S8 Species abundance distributions for disconnected local communities.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and 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.