

## LETTER

# A general framework for the distance–decay of similarity in ecological communities

**OnlineOpen:** This article is available free online at [www.blackwell-synergy.com](http://www.blackwell-synergy.com)

Hélène Morlon,<sup>1\*†</sup> George Chuyong,<sup>2</sup> Richard Condit,<sup>3</sup> Stephen Hubbell,<sup>4</sup> David Kenfack,<sup>5</sup> Duncan Thomas,<sup>6</sup> Renato Valencia<sup>7</sup> and Jessica L. Green<sup>1†</sup>

## Abstract

Species spatial turnover, or  $\beta$ -diversity, induces a decay of community similarity with geographic distance known as the distance–decay relationship. Although this relationship is central to biodiversity and biogeography, its theoretical underpinnings remain poorly understood. Here, we develop a general framework to describe how the distance–decay relationship is influenced by population aggregation and the landscape-scale species-abundance distribution. We utilize this general framework and data from three tropical forests to show that rare species have a weak influence on distance–decay curves, and that overall similarity and rates of decay are primarily influenced by species abundances and population aggregation respectively. We illustrate the utility of the framework by deriving an exact analytical expression of the distance–decay relationship when population aggregation is characterized by the Poisson Cluster Process. Our study provides a foundation for understanding the distance–decay relationship, and for predicting and testing patterns of beta-diversity under competing theories in ecology.

## Keywords

Beta-diversity, distance–decay relationship, Poisson Cluster Process, sampling biodiversity, Sørensen index, spatial aggregation, spatial turnover, species-abundance distribution, species–area relationship, tropical forests.

*Ecology Letters* (2008) 11: 904–917

## INTRODUCTION

One of the most widely used relationships in spatial biodiversity studies is the distance–decay, which describes how the similarity in species composition between two communities varies with the geographic distance that separates them. This relationship received the early interest

of Whittaker in his seminal study of vegetation in the Siskiyou mountains (Whittaker 1960, 1972) and Preston when examining the Galapagos flora (Preston 1962). The distance–decay relationship became increasingly popular after Nekola & White (1999) formalized its ability to describe, compare and understand biodiversity patterns. Considered one of the few ‘distributions of wealth’

<sup>1</sup>School of Natural Sciences, University of California, Merced, CA, USA

<sup>2</sup>Department of Life Sciences, University of Buea, P.O. Box 63, Buea, Republic of Cameroon

<sup>3</sup>Smithsonian Tropical Research Institute, Box 0843-03092 Balboa, Ancon, Republic of Panama

<sup>4</sup>Ecology and Evolutionary Biology, University of California, Los Angeles, CA, USA

<sup>5</sup>Missouri Botanical Garden, P.O. Box 299, St Louis MO 63166-0299, USA

<sup>6</sup>Department of Forest Science, Oregon State University, Corvallis OR 97331-2902, USA

<sup>7</sup>Laboratory of Plant Ecology, School of Biological Sciences, Aptado, 17-01-2184, Pontificia Universidad Católica del Ecuador, Quito, Ecuador

\*Correspondence: E-mail: [morlon.helene@gmail.com](mailto:morlon.helene@gmail.com)

†Present address: Center for Ecology and Evolutionary Biology, University of Oregon, Eugene, OR, USA.

Reuse of this article is permitted in accordance with the Creative Commons Deed, Attribution 2.5, which does not permit commercial exploitation.

characterizing communities (Nekola & Brown 2007), distance–decay curves have now been studied across a wide range of organisms, geographic gradients and environments (Nekola & White 1999; Condit *et al.* 2002; Tuomisto *et al.* 2003; Green *et al.* 2004; Novotny *et al.* 2007; Qian & Ricklefs 2007; Soininen & Hillebrand 2007).

There are many reasons to explain the success of the distance–decay relationship in ecology. Data required to plot the distance–decay curve are readily obtained by sampling at local scales across a landscape, making large-scale biodiversity studies empirically tractable (Harte *et al.* 1999; Condit *et al.* 2002; Green *et al.* 2004; Krishnamani *et al.* 2004). Because the distance–decay relationship reflects patterns of spatial distribution and autocorrelation, it is likely sensitive to key spatial processes such as dispersal limitation, making it a powerful tool for testing mechanistic ecological theories (Chave & Leigh 2002; Condit *et al.* 2002). Even in the absence of theoretical derivations, distance–decay data can be used to understand the forces driving community turnover patterns such as dispersal limitation and environmental heterogeneity (Tuomisto *et al.* 2003; Ferrier *et al.* 2007; see Legendre *et al.* (2005) and Tuomisto & Ruokolainen (2006) for discussion of statistical approaches). Finally, the recent incorporation of species' evolutionary history in distance–decay approaches offers a novel perspective for investigating the spatial turnover of phylogenetic composition across landscapes (Ferrier *et al.* 2007; Bryant *et al.* in press).

Despite a longstanding interest in the distance–decay relationship, its theoretical foundations remain poorly understood. The first theoretical derivation of the distance–decay relationship was based on dimensional analyses and the assumption of fractal species' spatial distributions (Harte & Kinzig 1997; Harte *et al.* 1999). More recent analyses stemming from the neutral theory of biodiversity provide predictions for the distance–decay relationship in an environmentally homogeneous landscape, under the assumption that species are demographically identical (Hubbell 2001; Chave & Leigh 2002; Condit *et al.* 2002). However, a theoretical framework for the distance–decay relationship free of assumptions about the spatial organization of individuals or community dynamics is still lacking. Such a general framework is necessary to interpret distance–decay curves observed in nature, where no particular clustering or assembly processes can be assumed *a priori*.

Sampling theory provides a foundation for understanding the spatial scaling of diversity with minimal assumptions (McGill *et al.* 2007). Sampling theory has been used to derive scaling relationships for many macroecological patterns including the species–area and endemics–area relationships (He & Legendre 2002; Green & Ostling 2003), the species–abundance distribution (Green & Plotkin 2007) and species

turnover (Plotkin & Muller-Landau 2002). Plotkin & Muller-Landau (2002) paved the way for integrating the distance–decay relationship into sampling theory by deriving the compositional similarity between two samples randomly drawn from a landscape, independent of their spatial location. However, the distance–decay relationship requires understanding how community similarity varies as a function of the geographic distance separating samples, and there currently exists no general sampling formula for this spatial pattern.

In this paper, we merge sampling theory and spatial statistics to develop a framework for understanding the distance–decay relationship. We begin by deriving a general formula for distance–decay as a function of the landscape-scale species–abundance distribution and intraspecific spatial autocorrelation. This general framework does not assume a particular type of population clustering or community dynamics. To illustrate the utility of this framework, we examine a specific model of clustering: the Poisson Cluster Process. This spatial-point process was chosen due to its mathematical tractability (Cressie 1993; Diggle 2003), its ability to reproduce species–area curves (Plotkin *et al.* 2000) and its potential to characterize the dispersal capacity of species (Seidler & Plotkin 2006). We compare our theoretical predictions to empirical data from three tropical forests with distance–decay curves that differ widely in their compositional similarity values, rate of decay and functional form. We conclude by discussing the implication of our results for biodiversity and biogeography studies.

## GENERAL FRAMEWORK

Our interest lies in the similarity between two sampled communities separated by a given geographic distance. We quantify community similarity using the incidence-based Sørensen index, which measures the number of species shared between two communities divided by the average number of species in each community. The analytical derivations outlined below could be readily adapted for other measures of similarity based on species presence/absence or abundance, but we focus on the Sørensen index because it is widely used in ecology (Magurran 2004), has been proposed as a means to estimate the species–area relationship (Harte & Kinzig 1997) and was adopted in the initial developments of beta-diversity sampling theory (Plotkin & Muller-Landau 2002).

### General sampling formula

Deriving a sampling formula for the distance–decay relationship requires knowledge about the abundance and

aggregation of species within a landscape. Biodiversity sampling theory has traditionally assumed that population aggregation is invariant across species (He & Legendre 2002; Plotkin & Muller-Landau 2002; Green & Ostling 2003) or a linear function of population abundance (Green & Plotkin 2007). For generality, we relax this assumption by introducing  $\xi(n, \gamma)$ , the joint probability that a given species in the landscape has abundance  $n$  and a set of clustering parameters  $\gamma$  (e.g. the parameter  $k$  of the negative binomial distribution, or the parameters  $\rho$  and  $\sigma$  of the Poisson Cluster Process).

Let  $\Psi(a, n, \gamma)$  denote the probability that a species with landscape-scale abundance  $n$  and aggregation  $\gamma$  is present in a sample that covers a proportion  $a$  of a landscape. Let  $\Psi^*(a, n, \gamma, d)$  denote the probability that a species with abundance  $n$  and aggregation  $\gamma$  is present in  $a$  situated at distance  $d$  from a focal individual. The expected Sørensen similarity  $\chi(a, d)$  is:

$$\chi(a, d) = \frac{\int \Psi(a, n, \gamma) \Psi^*(a, n, \gamma, d) \xi(n, \gamma) dn d\gamma}{\int \Psi(a, n, \gamma) \xi(n, \gamma) dn d\gamma}. \quad (1)$$

A summary of symbol notations and the theoretical underpinnings for eqn 1 can be found in Appendices SA and SB of the Supporting Information. The occurrence probability  $\Psi(a, n, \gamma)$  is commonly used to quantify macroecological patterns such as species range size distributions and species richness in a sampling area (Gaston & Blackburn 2000). The probability  $\Psi^*(a, n, \gamma, d)$ , which we refer to as the ‘neighbourhood occurrence probability’, is novel but closely related to the classical relative neighbourhood density  $\Omega(d)$  (Fig. 1).  $\Omega(d)$  is defined as the expected density of individuals in an annulus of radius  $d$  and thickness  $\Delta d$  centred on a focal individual, normalized by the density of individuals in the landscape (Condit *et al.* 2000; Ostling *et al.* 2000; Wiegand & Moloney 2004).  $\Omega(d)$  is also known as the pair correlation function in spatial statistics, and is interchangeable with other correlation metrics (Appendix SB).

In Appendix SB, we derive the distance–decay relationship in terms of the correlation metric  $\Omega(d)$ :

$$\chi(a, d) = \frac{\int \Psi(an, \gamma) \Psi(an\Omega(\gamma, d), \gamma) \xi(n, \gamma) dn d\gamma}{\int \Psi(an, \gamma) \xi(n, \gamma) dn d\gamma}. \quad (2)$$

Equation 2 provides the analytical link between abundance, clustering, sample area and the decay of community similarity with distance. Although the derivation of eqns 1 and 2 require the assumption that sampling areas are relatively small compared with the geographic distance separating them (for discussion see Appendix SB), we demonstrate in *Empirical Evaluation* that these equations provide an accurate approximation over a wide range of spatial scales.

## Qualitative predictions

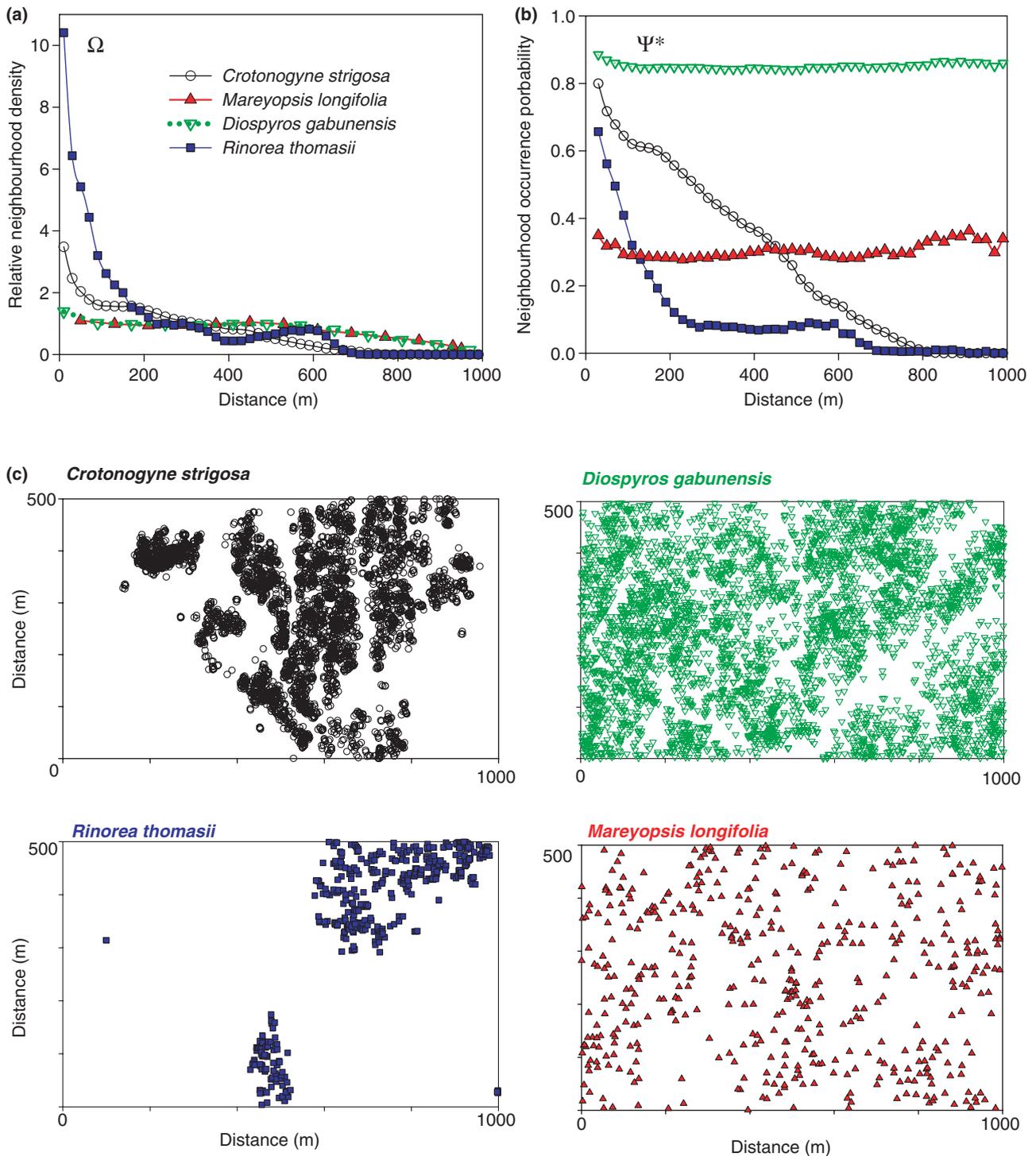
The general sampling formula above (eqn 2) leads to a suite of qualitative predictions that do not require assuming a specific form for the occurrence probability, spatial autocorrelation function, or landscape-scale species abundance distribution. Equation 2 does not involve the total number of species in the landscape, suggesting that the distance–decay relationship is insensitive to species richness. Equation 2 does not involve spatial correlations between species, suggesting that shuffling species in space would not affect the distance–decay relationship. Interspecific aggregation may thus only influence distance–decay curves indirectly through its influence on species’ abundances and intraspecific aggregation. Finally, the contribution of species to the integrals in eqn 2 is weighted by their landscape-scale abundance, suggesting that similarity at any distance is primarily determined by the most abundant species in a landscape and relatively insensitive to the rare ones.

