Meng Gao · Xinxiu Wang · De Wang

Species spatial distribution analysis using nearest neighbor methods: aggregation and self-similarity

Received: 20 August 2013 / Accepted: 27 January 2014 / Published online: 20 February 2014
© The Ecological Society of Japan 2014

Abstract: Spatial aggregation and self-similarity are two important properties in species spatial distribution analysis and modeling. The aggregation parameter k in the negative binomial distribution model and fractal dimension are two widely used measures of spatial aggregation and self-similarity, respectively. In this paper, we attempt to describe spatial aggregation and self-similarity using nearest neighbor methods. Specifically, nearest neighbor methods are used to calculate k and box-counting fractal dimension of species spatial distribution. First, five scaling patterns of k are identified for tree species in a tropical rainforest on Barro Colorado Island (BCI), Panama. Based on the scaling patterns and the means of the n th nearest neighbor distance (NND), the mean NND of higher ranks can be accurately predicted. Second, we describe how to use the theoretical probability distribution model of the n th NND for a homogeneous Poisson process on regular fractals to estimate the fractal dimension. The results indicate that the fractal dimensions estimated using the nearest neighbor method are consistent with those estimated using the scale–area method for 85 tree species on BCI (abundance ≥ 100 individuals and ≤ 5000 individuals). For other tree species, the breakdown of self-similarity in estimates of fractal dimension causes these two methods to be inconsistent. The applicability of the nearest neighbor method is also discussed.

Keywords Spatial aggregation · Scaling · Occurrence map · Fractal dimension · Barro Colorado Island, Panama

Introduction

Species spatial distribution, which is the result of different ecological processes and forces on multiple scales (Dungan et al. 2002; Wiegand and Moloney 2004), is the manner in which species are spatially arranged (Gaston 1994; Fortin and Dale 2005). Species spatial distribution analysis is crucial to understanding the relationship between ecological patterns and processes (Tilman and Kareiva 1997; Perry et al. 2002; Wiegand and Moloney 2004). Species distributions are commonly measured by the number of sites or geographic grid squares that the species occupy (Gaston and Blackburn 2000; Hui and McGee 2007a), and the total area of the occupied grids is defined as the range size (Harte et al. 2001). The relationship between species spatial range and abundance is an important research theme in ecology (Harte et al. 2001).

There are generally two categories of models that describe the link between species spatial range and abundance (Hui et al. 2009). The first category is occupancy–abundance relationship (OAR) models (Hui et al. 2009). The simplest OAR model is based on the Poisson model that corresponds to a random spatial distribution (He and Gaston 2000). For random (Poisson) distribution patterns, a remarkable characteristic is the power law relationship between the mean of the n th nearest neighbor distance (NND) and the rank of the NND (Thompson 1956). Hubbell et al. (2008) extended this power law to nonrandom spatial distribution patterns of tree species and estimated the range size based on the inverse relationship between species abundance and the power exponent; however, spatial aggregation of tree species was not included.

In our previous paper, an explicit expression of the mean of the n th NND (r_n) was derived, in which the

Electronic supplementary material The online version of this article (doi:10.1007/s11284-014-1131-8) contains supplementary material, which is available to authorized users.

M. Gao (✉) · X. Wang · D. Wang
Yantai Institute of Coastal Zone Research, Chinese Academy of Sciences, Yantai 264003, Shandong, China
E-mail: gaomeng03@hotmail.com
Tel.: +86-535-2109134
Fax: +86-535-2109000

X. Wang
Graduate University of Chinese Academy of Sciences,
Beijing 100049, China

aggregation parameter k was added (Gao 2013). In this paper, the NND specifically refers to the distance from a focal individual to a conspecific neighbor (event-to-event NND in Gao 2013). The expectation $E[r_n]$ can be computed based on the probability distribution model of the n th NND, and k is the same as the aggregation parameter in the negative binomial distribution (NBD) model. The parameter k has been verified to be a scale-variant measure of spatial aggregation (Hui et al. 2010; Zillio and He 2010). Theoretically, the k in the NBD model increases proportionally with the spatial scale (quadrat size) (Conlisk et al. 2007; He and Gaston 2007); however, much empirical evidence has shown that, in practice, the theoretical proportional dependence of k across scales is unlikely to hold (He and Gaston 2007). Therefore, exploring the scaling patterns of k is not only useful to accurately estimate $E[r_n]$ and species range size, but is also helpful to reveal the relationship between spatial distribution and abundance. In our previous paper, we verified that estimating k using the NND method was more robust than using the quadrat sampling method (Gao 2013).