Figure 2 illustrates qualitative predictions related to the influence of abundance, clustering and sample area on the distance–decay relationship. In a hypothetical landscape with even abundances and aggregation, the distance–decay formula suggests that the functional form of the relationship is primarily determined by species’ aggregation as measured by the decay of  $\Omega$  with distance, while landscape-scale species abundances and sample area primarily influence overall similarity (Appendix SB). In biologically realistic landscapes where species differ in their abundance and aggregation, the correlation between these two variables will substantially influence the predictions above. More generally, the aggregation–abundance relationship is expected to play a major role in shaping distance–decay curves. The relative contribution of rare species to the rate of decay is expected to be more important if rare species are highly aggregated, and steep decays should occur in landscapes where the dominant species are highly aggregated.

In *Empirical Evaluation*, we test these qualitative predictions in tropical forests. We now give an example of how the framework presented above can be used to derive the distance–decay relationship when a specific type of population aggregation is assumed.

## APPLICATION: POISSON CLUSTER PROCESS

Spatial statistics have received growing interest among ecologists with the acquisition of spatially explicit data, including the establishment of large tropical forest plots around the globe (John *et al.* 2007; Wiegand *et al.* 2007). Spatial point processes provide powerful tools for characterizing aggregation. The homogeneous Poisson Cluster



**Figure 1** Example of (a) the relative neighbourhood density  $\Omega$  and (b) the neighbourhood occurrence probability curves  $\psi^*$  for (c) four tropical forest species in Korup National Park, Cameroon.  $\Omega$  and  $\psi^*$  are tightly linked: when a species is aggregated (i.e. *Crotonogyne strigosa*, *Rinorea thomasii*), both the relative neighbourhood density  $\Omega$  and the neighbourhood occurrence probability  $\psi^*$  are decreasing functions of distance. When a species is uniformly distributed (i.e. *Diospyros gabunensis*, *Mareyopsis longifolia*), neither  $\Omega$  nor  $\psi^*$  depend on distance. Aggregation mainly influences the shape of  $\psi^*$ , and abundance its overall value. Here,  $\psi^*$  is calculated in a  $20 \times 20$  m quadrat nested in the 50-ha plot ( $a = 0.0008$ ).

Process is one of the simplest, and is described in detail elsewhere (Cressie 1993; Plotkin *et al.* 2000; Diggle 2003). In short, individuals of a species are assumed to be clumped in clusters according to the following rules:

- (1) Cluster centres are randomly distributed in the landscape  $X$  according to a Poisson process with constant density  $\rho$ .
- (2) Each cluster is assigned a random number of individuals, drawn independently from a Poisson distribution with intensity  $\mu$ .
- (3) The position of the individuals relative to the centre of their clusters is drawn independently from a radially symmetric Gaussian distribution  $b$  with variance  $\sigma^2$ , namely

$$b(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right). \quad (3)$$

Intuitively,  $\rho$  reflects the density of clusters,  $\sigma$  their spatial extent and  $\mu$  the number of individuals per cluster. A landscape where the homogeneous Poisson Cluster Process characterizes population aggregation consists of an independent superposition of individual species, so that inter-specific spatial correlations are ignored.

The homogeneous Poisson Cluster Process provides a simple, relatively realistic characterization of population clustering (Plotkin *et al.* 2000). In nature, several processes cause clusters to form. Dispersal limitation is among the strongest, as illustrated in tropical forests by the high correlation between cluster size (as measured by  $\sigma$ ) and a species' mode of dispersal (Seidler & Plotkin 2006). The spatial distribution of clusters depend mainly on environmental heterogeneity (Plotkin *et al.* 2000; Seidler & Plotkin 2006) or secondary dispersal (Wiegand *et al.* 2007) and the parsimonious assumption that clusters are randomly distributed with constant density  $\rho$  may not be accurate. The degree to which the model fails in reproducing empirical patterns in nature should yield insight into the importance of incorporating heterogeneity into the Poisson Cluster model.

In Appendix SC, we derive exact analytical expressions for a species' occurrence probability  $\psi$  and spatial correlation function  $\Omega$  under the Poisson Cluster Process. From eqn 2, we deduce the distance–decay relationship in a landscape where aggregation is characterized by the homogeneous Poisson Cluster Process:

$$\chi(a, d) = \frac{\int (1 - \exp(-anc(A)))(1 - \exp(-anc(A)\Omega(\rho, \sigma, d)))\xi(n, \rho, \sigma)dn d\rho d\sigma}{\int (1 - \exp(-anc(A)))\xi(n, \rho, \sigma)dn d\rho d\sigma} \quad (4)$$

with

$$c(A) = \frac{1}{\mu A} \int_X (1 - \exp(-\mu \int_A b(u - s)du)ds) \quad (5)$$

and

$$\Omega(\rho, \sigma, d) = 1 + \frac{1}{4\pi\rho\sigma^2} \exp\left(-\frac{d^2}{4\sigma^2}\right). \quad (6)$$

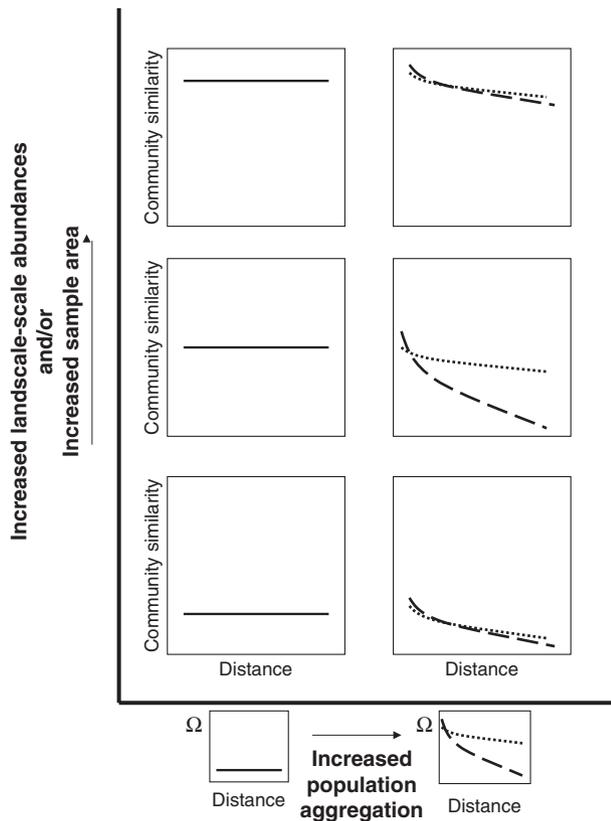
Here,  $b$  is given by eqn 3 ( $u$  and  $s$  represent two-dimensional coordinates in the landscape).  $c$  is a coefficient between 0 and 1 reflecting the deviation of the occurrence probability  $\psi$  from that expected under random placement. In Appendix SC, we derive the analytical link between  $c$  and the parameter  $k$  of the negative binomial distribution. Equation 4 provides the expression for the distance–decay relationship when population aggregation is characterized by the Poisson Cluster Process. The denominator in eqn 4 provides the expression for the species–area relationship.

### EMPIRICAL EVALUATION

We use data from three tropical forests to evaluate the predictions outlined above. First, we examine the qualitative predictions formulated in *General framework*, which make no *a priori* assumptions about population clustering or community dynamics. Second, we test the theoretical predictions derived in *Application: Poisson Cluster Process*. We test the accuracy of eqn 4 and the validity of the homogeneous Poisson Cluster Process as a model of clustering.

### Data

The three forest plots are part of the Center for Tropical Forest Studies network: Barro Colorado Island (Panama, 300 species), Yasuni National Park (Ecuador, 1132 species) and Korup National Park (Cameroon, 494 species). Within the 50-ha plot in Korup National Park and Barro Colorado Island, and the 25-ha plot in Yasuni, every stem > 1cm at breast height has been spatially mapped and identified to species. Detailed description of the plots and references are available on the CTFS web site <http://www.ctfs.si.edu/doc/plots>.



**Figure 2** Conceptual figure illustrating the hypothetical influence of landscape-scale abundances, sampling and population aggregation on the distance–decay relationship, as suggested by eqn 2. We consider abundance  $n$  and sample area  $a$  in parallel because they are expected to have the same effect on the distance–decay relationship (community similarity at a given distance is a function of the average number of individuals in a sample  $an$ ). *From left to right:* with comparable landscape-scale species abundances and sample area, increased aggregation (steeper decays of  $\Omega$  with distance) induces steeper decays in community similarity and lower similarity values at large distances. *From bottom to top:* with comparable aggregation, increased landscape-scale abundances (or equivalently increased sample areas), induce high overall similarity. *Dashed lines:* long dashed lines reflect high aggregation, dotted lines reflect moderate aggregation. In highly aggregated communities, the distance–decay slope can be influenced by abundance and sampling at the boundaries of low and high similarity values.

### General framework

To evaluate the general sampling formula (eqn 2) qualitative predictions, we first examine empirical distance–decay patterns in tropical forests using a sub-setting approach similar to Nekola & White (1999). We divide species into classes based on their landscape-scale abundance or degree of population aggregation (aggregation is measured using the  $\Omega$  statistic in the 0–10 m distance class  $\Omega_{0-10}$  following Condit *et al.* (2000)). We then compare distance–decay

relationships among various subsets of the data (e.g. subsets containing mostly dominant species or highly aggregated species). Second, we compare distance–decay relationships obtained in each forest with different sampling areas, ranging from  $A = 0.0004$  ha to  $A = 6.25$  ha.

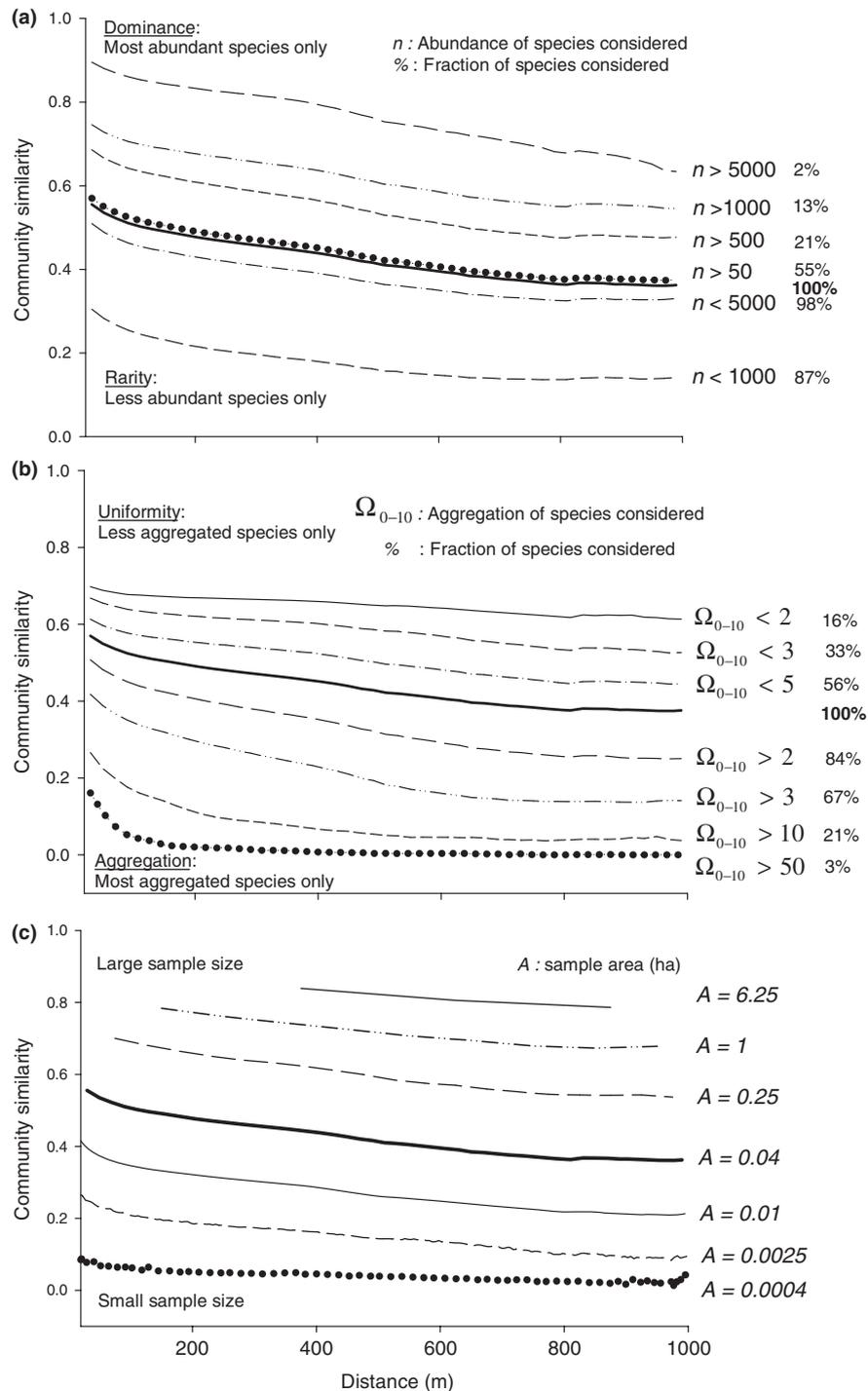
Similar results, consistent with our qualitative predictions, are found in the three forests (see Fig. 3 for results in Korup, and Appendix SD for results in BCI and Yasuni). The distance–decay relationship is mainly driven by the most abundant species in a forest, and is relatively insensitive to the rare ones (Fig. 3a). The functional form of the distance–decay relationship is largely controlled by population aggregation (Fig. 3b). Finally, landscape-scale abundances and sample area influence overall similarity, rather than the rate of decay (Fig. 3c). Although these results are expected from our qualitative predictions, two caveats are in order. First, as we show below, the aggregation metric  $\Omega_{0-10}$  is correlated with landscape-scale abundance in these tropical forests, making it difficult to infer the independent influence of aggregation versus abundance in shaping the distance–decay curves of subcommunities. Second, as illustrated in Fig. 2, sample area and landscape-scale abundances could have a stronger influence on the slope of the distance–decay relationship in landscapes where the degree of aggregation is higher than the forests studied here.

### Application: Poisson Cluster Process

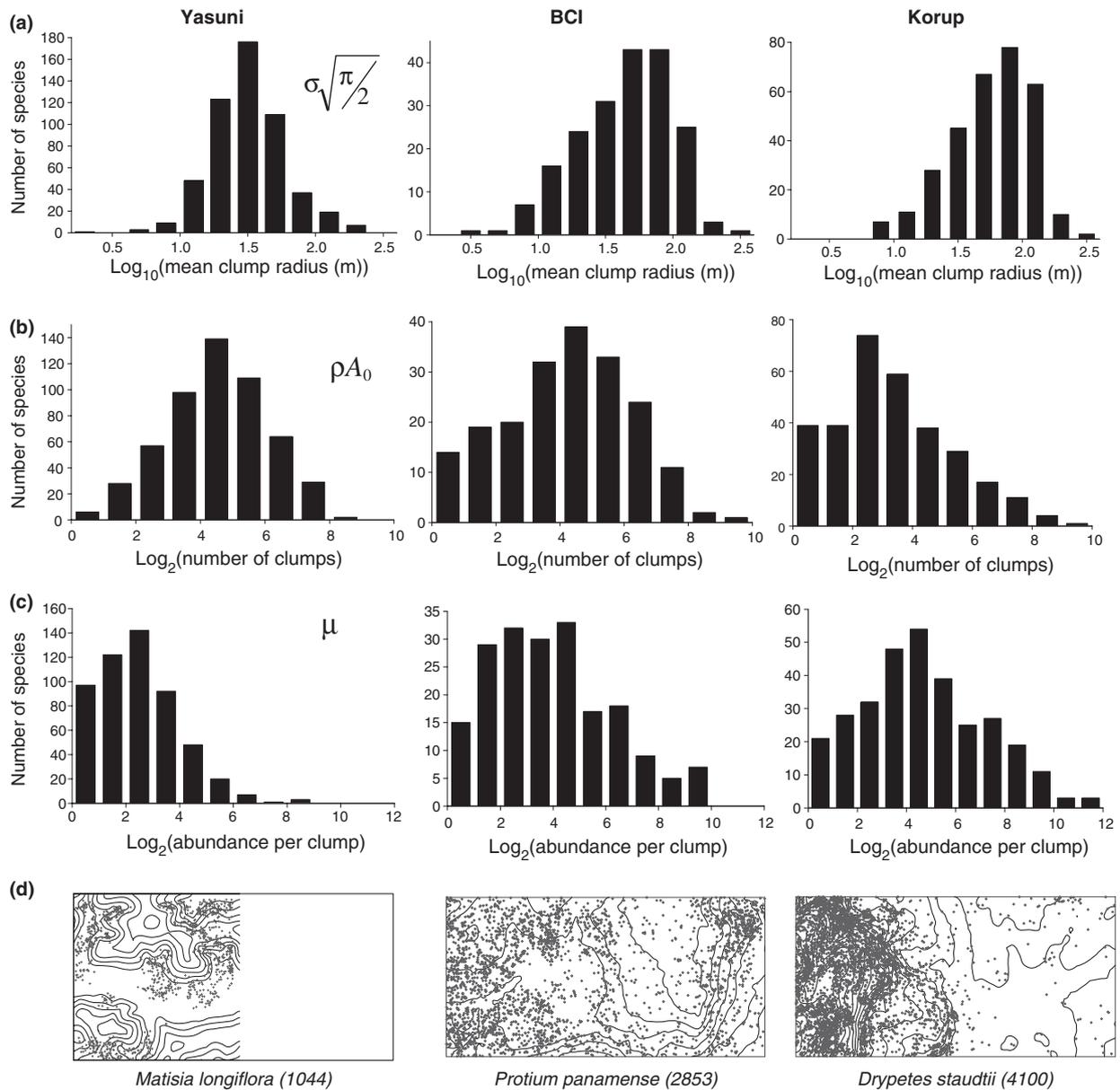
Here we test the accuracy of our analytical derivations (eqn 4) using simulations, and the ability of the homogeneous Poisson Cluster Process to reproduce distance–decay relationships observed in nature. The homogeneous Poisson cluster assumptions may not precisely reflect population aggregation in the forests. BCI is a forest with relatively homogeneous environment and many generalists, where these assumptions are justified. Yasuni and Korup support several habitat types that may influence species clustering patterns in an inhomogeneous way. In *Distance–decay relationships*, we evaluate the relevance of the Poisson Cluster assumptions in the forests.

#### *Clustering in tropical forests*

We fit the Poisson Cluster Process to spatial point data for each species in BCI, Korup and Yasuni (see Appendix SE for parameter estimation details). Figures 4 and 5 reveal important differences about population aggregation patterns among the three forest plots. In Yasuni, conspecifics tend to be grouped into small (Figs 4a and 5a) and numerous (Figs 4b and 5b) clusters containing few individuals (Figs 4c and 5c). This trend gets stronger as abundance increases. In Korup, conspecifics tend to be grouped into large and sparse clusters containing many individuals. These differences may be explained by differences in the ecology of each



**Figure 3** Influence of landscape-scale abundance, population aggregation and sampling on the distance–decay relationship in Korup. (a) An increasing proportion of the rarest (lines going up) or most abundant (lines going down) species are removed from the forest. Removing species with fewer than 50 individuals corresponds to considering only 55% of the landscape-scale species pool, yet this removal has very little effect on the relationship. At the other side of the spectrum, removing only 2% of the most abundant species substantially affects overall similarity. (b) An increasing proportion of the most aggregated (lines going up) or least aggregated (lines going down) species is removed from the forest. Only species with > 50 individuals are considered (Condit *et al.* 2000). (c) Sample area substantially influence rate of decays only at the smallest sample area. In (a) and (b), distance–decay plots correspond to 20 × 20 m samples nested in the 50 ha plot ( $A = 0.04$  ha,  $a = 0.0008$ ). See Appendix SD for similar results in BCI and Yasuni and log-linear plots emphasizing the effect of aggregation.



**Figure 4** Distributions of clustering parameters estimated by the Poisson Cluster Process (a) The distributions of mean clump radius  $\sigma\sqrt{\pi/2}$  appear log-normal (in Yasuni) to right-skewed log-normal (in BCI and Korup); plotted on a linear scale, they are characterized by left-skewed shapes similar to those observed by Plotkin *et al.* (2000) (their fig. 5; see Appendix SE). (b–c) The distributions of number of clumps  $\rho A_0$  and number of individuals per clump  $\mu$  vary greatly between forests: species with few clusters and many individuals per cluster are common in Korup, but scarce in Yasuni, where species tend to be clumped in more clusters with fewer individuals. (d) Topographic maps and typical spatial distributions for trees in Yasuni, BCI and Korup.

site. Korup is divided into two distinct regions: one steep/rocky ridge and one muddy/flat valley. Species tend to specialize in one of the two terrains, forming few large densely populated clumps (Fig. 4d). Environmental heterogeneity such as gullies, steep slopes, flats, wet and dry sections within these terrains likely form nested clusters. The Poisson Cluster Process, designed to characterize one scale of aggregation only, may fail to detect the smaller

nested clumps. In Yasuni, valleys and ridges also constrain the spatial repartition of flora, but they are narrower and less dramatic than in Korup, the soil is more homogeneous, and the species are more generalists (Valencia *et al.* 2004). As a result, species typically have numerous small clusters spanning the entire plot.

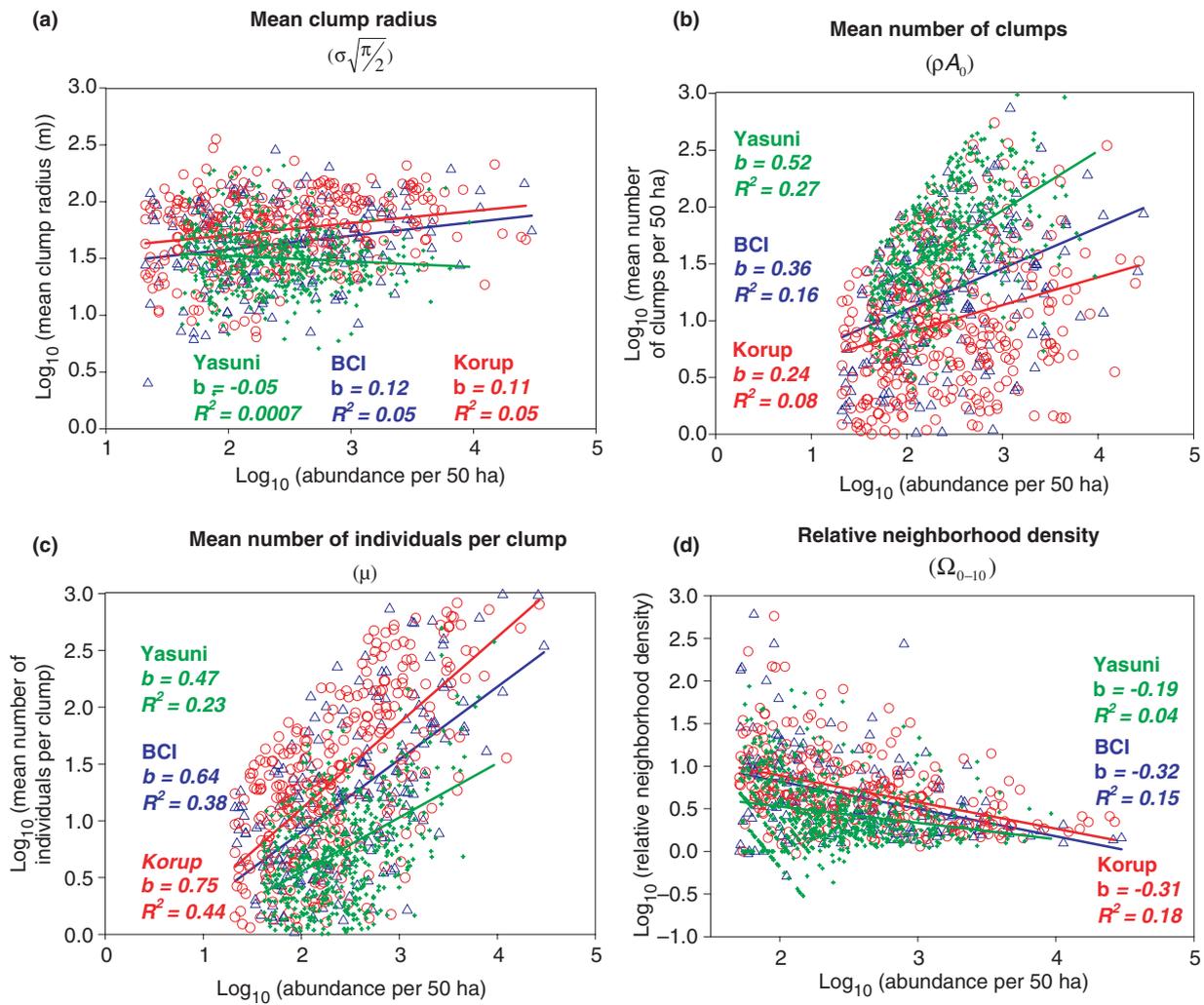
The correlation between clustering and abundance is fundamental in shaping distance–decay curves. Understand-

ing this correlation can also help in formulating hypotheses on the origin of rarity in tropical tree communities (Hubbell 1979). There is no consensus on how aggregation scales with abundance: positive (He *et al.* 1997), negative (Hubbell 1979; Condit *et al.* 2000) and insignificant (Plotkin *et al.* 2000) relationships have been proposed. The correlation between aggregation and abundance depends on how aggregation is quantified. Measuring aggregation in the forests using the mean clump size  $\sigma$  (Fig. 5a), we find a weak correlation between aggregation and abundance, consistent with Plotkin *et al.* (2000). Using the neighbourhood occurrence probability  $\Omega_{0-10}$  (Fig. 5d), we find a negative correlation between aggregation and abundance, consistent with Condit *et al.* (2000). This disparity can be understood from the expression for  $\Omega$  under the Poisson Cluster Process (eqn 6) (see Appendix SF for details). In

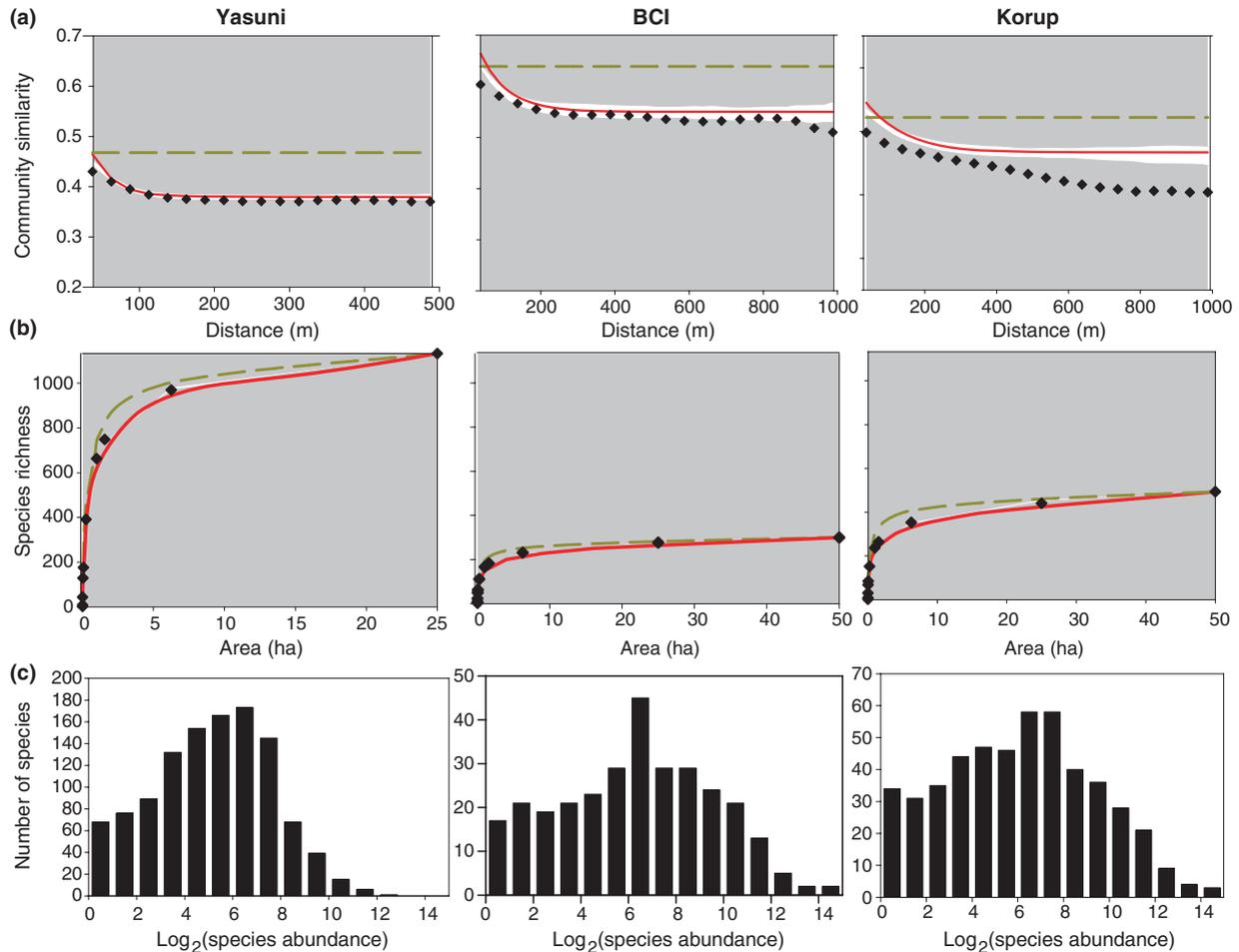
brief,  $\Omega$  reflects both the size of clusters ( $\sigma$ ), which is independent of abundance (Fig. 5a), and their density in the landscape ( $\rho$ ), which is correlated with abundance (Fig. 5b). Analysing the three Poisson Cluster Process parameters ( $\sigma$ ,  $\rho$  and  $\mu$ ) in concert provides the most comprehensive view of the abundance–aggregation relationship. A consequence of the observed high correlation between  $\rho$  and  $n$  relevant to our distance–decay analyses is that aggregation parameters in eqn 4 cannot be assumed invariant across species, thus justifying the consideration of the joint distribution  $\xi(n, \gamma)$ .

*Distance–decay relationships*

Using the data parameterized above, we test eqn 4 and the ability of the homogeneous Poisson Cluster Process to reproduce distance–decay relationships. Figure 6 illustrates the results obtained by sampling 25 × 25 m quadrats from



**Figure 5** Dependence of (a) the mean clump radius  $\sigma\sqrt{\pi/2}$ , (b) the number of clumps  $\rho A_0$ , (c) the mean number of individuals per clump  $\mu$  and (d) the relative neighbourhood density  $\Omega_{0-10}$  on a species' abundance  $n$ . All correlations are significant (Spearman test,  $P < 0.05$ );  $b$ -values correspond to the slope of the log–log regression of the parameters against abundance.



**Figure 6** Comparison of theory with data in Yasuni, BCI and Korup (a) distance–decay curves reported for  $25 \times 25$  m samples ( $A = 0.0625$  ha), (b) species–area curves, (c) species–abundance distributions. The diamonds represent observed data. The red solid lines represent curves predicted by the Poisson Cluster Process (eqn 5). The white area represents the 95% confidence intervals produced by simulation of the Poisson Cluster Process. The green dashed lines represent curves predicted when assuming random placement (Appendix SF). The sensitivity of the results to sample area is presented in Appendix SF.

the landscape. Results for a wider range of sampling areas are presented in Appendix SF. To put our results in context with previous studies (Plotkin *et al.* 2000), we also examine species–area relationships. To test eqn 4, we compare the predicted distance–decay and species area curves to the mean and 95% confidence envelope obtained by simulations of the process (see Appendices SE and SF for details). We find that predictions and simulations agree, with only a slight overestimation for community similarity at small distances, showing that approximations made in eqns 1 and 2 are relevant, and demonstrating the accuracy of the framework and specific derivations under the process.

To test the ability of the Poisson Cluster Process to reproduce distance–decay and species–area relationships, we compare the curves directly obtained from the raw data to

those predicted by eqn 4, and we use simulations to test for the significance of the results (see Appendix SF for statistical methods). Consistent with previous studies (Plotkin *et al.* 2000), we find that the Poisson Cluster Process accurately reproduces observed species–area relationships ( $P > 0.05$ ). The accuracy of the Poisson Cluster Process to reproduce observed distance–decay relationships is less straightforward. The hypotheses that aggregation can be modelled with the process is rejected in the three forests ( $P < 0.05$ ), except in Yasuni and BCI with  $25 \times 25$  m sample areas. The process tends to overestimate similarity values in the forests for small sample areas, and to underestimate them for larger sample areas (see Appendix SF). In Yasuni and BCI the Poisson Cluster Process is nevertheless a reasonable first approximation of clustering patterns. In Korup, however, similarity values are largely

overestimated at any distance-class and all but the  $100 \times 100$  m sample area.

Korup appears to be an outlier: population aggregation in this forest is not well characterized by the simple homogeneous Poisson Cluster Process. The inability of Poisson Cluster Process to reproduce distance–decay relationships in Korup probably lies in its inability to reproduce species' spatial autocorrelation (decays of  $\Omega$  with distance). The species–area relationship, which does not reflect  $\Omega$ , is well reproduced by the process. Species' spatial autocorrelation in Korup may be poorly reproduced as a result of species having more than one scale of aggregation, as suggested by the ecology of the site (see *Clustering in tropical forests*). The shape of the distance–decay curve in Korup supports this hypothesis: the curve is characterized by two distinct range of distances where the decay is steeper ( $0 \leq d \leq 200$  and  $400 \leq d \leq 600$ ), suggesting that two scales of aggregation occur in this forest.

## DISCUSSION

The distance–decay relationship reflects how diversity is spatially distributed and has consequences for conservation and our general understanding of community assembly. Interpreting this relationship and using it to test theories in ecology requires understanding how patterns in the distribution and abundance of species influence its shape. Our general distance–decay framework provides a theoretical foundation for addressing this need. The derivation under the Poisson Cluster Process illustrates a specific application of this general framework, and the efficiency of the distance–decay relationship in falsifying theories.

### General framework

Our distance–decay framework provides a theoretical foundation for interpreting earlier analyses of beta-diversity based on empirical and simulated data. Equation 1 shows that the distance–decay curve follows from a weighted combination of species-level neighbourhood occurrence curves. This prediction is in agreement with neutral theory predictions of a 'compound' curvilinear distance–decay curve (Hubbell 2001). Hubbell (2001) also proposed that the initial steep decay of similarity at short distances is induced by rare species, while the following shallow decay is induced by more abundant ones. In contrast, our results (eqn 3 and Fig. 3a) support the hypothesis that rare species have a weak influence on the distance–decay relationship. These results might be specific to the incidence-based Sørensen index of similarity we considered in our study. However, Nekola & White (1999) measured community similarity with the Jaccard index and also found that removing the rare species in a landscape

(measured as species with low occurrence) does not affect the slope of the relationship. We expect abundance-based metrics to be even less sensitive to the rare species since they give more weight to dominant species. The distance–decay relationship should thus be robust to the potential bias caused by sampling the most abundant species in a landscape, as is common, for example, in microbial ecology.

A central hypothesis stemming from our analyses is that the slope of the distance–decay relationship alone is a poor indicator of species spatial turnover (or  $\beta$ -diversity) and total species richness in a landscape (or  $\gamma$ -diversity). Understanding how turnover in community composition across a landscape relates to the rate of species gain with sampling area has been the focus of many studies (Harte *et al.* 1999; Lennon *et al.* 2001). It is commonly believed that a shallow distance–decay slope reflects a low rate of species turnover, leading to low diversity at large spatial scales. This idea was formalized by Harte *et al.* (1999) in the context of self-similarity and proposed as a means to estimate diversity at large spatial scale from the sampling of small plots (Harte *et al.* 1999; Green *et al.* 2004; Krishnamani *et al.* 2004). Our results suggest that the slope of the distance–decay relationship is a poor indicator of landscape-scale species richness, complementing previous results showing that a significant taxa–area relationship can hold even when the distance–decay relationship is flat (Woodcock *et al.* 2006), or that richness estimators based on the rate of decay in similarity perform poorly (Jobe 2008). For example, Fig. 3a shows that the slope of the distance–decay curve can be conserved even when only a small fraction of the species is considered. Figure 3b shows that the slope of the distance–decay at small spatial scales is the steepest for highly aggregated communities, also known to display the shallowest species–area slopes at this scale (He & Legendre 2002). Finally, Fig. 6 shows that the most species rich forest in our study (Yasuni), has the shallowest distance–decay slope. We suggest that steep decays characterize communities where abundant species are highly aggregated rather than communities with high spatial turnover, and that  $\beta$ -diversity is better described by overall similarity than by rates of decay. We support the idea that the focus on the slope of the relationship (e.g. Qian *et al.* 2005; Qian & Ricklefs 2007) must be expanded to include a focus on intercepts and half-distances (Soininen & Hillebrand 2007), or average similarity (Plotkin & Muller-Landau 2002).

Our analyses illustrate the superiority of the distance–decay to the species–area relationship in testing spatial ecology theories, and provide the analytical basis for deriving expectations for this relationship under competing ecological hypotheses. While species–area relationships can be derived without precise information on species-level spatial autocorrelation, we show that this information is

crucial in shaping distance–decay curves (eqn 2), suggesting that distance approaches are particularly informative of spatial structure in ecological communities. Analytical derivations for species-level spatial autocorrelations exist under theories such as neutrality (in absence of speciation) (Houchmandzadeh & Vallade 2003), self-similarity (Ostling *et al.* 2000), multiscale or inhomogeneous point processes (Diggle *et al.* 2007; Wiegand *et al.* 2007). These expectations could be combined with our framework to predict the community level distance–decay relationship expected under different scenarios of spatial organization.

### Poisson Cluster Process

Specific derivations under the Poisson Cluster may inform future research aimed at understanding the role of dispersal mechanisms in shaping the decay of similarity in ecological communities. Nekola & White (1999) first noted that the mode of dispersion influences distance–decay slopes, with more vagile communities displaying a shallower decay. Hubbell's (2001) neutral theory predicts that dispersal limitation and speciation alone can drive species turnover in a homogeneous landscape. Finally, source-sink meta-communities predict a decrease in beta-diversity with increasing dispersal (Mouquet & Loreau 2003). The Poisson Cluster Process is phenomenological, not mechanistic, and should not be used as a model of community assembly (but see Potts *et al.* 2004; John *et al.* 2007). However, the parameter  $\sigma$  reflecting the size of clusters is strikingly correlated with the dispersal capacity of species (Seidler & Plotkin 2006), and is incorporated explicitly in our expression for the distance–decay relationship (eqn 4). The equation along with findings by Nekola & White (1999) and neutral theory (Hubbell 2001; Chave & Leigh 2002) suggests that strong dispersal limitation (small  $\sigma$  values) induces a steep decay in community similarity.

Combining distance–decay analyses to the Poisson Cluster Process reveals limitations of this process as a model of clustering that had not been previously demonstrated. After it was shown that tropical tree populations are spatially aggregated (Hubbell 1979; He *et al.* 1997; Condit *et al.* 2000), Plotkin *et al.* (2000) proposed that randomly distributed population clusters (i.e. the Poisson Cluster Process) could be a good model of spatial organization and showed that this model accurately reproduced species area curves in 50 ha tropical forest plots, a result reproduced in our study (Fig. 6). The comparison of the distance–decay relationships observed in the forests to those produced by the Poisson Cluster Process suggests that this process does not universally reproduce clustering patterns. Several assumptions underlying the homogeneous Poisson Cluster Process are violated in natural systems. First, the Poisson Cluster Process assumes one scale of aggregation only, while

ecological processes act at multiple spatial scales (e.g. adaptation to a heterogeneous landscape, dispersal limitation, intra- and inter-specific competition, facilitation and localized pest pressure) to induce nested clustering (Levin 1992; Plotkin *et al.* 2002; Cornell *et al.* 2007; Scanlon *et al.* 2007; Wiegand *et al.* 2007). Second, the process assumes a constant density of conspecifics across the landscape, whereas abundances are known to vary widely across a species' range (Brown *et al.* 1995). It is therefore not surprising that the Poisson Cluster Process performs better in a more homogeneous environment (e.g. BCI), or when clumps span the landscape despite environmental heterogeneity (e.g. Yasuni), than when the density of trees is inhomogeneously distributed in the landscape (e.g. Korup).

The limits of the Poisson Cluster Process outlined above should not overshadow its utility, and the benefits gained from merging this model with sampling theory. Although the Poisson Cluster Process is not mechanistic and does not always reproduce patterns accurately, considering this process allowed us to develop theoretical basis for introducing spatial statistics into  $\beta$ -diversity studies. This approach could be extended to integrate processes across spatial scales, which remains a major challenge in ecology. To capture biodiversity patterns at both small and large scales, the assumption of a constant density of individuals over space, as specified by the homogeneous Poisson Cluster Process, could be relaxed. One could consider an inhomogeneous Poisson Cluster Process (Diggle *et al.* 2007), allowing the intensity of the process to vary with environmental variables, or to follow a 'peak and tail' distribution (McGill & Collins 2003) with population abundance 'hotspots' across the landscape. Considering the Poisson Cluster Process allowed for the analytical derivation, in a common framework, of two of the most widely studied spatial biodiversity patterns in ecology. This first step towards theoretically linking the increase of richness with area and the decay of community similarity with distance offers the promise of estimating diversity at large spatial scale with feasible sampling effort.

### CONCLUSION

Our study illustrates the power of the distance–decay relationship in falsifying models, and renders the relationship analytically tractable, offering a promising framework for testing theories in ecology. Theoretical ecology has placed great emphasis on the species-abundance distribution and species–area relationship, leaving the distance–decay relationship largely ignored. Our analyses provide a unified framework for systematic analysis of spatial biodiversity patterns in relation to abundance and aggregation that may inform future research aimed at understanding how biodiversity is distributed and maintained.

## ACKNOWLEDGEMENTS

We are grateful to Joshua Plotkin and John Harte for their comments on this work throughout its development. We also thank Jérôme Chave and three anonymous referees for very useful comments on our manuscript. We acknowledge Brendan Bohannan, Jessica Bryant, Hector Garcia-Martin and James O'Dwyer for stimulating discussions. This work was supported in part by the U.S. Department of Energy, Office of Science, Offices of Advanced Scientific Computing Research and Biological & Environmental Research through the U.C. Merced Center for Computational Biology thanks to Mike Colvin, who also provided scientific support. J. L. G. acknowledges support from the National Science Foundation. The tropical forests studied in this manuscript are part of the Center for Tropical Forest Science, a global network of large-scale demographic tree plots. We gratefully acknowledge the contribution of the many dedicated people and generous founders that have made this data set possible.

## REFERENCES

- Brown, J.H., Mehlman, D.W. & Stevens, G.C. (1995). Spatial variation in abundance. *Ecology*, **76**, 2028–2043.
- Bryant, J., Lamanna, C., Morlon, H., Kerkhoff, A.J., Enquist, B. & Green, J. (2008, in press) Microbes on mountainsides: contrasting elevational patterns of bacterial and plant diversity. *PNAS*.
- Chave, J. & Leigh, E.G. (2002). A spatially explicit neutral model of beta-diversity in tropical forests. *Theor. Popul. Biol.*, **62**, 153–168.
- Condit, R., Ashton, P.S., Baker, P., Bunyavejchewin, S., Gunatilleke, S., Gunatilleke, N. *et al.* (2000). Spatial patterns in the distribution of tropical tree species. *Science*, **288**, 1414–1418.
- 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.
- Cornell, H.V., Karlson, R.H. & Hughes, T.P. (2007). Scale-dependent variation in coral community similarity across sites, islands, and island groups. *Ecology*, **88**, 1707–1715.
- Cressie, N.A.C. (1993). *Statistics for Spatial Data*. Wiley & Sons, New York.
- Diggle, P.J. (2003). *Statistical Analyses of Spatial Point Patterns*, 2nd edn. Edward Arnold, London.
- Diggle, P.J., Gomez-Rubio, V., Brown, P.E., Chetwynd, A.G. & Gooding, S. (2007). Second-order analysis of inhomogeneous spatial point processes using case-control data. *Biometrics*, **63**, 550–557.
- Ferrier, S., Manion, G., Elith, J. & Richardson, K. (2007). Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. *Divers. Distrib.*, **13**, 252–264.
- Gaston, K.J. & Blackburn, T.M. (2000). *Pattern and Process in Macroecology*. Blackwell Science, Malden.
- Green, J.L. & Ostling, A. (2003). Endemics–area relationships: the influence of species dominance and spatial aggregation. *Ecology*, **84**, 3090–3097.
- Green, J.L. & Plotkin, J.B. (2007). A statistical theory for sampling species abundances. *Ecol. Lett.*, **10**, 1037–1045.
- Green, J.L., Holmes, A.J., Westoby, M., Oliver, I., Briscoe, D., Dangerfield, M. *et al.* (2004). Spatial scaling of microbial eukaryote diversity. *Nature*, **432**, 747–750.
- Harte, J. & Kinzig, A.P. (1997). On the implications of species–area relationships for endemism, spatial turnover, and food-web patterns. *Oikos*, **80**, 417–427.
- Harte, J., McCarthy, S., Taylor, K., Kinzig, A. & Fischer, M.L. (1999). Estimating species–area relationship from plot to landscape scale using species spatial-turnover data. *Oikos*, **86**, 45–54.
- He, F. & Legendre, P. (2002). Species diversity patterns derived from species–area models. *Ecology*, **83**, 1185–1198.
- He, F., Legendre, P. & LaFrankie, J. (1997). Distribution patterns of tree species in a Malaysian tropical rain forest. *J. Veg. Sci.*, **8**, 105–114.
- Houchmandzadeh, B. & Vallade, M. (2003). Clustering in neutral ecology. *Phys. Rev. E*, **68**, 061912.
- 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.
- Jobe, R.T. (2008). Estimating landscape-scale species richness: reconciling frequency- and turnover-based approaches. *Ecology*, **89**, 174–182.
- John, R., Dalling, J.W., Harms, K.E., Yavitt, J.B., Stallard, R.F., Mirabello, M. *et al.* (2007). Soil nutrients influence spatial distributions of tropical tree species. *Proc. Natl Acad. Sci. USA*, **104**, 864–869.
- Krishnamani, R., Kumar, A. & Harte, J. (2004). Estimating species richness at large spatial scales using data from small discrete plots. *Ecography*, **27**, 637–642.
- Legendre, P., Borcard, D. & Peres-Neto, P. (2005). Analyzing beta diversity: partitioning the spatial variation of community composition data. *Ecol. Monogr.*, **4**, 435–450.
- Lennon, J.J., Koleff, P., Greenwood, J.J.D. & Gaston, K.J. (2001). The geographical structure of British bird distributions: diversity, spatial turnover and scale. *J. Anim. Ecol.*, **70**, 966–979.
- Levin, S.A. (1992). The problem of pattern and scale in ecology. *Ecology*, **73**, 1943–1967.
- Magurran, A.E. (2004). *Measuring Biological Diversity*. Blackwell Publishing, Malden.
- McGill, B. & Collins, C. (2003). A unified theory for macroecology based on spatial patterns of abundance. *Evol. Ecol. Res.*, **5**, 469–492.
- McGill, B.J., Etienne, R.S., Gray, J.S., Alonso, D., Anderson, M.J., Benecha, H.K. *et al.* (2007). Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. *Ecol. Lett.*, **10**, 995–1015.
- Mouquet, N. & Loreau, M. (2003). Community patterns in source-sink metacommunities. *Am. Nat.*, **162**, 544–557.
- Nekola, J.C. & Brown, J.H. (2007). The wealth of species: ecological communities, complex systems and the legacy of Frank Preston. *Ecol. Lett.*, **10**, 188–196.
- Nekola, J.C. & White, P.S. (1999). The distance decay of similarity in biogeography and ecology. *J. Biogeogr.*, **26**, 867–878.
- Novotny, V., Miller, S.E., Hulcr, J., Drew, R.A.I., Basset, Y., Janda, M. *et al.* (2007). Low beta diversity of herbivorous insects in tropical forests. *Nature*, **448**, 692–697.

- Ostling, A., Harte, J. & Green, J.L. (2000). Self-similarity and clustering in the spatial distribution of species. *Science*, 290, 671a.
- 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. *J. Theor. Biol.*, 207, 81–99.
- Plotkin, J., Chave, J. & Ashton, P. (2002). Cluster analysis of spatial patterns in Malaysian tree species. *Am. Nat.*, 160, 629–644.
- Potts, M., Davies, S., Bossert, W., Tan, S. & Supardi, M. (2004). Habitat heterogeneity and niche structure of trees in two tropical rain forests. *Oecologia*, 139, 1432–1439.
- Preston, F.W. (1962). The canonical distribution of commonness and rarity: Part II. *Ecology*, 43, 410–432.
- Qian, H., Ricklefs, R.E. & White, P.S. (2005). Beta-diversity of angiosperms in temperate floras of eastern Asia and eastern North-America. *Ecol. Lett.*, 8, 15–22.
- Qian, H. & Ricklefs, R.E. (2007). A latitudinal gradient in large-scale beta diversity for vascular plants in North America. *Ecol. Lett.*, 10, 737–744.
- Scanlon, T.M., Caylor, K.K., Levin, S.A. & Rodriguez-Iturbe, I. (2007). Positive feedbacks promote power-law clustering of Kalahari vegetation. *Nature*, 449, 209–213.
- Seidler, T.G. & Plotkin, J. (2006). Seed dispersal and spatial pattern in Tropical trees. *PLoS Biol.*, 4, 2132–2137.
- Soininen, J.M.R. & Hillebrand, H. (2007). The distance decay of similarity in ecological communities. *Ecography*, 30, 3–12.
- Tuomisto, H. & Ruokolainen, K. (2006). Analyzing or explaining beta diversity? Understanding the targets of different methods of analysis. *Ecology*, 11, 2697–2708.
- Tuomisto, H., Ruokolainen, K. & Yli-Halla, M. (2003). Dispersal, environment, and floristic variation of Western Amazonian forests. *Science*, 299, 241–244.
- Valencia, R., Foster, R., Villa, G., Condit, R., Svenning, J.C., Hernandez, C., Romoleroux, K., Losos, E., Magard, E. & Balslev, H. (2004). Tree species distributions and local habitat variation in the Amazon: large forest plot in Eastern Ecuador. *J. Ecol.*, 92, 214–229.
- Whittaker, R.H. (1960). Vegetation of the Siskiyou mountains, Oregon and California. *Ecol. Monogr.*, 30, 279–338.
- Whittaker, R.H. (1972). Evolution and measurement of species diversity. *Taxon*, 21, 213–251.
- Wiegand, T. & Moloney, K.A. (2004). Rings, circles, and null-models for point pattern analysis in ecology. *Oikos*, 104, 209–229.
- Wiegand, T., Gunatilleke, S., Gunatilleke, N. & Okuda, T. (2007). Analyzing the spatial structure of a Sri Lankan tree species with multiple scales of clustering. *Ecology*, 88, 3088–3102.
- Woodcock, S., Curtis, T.P., Head, I.M., Lunn, M. & Sloan, W.T. (2006). Taxa–area relationships for microbes: the unsampled and the unseen. *Ecol. Lett.*, 9, 805–812.

## SUPPORTING INFORMATION

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

**Appendix SA** Symbol notations used in the distance–decay theoretical framework.

**Appendix SB** Derivations of the general sampling formula for the distance–decay relationship and related predictions.

**Appendix SC** Poisson Cluster Process derivations.

**Appendix SD** Figures relevant to empirical evaluation of *Qualitative predictions* in BCI, Yasuni and Korup.

**Appendix SE** Estimation and distribution of Poisson Cluster Process parameters.

**Appendix SF** Supplementary information relevant to *Empirical evaluation*.

**Appendix SG** Supplementary material references.

Please note: Blackwell publishing is not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

Editor, Jerome Chave

Manuscript received 21 December 2007

First decision made 23 January 2008

Manuscript accepted 4 April 2008

**APPENDIX A. Symbol notations used in the distance-decay theoretical framework**

Symbol	Definition
$A_0$	area of the landscape
$S_0$	total number of species in the landscape
$A$	area of a sample
$a$	sample fraction area = $\frac{A}{A_0}$
$n$	species landscape scale abundance
$\gamma$	set of parameters reflecting a species aggregation
$\xi$	joint distribution of species abundance and aggregation
$\chi$	Sørensen index of similarity : fraction of species shared between two communities
$\psi$	occurrence probability: probability that a species is present in a sample
$\psi^*$	neighborhood occurrence probability: probability that a species is present in a sample in the neighborhood of a conspecific
$\Omega$	relative neighborhood density : density of individuals in an annulus around a conspecific normalized by the density of individuals in the landscape
$\rho$	density of clusters in the landscape under a Poisson Cluster Process
$\sigma$	$\sigma\sqrt{\pi/2}$ : mean clump radius $\sigma^2$ :variance of the Gaussian distribution under a Poisson Cluster Process
$\mu$	average number of individuals per cluster under a Poisson Cluster Process

**APPENDIX B. Derivations of the general sampling formula for the distance-decay relationship and related predictions**

**B1. Generalization of Plotkin & Muller Landau's (2002) sampling formula of beta-diversity (derivation of Equation 1)**

Consider two samples whose areas constitute proportions  $a$  and  $b$ , respectively, of a landscape. With the notations introduced in the manuscript, Plotkin & Muller-Landau's formula for the expected fraction of species shared between two samples, also known as the classic incidence-based Sørensen index, takes the form (their Equation 10):

$$\chi(a,b) = \frac{\int \psi(a,n,\gamma)\psi(b,n,\gamma)\xi(n,\gamma)dnd\gamma}{\frac{1}{2} \left[ \int \psi(a,n,\gamma)\xi(n,\gamma)dnd\gamma + \int \psi(b,n,\gamma)\xi(n,\gamma)dnd\gamma \right]} \quad [\text{B.1}]$$

$\xi(n,\gamma) = \phi(n)\theta(\gamma|n)$ , where  $\phi(n)$  is the landscape scale species abundance distribution, and  $\theta(\gamma|n)$  is the distribution of species clustering in the abundance class  $n$ . Alternatively,  $\xi(n,\gamma) = \theta(\gamma)\phi(n|\gamma)$  where  $\theta(\gamma)$  is the distribution of species clustering and  $\phi(n|\gamma)$  is the distribution of abundances in the aggregation class  $\gamma$ . The integrals in the numerator and denominator of Equation B.1 are integrals with respect to both  $n$  and  $\gamma$ . The integral over  $n$  goes from 0 to  $\infty$ . The boundaries with respect to  $\gamma$  depend on the measure chosen for aggregation. This formula assumes that samples are drawn from one landscape characterized by a single species-abundance distribution, but one could relax

this hypotheses and consider the joint distribution of abundances in multiple biomes or habitat types (Plotkin & Muller-Landau 2002).

Equation B.1 provides the expected community similarity between any two samples randomly drawn from the landscape, regardless of their spatial location. To derive an expectation of the decay in similarity with geographic distance  $d$ , we introduce the “neighborhood occurrence probability”,  $\psi^*(a, n, \gamma, d)$ , defined as the probability, for a species with landscape-scale abundance  $n$  and aggregation parameter  $\gamma$ , that at least one conspecific occurs in a sample area constituting a fraction  $a$  of the landscape situated at a distance  $d$  from a focal individual. When the sample size  $a$  is small compared to the landscape, the neighborhood occurrence probability is equivalent to the probability that at least one conspecific occurs in sample area  $a$  situated at a distance  $d$  from a non-overlapping sample of area  $a$  containing the species.

The expected Sørensen similarity between two non-overlapping samples whose areas constitute proportion  $a$  of a landscape and separated by a distance  $d$  is thus given by:

$$\chi(a, d) = \frac{\int \psi(a, n, \gamma) \psi^*(a, n, \gamma, d) \xi(n, \gamma) dn d\gamma}{\int \psi(a, n, \gamma) \xi(n, \gamma) dn d\gamma} \quad [\text{B.2}]$$

Equation B.2 will likely break down for large sample sizes, when the occurrence probability at a distance  $d$  from a sample cannot be approximated by the neighborhood occurrence probability  $\psi^*$ . Equation B.2 shows that the shape of the decay curve is

principally the result of the compound shapes of the curves describing the decay of neighborhood occurrence probability with distance for each species, weighted by the species occurrence probability. Interspecific heterogeneity and multiple spatial scales of aggregation will likely cause the functional form of  $\psi^*$  to vary widely between species and across geographic scales, making the distance decay relationship a compound of various curves with potentially different functional forms.

## **B2. Relationship between the neighborhood occurrence probability $\psi^*$ and $\Omega$**

### **(derivation of Equation 2)**

Here, we derive the relationship between a species' neighborhood occurrence probability  $\psi^*(a, n, \gamma, d)$  and a species' occurrence probability  $\psi(a, n, \gamma)$ , and the spatial autocorrelation function  $\Omega(d)$ . Recall that  $\psi^*(a, n, \gamma, d)$  is the expected probability that a species with abundance  $n$  and aggregation parameters  $\gamma$  occurs in a sample  $a$  situated at a distance  $d$  from an arbitrary individual. Consider an annulus with inner radius  $d$  and area  $A(d)$  at a distance  $d$  around an individual, and a sampled region in the annulus whose area constitutes proportion  $a$  of the landscape  $A_0$ . The fraction of annulus constituted by

this sample is:  $a \frac{A_0}{A(d)}$ , and, by definition of  $\Omega(d)$ , the number of individuals in the

annulus is:  $n\Omega(d) \frac{A(d)}{A_0}$ . We make the reasonable assumption that clustering in the

annulus can be modeled by the set of parameters  $\gamma$  that reflects clustering in the landscape. For a species with abundance  $n$  and set of aggregation parameters  $\gamma$ , the expected neighborhood occurrence probability in sample  $a$  at a distance  $d$  from an individual is given by:

$$\psi^*(a, n, \gamma, d) = \psi\left(a \frac{A_0}{A(d)}, n\Omega(d) \frac{A(d)}{A_0}, \gamma\right) \quad [\text{B.3}]$$

If the occurrence probability is a function of the average number of individuals in a sample  $an$  (example: Poisson and negative binomial distribution, Poisson Cluster Process; counter-example: self-similarity), we can write:

$$\psi(a, n, \gamma) = \psi(an, \gamma)$$

and:

$$\psi^*(a, n, \gamma, d) = \psi(an\Omega(\gamma, d), \gamma) \quad [\text{B.4}]$$

Equation B.4 is intuitive: it shows that the probability that a species with abundance  $n$  and clustering  $\gamma$  is present in a sample area  $a$  at a distance  $d$  from a conspecific is equivalent to the probability that a species with abundance  $n\Omega(d)$  and same clustering  $\gamma$  is present in a sample area  $a$ . It follows that:

$$\chi(a, d) = \frac{\int \psi(an, \gamma) \psi(an\Omega(\gamma, d), \gamma) \xi(n, \gamma) dn d\gamma}{\int \psi(an, \gamma) \xi(n, \gamma) dn d\gamma} \quad [\text{B.5}]$$

### B3. General framework assumptions

Equation B.2 (Equation 1 in the manuscript) makes the assumption that the probability of occurrence in a sample situated at a distance  $d$  from an *occupied sample* is equivalent to the probability of occurrence in a sample situated at a distance  $d$  from a *conspecific* ( $\psi^*$ ). This assumption requires that sample sizes are relatively small.

Equation B.4 (and thus Equation B.5 and Equation 2 in the manuscript) makes the assumption that the density of individuals in a sample area  $a$  situated at a distance  $d$  of an individual is constant in the sample. Therefore, Equation 4 assumes that  $a$  is sufficiently small relative to  $d$ .

Hence, our derivations are expected to be valid for relatively small sample sizes and relatively large distances. In practice, distance-decay curves are commonly designed to study the spatial organization of community composition at large spatial scales, with small sample areas and wide geographic coverage. The accuracy of the equations is tested in Figure 6 in the manuscript and Appendix F below.

#### **B4. Other measures of conspecific spatial autocorrelation and their relationship to the neighborhood occurrence probability $\Omega$**

Our derivations use the statistic  $\Omega(d)$  to describe spatial autocorrelation. Here, we provide the exact analytical link between  $\Omega(d)$  and two other measures of spatial autocorrelation: 1) the correlation function  $\rho(d)$ , defined as the probability that two individuals from a species are at a distance  $d$  from each other 2) the Ripley's statistic  $K(d)$ , defined as the number of individuals in a circle of radius  $d$ , normalized by the density of individuals.

##### Relationship between $\Omega(d)$ and $\rho(d)$

Consider a species with total abundance  $n$  in a landscape of area  $A_0$ . Consider annuli with inner radius  $d$  and area  $A(d)$  around each individual, and let  $\langle n(d) \rangle$  be the average number of individuals in those annuli. By definition:

$$\Omega(d) = \frac{\langle n(d) \rangle / A(d)}{n / A_0} \quad [\text{B.6}]$$

$$\langle n(d) \rangle \text{ is given by: } \langle n(d) \rangle = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \delta(d(i, j) \in [d - \Delta d, d + \Delta d])$$

Here  $\delta$  is Dirac's delta function ( $\delta(B) = 1$  if statement  $B$  is true, 0 otherwise)

By definition,  $\rho(d) = \frac{p(d)}{p}$  where  $p(d)$  is the number of pairs separated by a distance

comprised between  $d - \Delta d$  and  $d + \Delta d$  and  $p$  is the total number of pairs.

$$\text{The total number of pairs } p \text{ is given by } p = \frac{n(n-1)}{2}$$

$$p(d) \text{ is given by: } p(d) = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \delta(d(i, j) \in [d - \Delta d, d + \Delta d])$$

$$\text{Hence, } \rho(d) = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j=1}^n \delta(d(i, j) \in [d - \Delta d, d + \Delta d])$$

$$\text{And } \langle n(d) \rangle = (n-1)\rho(d)$$

Finally:

$$\Omega(d) = \frac{(n-1)\rho(d) / A(d)}{n / A_0} \quad [\text{B.7}]$$

### Relationship between $\Omega(d)$ and $K(d)$

Let  $\langle N(d) \rangle$  be the average number of individuals in the circles of radius  $d$  surrounding each individual in the landscape, and let  $\lambda$  be the density of individuals in the landscape.

$$\text{By definition: } K(d) = \frac{1}{\lambda} \langle N(d) \rangle$$

In the annulus delimited by radii  $d_1$  and  $d_2$ , by definition:

$$\Omega(d_{1,2}) = \frac{\langle N(d_2) \rangle - \langle N(d_1) \rangle}{\lambda(\pi d_2^2 - \pi d_1^2)}$$

$$\text{Thus: } \Omega(d_{1,2}) = \frac{K(d_2) - K(d_1)}{\pi(d_2^2 - d_1^2)} \quad [\text{B.8}]$$

and

$$\Omega(d) = \lim_{d_1 \rightarrow d_2 = d} \Omega(d_{1,2}) = \lim_{d_1 \rightarrow d_2 = d} \frac{K(d_2) - K(d_1)}{\pi(d_2^2 - d_1^2)}$$

## B5. General framework predictions

To interpret Equation B.5 (Equation 2 in the manuscript), it may be more intuitive to consider its discrete equivalent:

$$\chi(a, d) = \frac{\sum_{i=1}^{S_0} \psi(an_i, \gamma_i) \psi(an_i \Omega_i(d), \gamma_i, d)}{\sum_{i=1}^{S_0} \psi(an_i, \gamma_i)} \quad [\text{B.9}]$$

Here,  $S_0$  is the total number of species in the landscape, and  $i$  stands for species  $i$ .

In a hypothetical community where all the species have same abundance  $n$  and aggregation  $\gamma$ ,  $\chi(a, d) = \psi(an\Omega(\gamma, d), \gamma)$ , so that as a first approximation, the dependence of similarity on abundance, aggregation, and sampling should follow that of the neighborhood occurrence probability  $\psi(an\Omega(\gamma, d), \gamma)$ . The distance-dependence of community similarity is entirely embedded in the distance-dependence of the neighborhood occurrence density  $\Omega(\gamma, d)$  for each species (distance does not appear

anywhere else in the formula than in  $\Omega$ , so that the functional form of the distance-dependence of similarity is entirely determined by the functional form of the decay of  $\Omega(\gamma, d)$  with distance.  $\psi(an\Omega(\gamma, d), \gamma)$  at large distances is a decreasing function of aggregation, so that more aggregated communities are expected to display lower similarity values at large distances.  $\psi(an\Omega(\gamma, d), \gamma)$  at any distance increases with landscape-scale abundance  $n$  (or equivalently sample area  $a$ ), so that high similarity values across the range of distances are expected when species have high landscape-scale abundances (or sampled with large sample areas), while low similarity values across the range of distances are expected when species have low landscape-scale abundances.

In a hypothetical community where all the species have same landscape-scale abundance  $n$  and aggregation  $\gamma$ , the slope of the distance-decay relationship at a given distance is expected to vary as  $an \frac{\partial \Omega(\gamma, d)}{\partial d} \psi'(an\Omega(\gamma, d), \gamma)$  (Here  $\psi'$  is the derivate of  $\psi$ ).  $\psi'$  is expected to be a decreasing function, making it difficult to predict how the slope varies with  $an$  (the slope is a product of an increasing function and a decreasing function of the product  $an$ ). We would need to assume a functional form for  $\psi$  to assess the effect of sample area and landscape-scale abundances on the slope of the relationship. However, in the three tropical forests studied in the manuscript, results presented in the section *Empirical Evaluation* suggest that landscape-scale abundances and sample area have a weak influence on the rate of decay, except possibly for extreme values of landscape-scale abundances or sample areas.

### **APPENDIX C. Poisson Cluster Process derivations**

Here, basing upon Cressie (1993), we derive a species occurrence probability  $\psi(a, n, \gamma)$  and  $\Omega(d)$  under a Poisson Cluster Process with parameters  $\rho$ ,  $\sigma$  and  $\mu$  and bivariate Gaussian Kernel  $h$ .

#### **C1. Occurrence probability**

Let  $N$  be any point process on space  $X$ . The probability generating function of  $N$  is defined by:

$$G_N(\omega) \equiv E \left[ \exp \left\{ \int_X \log \omega(s) N(ds) \right\} \right]$$

Where  $\omega$  is any function  $X \rightarrow [0,1]$

Let  $t$  be any real number,  $x \in X$ , and  $\omega_t(x) = 1 - (1-t)\delta(x \in A)$

$$G_N(\omega_t) \equiv E \left[ \exp \left\{ \int_A \log t N(ds) \right\} \right] = E(t^{N(A)})$$

Therefore,  $t \rightarrow G_N(\omega_t)$  is the generating function of  $N(A)$

And  $P(N(A) = 0) = G_N(\omega_t)|_{t=0}$

The occurrence probability in  $A$  is:

$$\psi(a, n, \rho, \sigma) = 1 - P(N(A) = 0) = 1 - G_N(\omega_t)|_{t=0} \quad [\text{C.1}]$$

The probability generation function of the Poisson Cluster Process defined in the manuscript is (directly derived from Cressie 1993):

$$G_N(\omega) = \exp \left\{ -\rho \int_X \left[ 1 - \eta \left\{ \int_X \omega(u+s)h(u)du \right\} \right] ds \right\} \quad [\text{C.2}]$$

Where  $\eta$  is the probability generating function of the number of events per cluster.

With  $\omega_t(x) = 1 - (1-t)\delta(x \in A)$

$$\int_X \omega_t(u+s)h(u)du = \int_X \omega_t(u)h(u-s)du = \int_A h(u-s)du + t \int_A h(u-s)du$$

Equation C.2 gives the generating function of  $N(A)$ :

$$G_N(\omega_t) = \exp \left\{ -\rho \int_X \left[ 1 - \eta \left\{ \int_A h(u-s)du + t \int_A h(u-s)du \right\} \right] ds \right\}$$

For any  $s$  in  $X$ :  $\int_A h(u-s)du + \int_A h(u-s)du = \int_X h(u-s)du = 1$

$$\text{Thus, } G_N(\omega_t) = \exp \left\{ -\rho \int_X \left[ 1 - \eta \left\{ 1 + (t-1) \int_A h(u-s)du \right\} \right] ds \right\}$$

The number of offspring per parent is Poisson-distributed with mean  $\mu$ , so that:

$$\eta(t) = \exp(\mu(t-1))$$

$$\text{and } \eta \left\{ 1 + (t-1) \int_A h(u-s)du \right\} = \exp(\mu(t-1) \int_A h(u-s)du)$$

$$\text{Thus, } G_N(\omega_t) = \exp \left\{ -\rho \int_X \left[ 1 - \exp(\mu(t-1) \int_A h(u-s)du) \right] ds \right\}$$

And from Equation C.1 we obtain the expression of the occurrence probability in a patch

under a Poisson Cluster Process:

$$\psi(a, n, \rho, \sigma) = 1 - \exp \left\{ -\rho \int_X \left[ 1 - \exp(-\mu \int_A h(u-s)du) \right] ds \right\}$$

With  $n = \rho\mu A_0$ , we find that the occurrence probability in a sample of area  $A$  for a species' with abundance  $n$  and parameters  $\rho$  and  $\sigma$  is given by:

$$\psi(a, n, \rho, \sigma) = 1 - \exp[-anc(A)] \quad [C.3]$$

With:

$$c(A) = \frac{1}{\mu A} \int_X (1 - \exp(-\mu \int_A h(u-s) du)) ds. \quad [C.4]$$

The coefficient  $c(A)$  ranges between 0 and 1 and reflects clustering at sample scale  $A$ . It depends on  $\sigma$  and the number  $\mu$  of individuals per cluster. The integral  $\int_A h(u-s) du$  measures the probability that a “parent” with coordinate  $s$  gives rise to an “offspring” in sample  $A$ .

## C2. Omega statistic

The derivation of Ripley's K statistic under the Poisson Cluster Process is known (Cressie 1993; Plotkin *et al.* 2000):

$$K(d) = \pi d^2 + \frac{1}{\rho} (1 - \exp(-d^2 / 4\sigma^2)) \quad [C.5]$$

Combining Equation B.8 and Equation C.5 gives:

$$\Omega(d_{1,2}) = 1 - \frac{1}{\pi\rho} \left[ \frac{\exp(-\frac{d_2^2}{4\sigma^2}) - \exp(-\frac{d_1^2}{4\sigma^2})}{d_2^2 - d_1^2} \right]$$

Writing  $D = d^2$ , we find:

$$\Omega(D_{1,2}) = 1 - \frac{1}{\pi\rho} \left[ \frac{\exp(-\frac{D_2}{4\sigma^2}) - \exp(-\frac{D_1}{4\sigma^2})}{D_2 - D_1} \right]$$

$$\Omega(D) = \lim_{D_1 \rightarrow D_2 = D} \Omega(D_{1,2})$$

$$\lim_{D_1 \rightarrow D_2 = D} \frac{\exp(-\frac{D_2}{4\sigma^2}) - \exp(-\frac{D_1}{4\sigma^2})}{D_2 - D_1} = f'(D) \text{ with } f(x) = \exp(-\frac{x}{4\sigma^2})$$

$$f'(x) = -\frac{1}{4\sigma^2} \exp(-\frac{x}{4\sigma^2})$$

$$\text{So that } \Omega(D) = 1 + \frac{1}{4\pi\rho\sigma^2} \exp(-\frac{D}{4\sigma^2})$$

And finally:

$$\Omega(d) = 1 + \frac{1}{4\pi\rho\sigma^2} \exp(-\frac{d^2}{4\sigma^2}) \quad [\text{C.6}]$$

Combining Equation 2 in the manuscript with Equation C.3 and Equation C.6 lead to the derivation for the distance-decay relationship under the Poisson Cluster Process

(Equations 4 to 6 in the manuscript):

$$\chi(a, d) = \frac{\int (1 - \exp(-anc(A)))(1 - \exp(-anc(A)\Omega(\rho, \sigma, d)))\xi(n, \rho, \sigma) dnd\rho d\sigma}{\int (1 - \exp(-anc(A)))\xi(n, \rho, \sigma) dnd\rho d\sigma} \quad [\text{C.7}]$$

$$\text{with } c(A) = \frac{1}{\mu A} \int_X (1 - \exp(-\mu \int_A h(u-s) du) ds$$

$$\text{and } \Omega(\rho, \sigma, d) = 1 + \frac{1}{4\pi\rho\sigma^2} \exp(-\frac{d^2}{4\sigma^2})$$

### **C3. Relationship between Poisson Cluster Process coefficient of aggregation $c$ and negative binomial clustering parameter $k$**

The probability that a species with landscape-scale abundance  $n$  occurs in a sample which area constitutes a fraction  $a$  of the landscape is given by:

For the negative binomial distribution with parameter  $k$ :

$$\psi(a, n) = 1 - \left(1 + \frac{an}{k}\right)^{-k}$$

For the Poisson cluster process:

$$\psi(a, n) = 1 - \exp(-anc)$$

It follows:  $\left(1 + \frac{an}{k}\right)^{-k} = \exp(-anc)$

$$-k \ln\left(1 + \frac{an}{k}\right) = -anc$$

And finally:

$$c = \frac{k}{an} \ln\left(1 + \frac{an}{k}\right) \tag{C.8}$$

Equation C.8 provides the exact analytical link between  $c$  and  $k$ .

**APPENDIX D. Figures relevant to empirical evaluation of *Qualitative predictions in***

**BCI, Yasuni and Korup**

**Figure D1. Empirical evaluation in BCI**

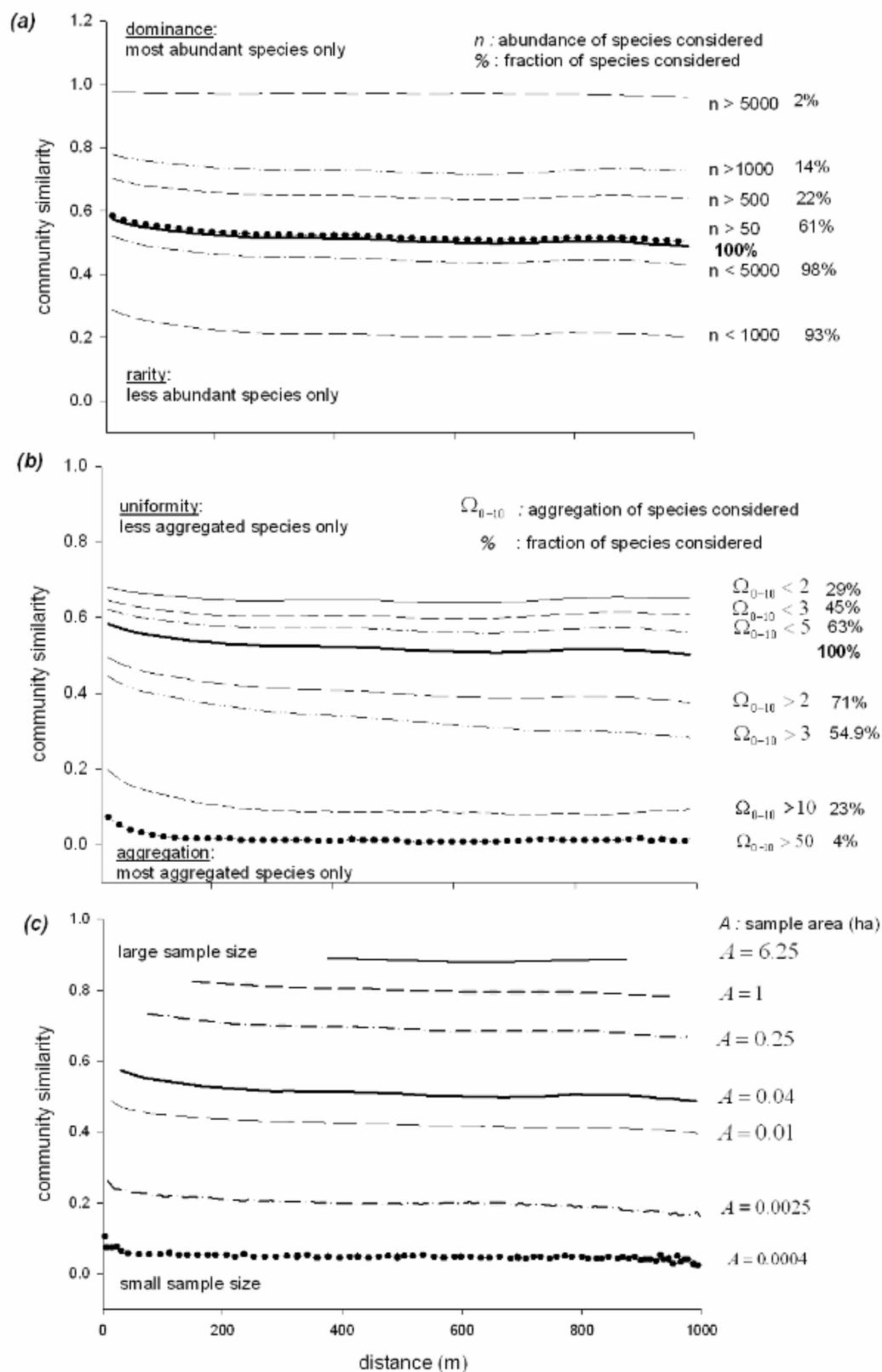
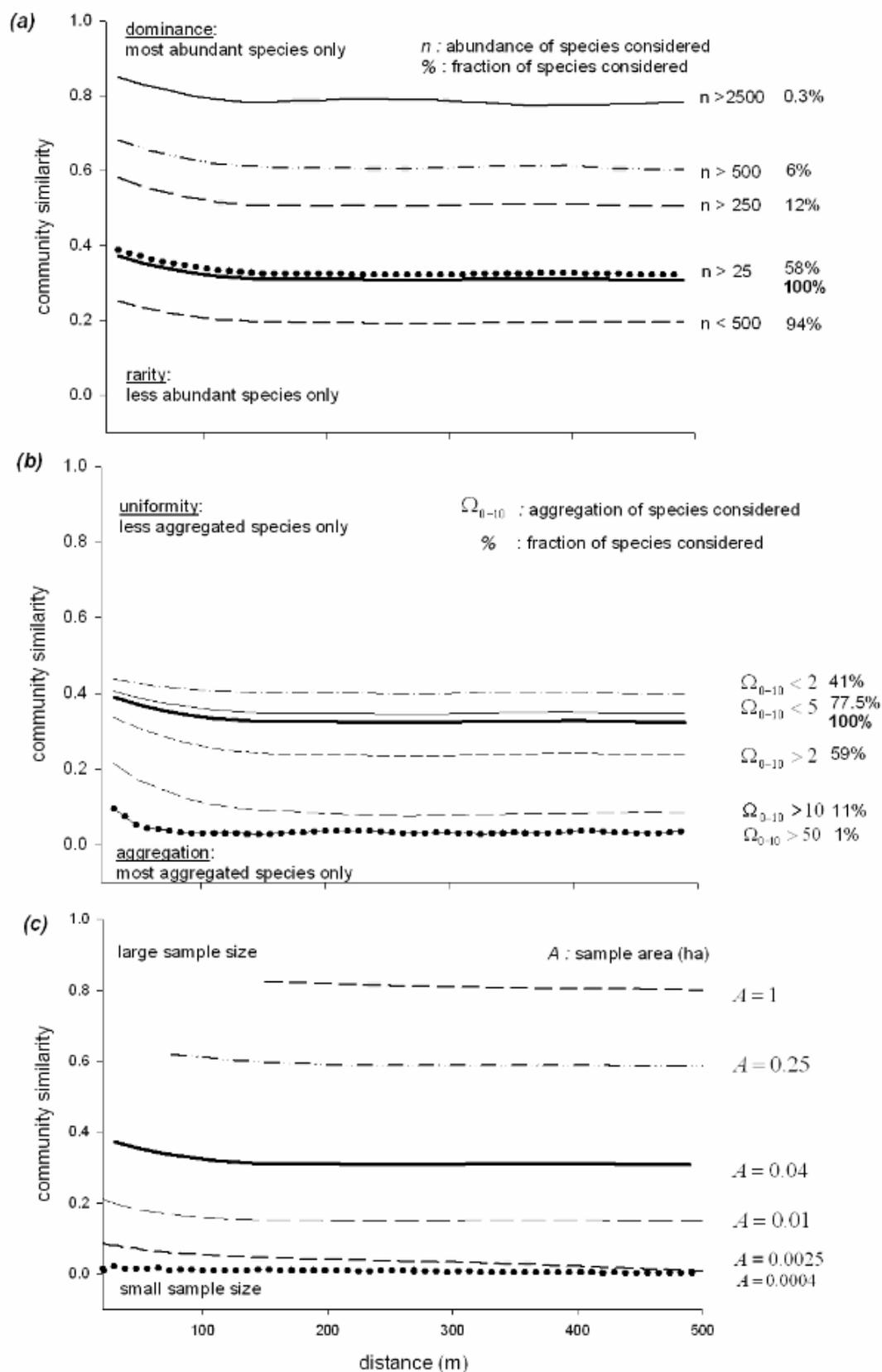
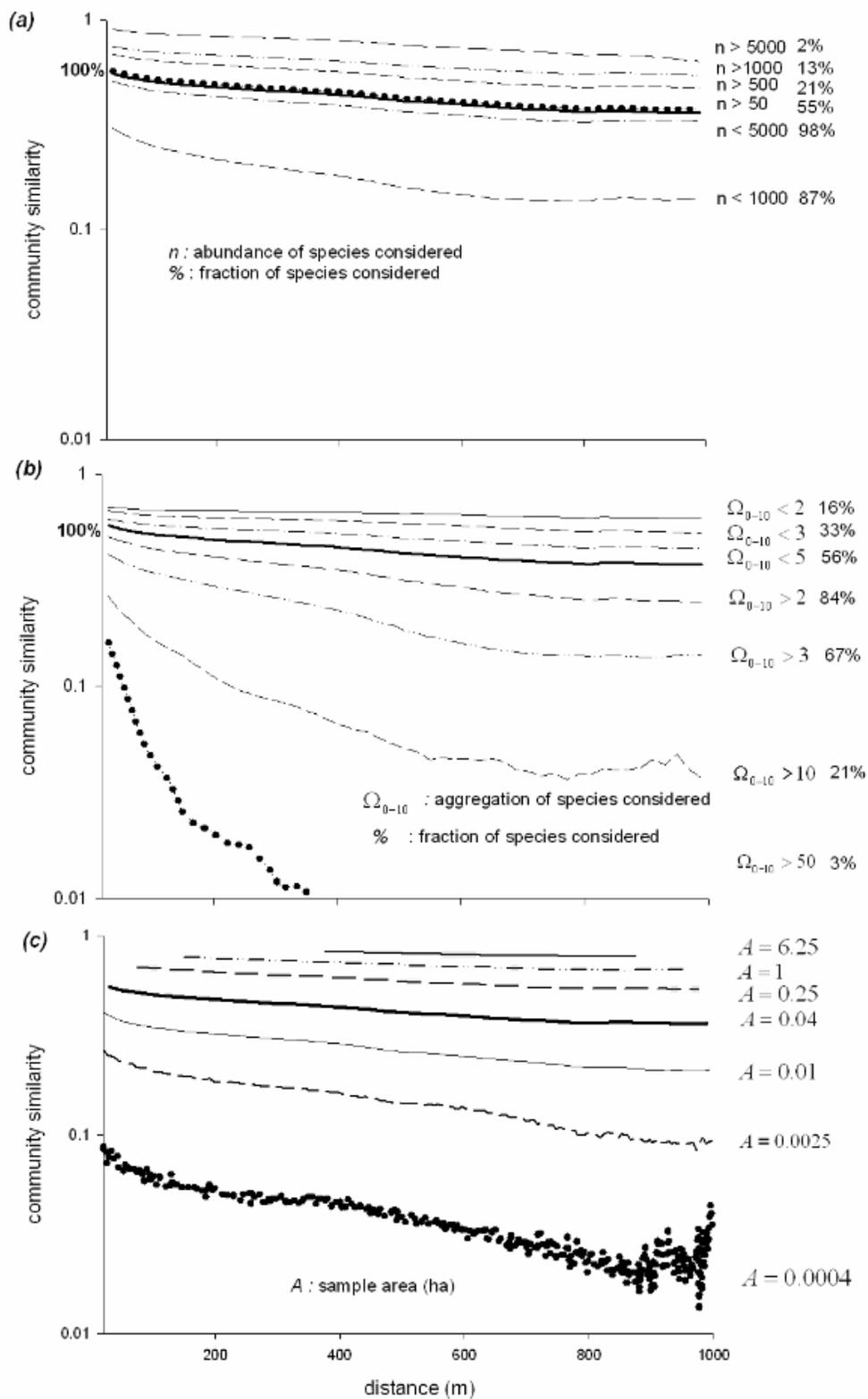


Figure D2. Empirical evaluation in Yasuni



**Figure D3. Linear-log plots emphasizing the importance of aggregation in shaping the distance-decay relationship in Korup**



## **APPENDIX E. Estimation and distribution of Poisson Cluster Process parameters**

All methods presented here are derived from Cressie (1993), Plotkin (2000) and Diggle (2003).

### **E1. Estimation of Poisson Cluster Process parameters**

For each species in a forest, we first estimated the parameters  $\rho$  and  $\sigma$  as described below, and then deduced the parameter  $\mu$  from the relationship  $n = \rho\mu A_0$ . For reasons explained below, we treated differently species with less or more than 20 individuals in the plot. Species with more than twenty individuals represented 63% of the pool of species in Yasuni, 72% in BCI, and 66% in Korup.

a) For species with more than 20 individuals in the plot, the parameters  $\rho$  and  $\sigma$  were estimated by best-fit of the observed Ripley's  $K$  curve  $K(d_1), \dots, K(d_{\max})$  to the expected  $K^{PCP}(d_1), \dots, K^{PCP}(d_{\max})$ , where  $K^{PCP}(d) = \pi d^2 + \frac{1}{\rho}(1 - \exp(-d^2 / 4\sigma^2))$ .

The best fit was found by minimizing the integral  $\int_0^{d_{\max}} (K(h)^c - (K(h)^{PCP})^c)^2 dh$  with  $c = 0.25$ , based on Plotkin *et al.* (2000). Following Plotkin *et al.* (2000),  $d_{\max}$  was chosen based on the observation of the curves describing  $P_d - P_{rand}$ , where  $P_d$  is the proportion of trees in the plot, distance  $d$  apart, which are the same species, and  $P_{rand}$  is the expected value of  $P_d$  under random placement. These curves suggested choosing  $200 \leq d_{\max} \leq 400$ . We performed all analyses (fitting the parameters of the Poisson Cluster Process and calculating the corresponding distance-decay relationships), for  $d_{\max} = 200$ ,  $d_{\max} = 300$ , and  $d_{\max} = 400$ . Choosing a different value for  $d_{\max}$  did not change the resulting distance-

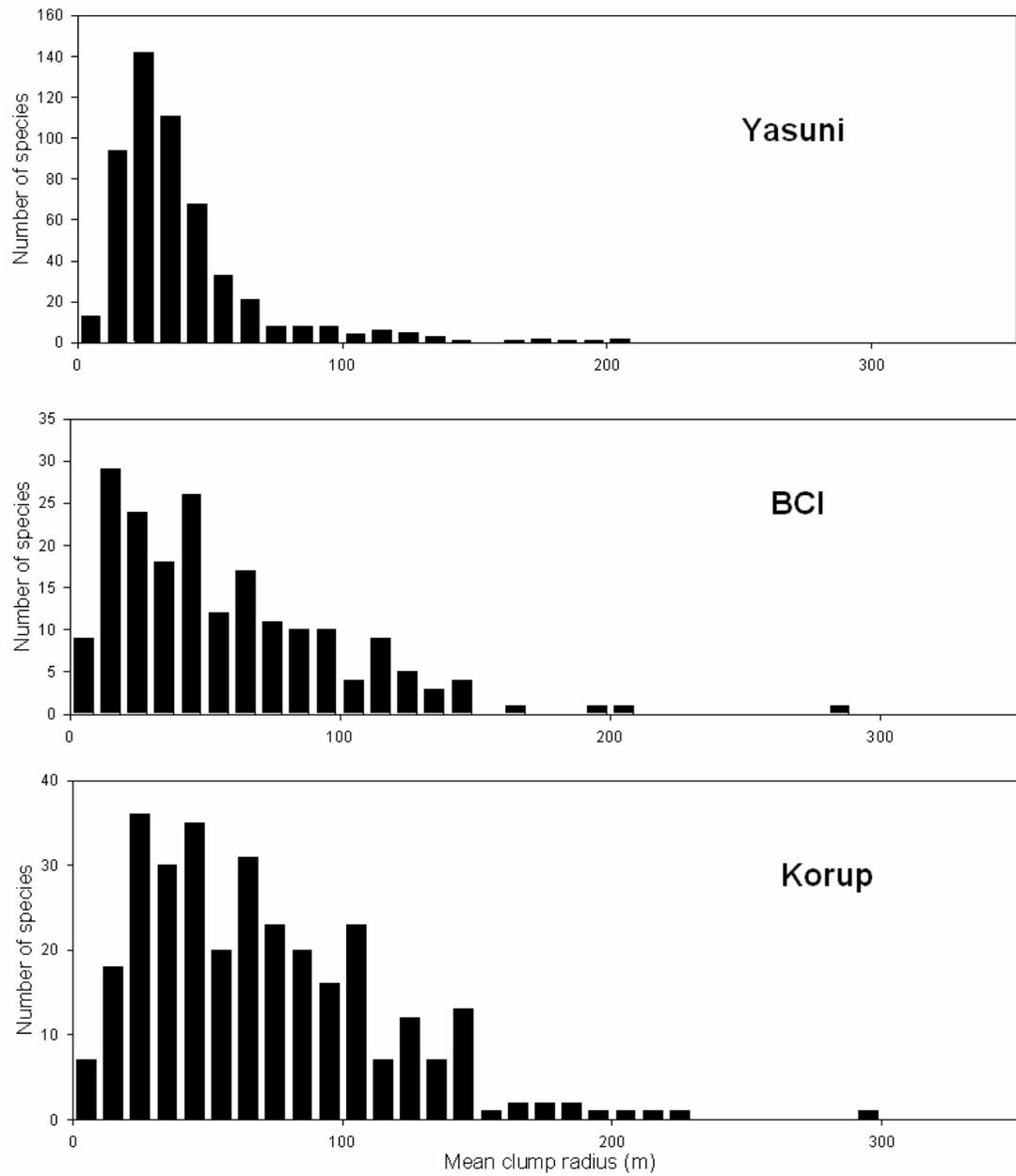
decay curves significantly, and we reported in the manuscript results obtained with the intermediate value  $d_{max} = 300$ . However, we do not exclude the possibility that the fit of the Poisson Cluster parameters with  $d_{max}$  chosen specifically species by species would improve the community level fit of the distance-decay.

b) When the spatial distribution of species is close to random, the value of the estimated parameters can be unrealistic. When we found a value for mean cluster diameter larger than the length of the plot, we set this value to the length of the plot. When we found more clusters than individuals in the plot, we set the number of clusters to the number of individuals in the plot. These arbitrary values were designed to reproduce the random distribution of the species and are not expected to influence our results. Reasonable parameter estimations were found for 74% of the species with more than 20 individuals in Yasuni, 91% in BCI and 95% in Korup. In Figures 4 and 5 in the manuscript, we consider these species only.

c) We consistently observed across forests that, for a high proportion of species with less than 20 individuals, fitting Ripley's  $K$  curve did not lead to a realistic estimation of the parameters. We therefore arbitrarily chose  $n = 20$  as a cut-off above which we did not use the statistical estimation technique described above. Instead, we estimated the parameters for rare species using those estimated in a) for species with more than 20 individuals and realistic parameter values. We assigned  $\rho$  and  $\sigma$  values to species with less than 20 individuals based on the log-log regression between  $\rho$  and  $n$  and the log-log regression between  $\sigma$  and  $n$  (Figure 5 in the manuscript). Log-log regressions were suggested by the

log-normal shape of the parameters distributions (Figures 4 and 6). This estimation for rare species is rough and concerns a high fraction of the species in the plot. However, the rare species are not expected to influence distance-decay relationships strongly (Figure 3 in the paper and Appendix D). We tested the possibility that the bad fit of the Poisson Cluster Process distance-decay relationship in Korup to the observed curve results from a bad estimation of the Poisson Cluster Process parameters for rare species. We performed in Korup distance-decay analyses similar to those presented in Figure 6 but excluding the rare species and found, consistently with Figure 6, that the curves obtained with the Poisson Cluster Process did not fit the observed curves. .

## E2. Distributions of mean clump radius on a linear-linear scale



## APPENDIX F. Supplementary information relevant to “Empirical evaluation”

### **F1. Comparing measures of aggregation**

#### Intuitive interpretation of $c$

$c(A)$  is a coefficient ranging between 0 and 1 that controls the deviation of the occurrence probability from the one expected under random placement ( $\psi(an) = 1 - e^{-an}$ ). Values of  $c(A)$  close to 0 correspond to high clustering. Values of  $c(A)$  close to 1 correspond to random placement. Intuitively, the number of individuals in a population should be multiplied by  $\frac{1}{c(A)}$  to obtain an occurrence probability corresponding to random placement.  $c(A)$  is scale dependent, and depends on  $\mu$  and  $\sigma$  only:

$$c(A) = \frac{1}{\mu A} \int_x (1 - \exp(-\mu \int_A h(u-s) du)) ds$$

For  $\sigma$  constant, aggregation increases (according to  $c$ ) if and only if the density of individuals per cluster  $\mu$  increases.

#### Intuitive interpretation of $\Omega$

$\Omega$  is a decreasing function of distance that reflects the deviation of the neighborhood occurrence probability from the occurrence probability. High values of  $\Omega$  at small distances and steep decays of  $\Omega$  correspond to high clustering. Under random placement,  $\Omega$  is equal to 1 at any distance. Intuitively,  $\Omega$  measures the effect of being close to a conspecific. It depends on  $\rho$  and  $\sigma$  only:

$$\Omega(\rho, \sigma, d) = 1 + \frac{1}{4\pi\rho\sigma^2} \exp\left(-\frac{d^2}{4\sigma^2}\right)$$

At constant  $\sigma$ , aggregation decreases (according to  $\Omega$ ) if and only if the density of clusters increases. Note that, despite the fact that  $\Omega$  is designed to investigate local densities around conspecifics, it is independent of the number of individuals per cluster ( $\mu$ ). Note also that  $\Omega$  values in adjacent distance classes are highly correlated (Condit *et al.* 2000), justifying measuring  $\Omega$  values in the arbitrary 0-10 m class. This choice is not expected to influence our results.

When abundance increases, both the number of clusters and the number of individuals per cluster increase (Figure 5 in the paper). The increase in number of clusters tends to reduce aggregation, while the increase in number of individuals per cluster tends to increase aggregation.  $\sigma$ ,  $\Omega$  and  $c$  emphasize different aspects of aggregation (size, number or density of clusters), each of which scale differently with abundance.

## **F2. Testing Equation 4 against simulations and the Poisson Cluster Process against data**

### Simulating the Poisson Cluster Process in each forest

Once the parameters  $\rho$  and  $\sigma$  are estimated for each species in a forest, the Poisson Cluster equivalent of the forest is obtained by overlaying each independently simulated species. To avoid edge effects, each species is simulated according to the Poisson Cluster in a larger area surrounding the actual plot, corresponding here to a 750 hectare plot surrounding the 50 hectare plot in Korup and BCI and the 25 hectare plot in Yasuni. We tested that increasing the area surrounding the plots did not influence the results.

## Testing the Poisson Cluster Process against data

### *Discrete equivalent of Equation 4*

To compute expected Poisson Cluster Process similarity values (Eq. 4) against data, we worked with the discrete equivalent of Equation 4, which reads:

$$\chi(a, d) = \frac{\sum_{i=1}^{S_0} (1 - \exp(-an_i c_i(A)))(1 - \exp(-an_i c(A)\Omega(\rho_i, \sigma_i, d)))}{\sum_{i=1}^{S_0} (1 - \exp(-an_i c_i(A)))} \quad [\text{F.1}]$$

with

$$c_i(A) = \frac{1}{\mu_i A} \int_X (1 - \exp(-\mu_i \int_A h_i(u-s) du)) ds$$

$$h_i(x, y) = \frac{1}{2\pi\sigma_i^2} \exp\left(-\frac{(x^2 + y^2)}{2\sigma_i^2}\right)$$

The discrete equivalent for the species-area relationship is given by the denominator of Equation F2.1

### *Expectation of Sørensen similarity under random placement*

Expectations under random placement were computed using Plotkin & Muller-Landau's (2002) sampling formula: under random placement,  $\psi^* = \psi$  at any distance (or equivalently  $\Omega(d) = 1$  at any distance), and  $\chi$  is equal to the average similarity across the landscape at any distance. The discrete equivalent of equation B.1 reads, for  $a = b$  :

$$\chi(a) = \frac{\sum_{i=1}^{S_0} \psi(an_i, \gamma_i)^2}{\sum_{i=1}^{S_0} \psi(an_i, \gamma_i)}$$

Under random placement:  $\psi(an_i, \gamma_i) = 1 - \exp(-an_i)$

and:

$$\chi(a) = \frac{\sum_{i=1}^{S_0} (1 - \exp(-an_i))^2}{\sum_{i=1}^{S_0} (1 - \exp(-an_i))} \quad [\text{F.2}]$$

*Species-area and distance-decay relationship 95% confidence intervals*

The 95% confidence intervals of the species-area and distance-decay relationships are obtained by Monte-Carlo simulations of Poisson Cluster Process communities. The 95% confidence envelop is composed of the 38 out of 40 intermediate values obtained by simulations at each area (species area relationship) or distance class (distance-decay relationship).

*Test of Poisson Cluster Process hypotheses and sensitivity analyses*

We test, for each forest, the null hypothesis H0 that population aggregation can be characterized by the Poisson Cluster Process, using both species-area and distance-decay curves. Below, we describe this test for the distance-decay relationship. The approach for species-area curves is similar, replacing community similarity by species richness, and distances by areas.

We use a test inspired from spatial statistics (Diggle 2003; Green *et al.* 2003): we measure the deviation of the observed similarity values from the theoretical expectations under the Poisson Cluster Process, using:

$$k_{observed} = \sum_i [\chi_{observed}(d_i) - \chi_{predicted}(d_i)]^2$$

where  $\chi(d_i)$  is the average similarity in distance-class  $i$ .

We also measure, for 40 simulated communities, the deviation of the similarity values from the theoretical expectations under the Poisson Cluster Process, using:

$$k_{simulated} = \sum_i [\chi_{simulated}(d_i) - \chi_{predicted}(d_i)]^2.$$

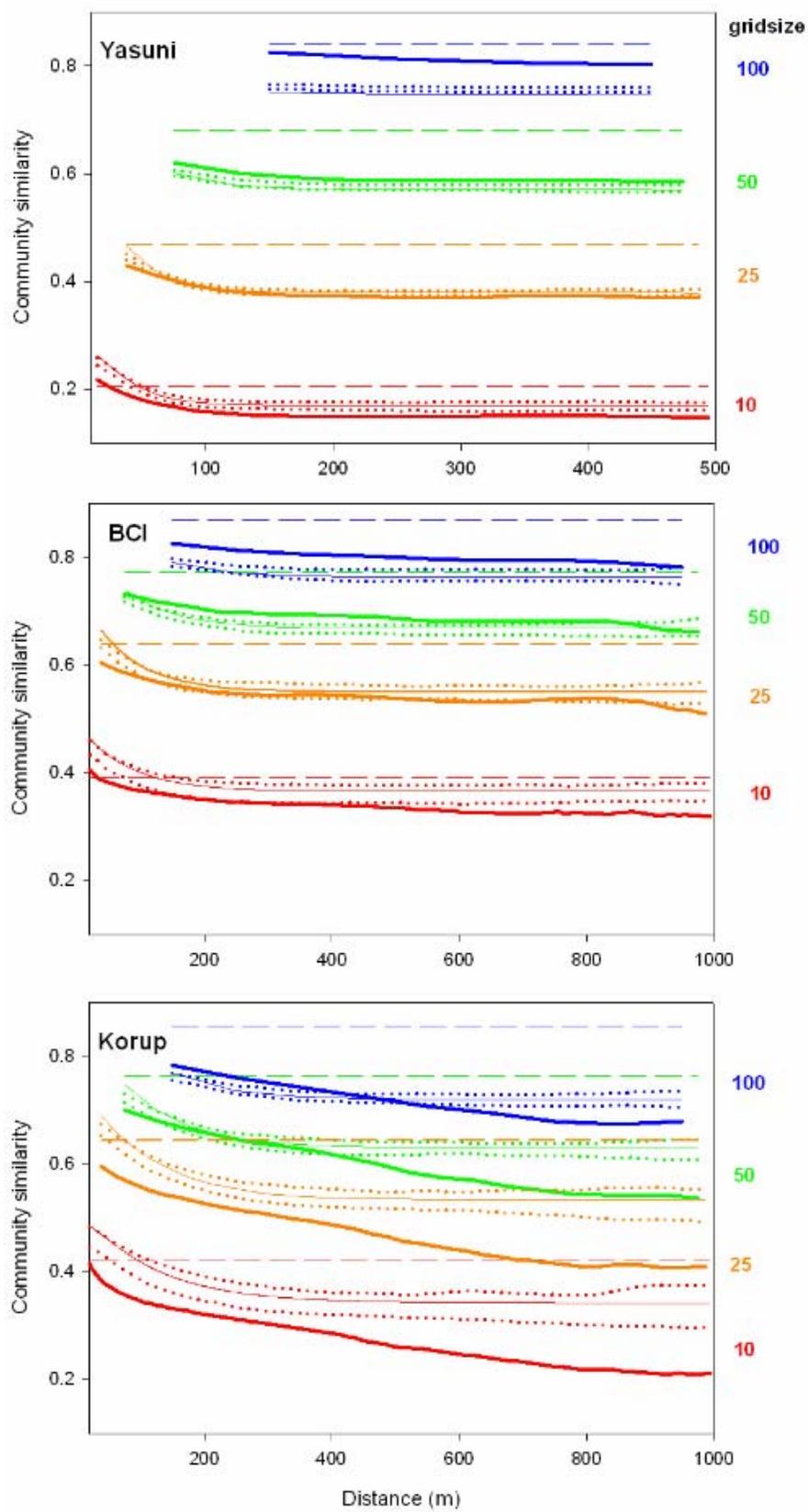
The null hypotheses H0 that a distance-decay curve arise from the Poisson Cluster Process is rejected at the 5% level confidence if  $k_{observed} > k_{simulated}$  for at least 38 out of the 40 simulations.

With these criteria, H0 is not rejected with the species-area relationship in any of the forests. It is rejected in all three forests with the distance-decay relationship, except in Yasuni and BCI for the 25 × 25 meter sample size.

**Figure F1. Comparison of theory with data in Yasuni, BCI, and Korup for different sample sizes**

Thick solid lines represent observed curves. Thin solid lines represent curves predicted by the Poisson Cluster Process (Eq. F.1 above). Thin dotted lines represent 95% confidence intervals produced by simulation of the Poisson Cluster Process. Dashed lines represent curves predicted from the species abundance distribution in each forest assuming random placement (Eq. F.2 above). The Poisson Cluster Process improves predictions under random placement in the three forests, and offers a good first approximation in Yasuni and BCI. The process tends to overestimate similarity values in the forests for small sample sizes, and to underestimate them for larger sample sizes. In Korup, the Poisson Cluster Process fails to reproduce the shape of observed distance-decay curves, and largely overestimate similarity values, particularly for the smallest sample sizes.

Figure F1.



**APPENDIX G. Supplementary Material References**

- Condit R., Ashton P.S., Baker P., Bunyavejchewin S., Gunatilleke S., Gunatilleke N., Hubbell S.P., Foster R.B., Itoh A., LaFrankie J.V., Seng L.H., Losos E., Manokaran N., Sukumar R. & Yamakura T. (2000) Spatial patterns in the distribution of tropical tree species. Science, 288, 1414-1418*
- Cressie N.A.C. (1993) Statistics for Spatial Data. Wiley & Sons, New-York.*
- Diggle P.J. (2003) Statistical Analyses of Spatial Point Patterns. 2nd edn. Edward Arnold, London.*
- Green J.L., Harte J. & Ostling A. (2003) Species richness, endemism, and abundance patterns: tests of two fractal models in a serpentine grassland. Ecology Letters, 6, 919-928*
- 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. J. Theor. Biol., 207, 81-99*