In this paper, we will first study the relationship between k and the spatial scale using the nearest neighbor method (here the spatial scale is the rank of NND), and then identify and verify the scaling patterns of k . These tasks will be implemented using a data set of tree species in a tropical rain forest on Barro Colorado Island (BCI), Panama. Examining this relationship is the first objective of this paper.

The second category of models that link spatial range and abundance are scaling pattern of occupancy (SPO) models (Hui et al. 2006, 2009; Hui 2009). Kunin's (1998) log-log linear model (LLM), which describes a power law between species occupancy and grain size, is one SPO model. This model assumes a self-similar (fractal) structure of species spatial distribution, with a box-counting fractal dimension representing the scale-independence of species spatial distribution (Kunin 1998). Harte et al. (2001) also showed that species with self-similar distributions could yield a log-log linear relationship between occupancy and grain size. However, the self-similar structure of a species distribution usually breaks down at fine scales, and self-similarity only holds for a particular species over some spatial scales (Hui and McGeoch 2007b). Although there is debate about the assumption of self-similarity in species spatial distributions, fractals remain the most readily applied approach to quantifying the scaling properties of species (Harte et al. 1999, 2001; Ostling et al. 2000; Hui and McGeoch 2007a, 2008).

Fractal dimension is an important index of self-similar structure in species spatial analysis (Kunin 1998; Halley et al. 2004; Fortin and Dale 2005), and it also measures the degree of occupancy in space (Mladenoff et al. 1993, Hui et al. 2010). In mathematics, the calculation of box-counting fractal dimension is done using balls or squares to cover these points (individuals of species) and find the scaling rule (Falconer 1990).

Kunin's (1998) scale-area method is a practical way to estimate the box-counting fractal dimension. Hui et al. (2010) verified that the box-counting fractal dimension was also a scale-variant measure of species spatial aggregation.

In this paper, we also use the nearest neighbor method to estimate the fractal dimension of species spatial distribution and determine the conditions under which fractal dimensions can be computed using this method in practice. This second objective of this paper was also accomplished using the data set of tree species on BCI.

Models and methods

Models

In species spatial distribution analysis, the locations of individuals are usually mapped as points in space, resulting in a spatial point pattern (Legendre and Fortin 1989). In spatial statistics, a spatial point pattern is considered a realization of a spatial point process that generates countable events in the plane (Gatrell et al. 1996; Perry et al. 2006). For quadrat sampling, the Poisson and NBD models are the most widely used probability distribution models for random and aggregated spatial distribution patterns, respectively (Pielou 1960; He and Gaston 2000).

For distance sampling, NND is easier to understand and measure in situ (Stoyan and Penttinen 2000); therefore, the nearest neighbor method is widely used in quadrat sampling. The probability distribution model of distance from a focal individual to its n th conspecific neighbor was derived for aggregated spatial distributions (Gao 2013) as

$$f_n(r) = \frac{2(\lambda\pi)^n r^{2n-1}}{k^n} \frac{\Gamma(n+k+1)}{\Gamma(n)\Gamma(k+1)} \left(1 + \frac{\lambda\pi r^2}{k}\right)^{-n-k-1} \quad (1)$$

where r is the distance, n (≥ 1) is the rank of the NND, λ is the density of individuals, k (> 0) is the aggregation parameter, and $\Gamma()$ is the gamma function. With this model, the expectation of the n th NND can be simply computed as

$$E[r_n] = \frac{1}{\sqrt{\lambda\pi}} \frac{\Gamma(n+0.5)}{\Gamma(n)} \frac{\Gamma(k+0.5)\sqrt{k}}{\Gamma(k+1)} \quad (2)$$

To our knowledge, few theoretical methods can link fractals and NND except Sakhr and Nieminen's (2006) probability distribution model of the n th NND for a homogeneous Poisson process on regular fractals. The dimensionless form of Sakhr and Nieminen's (SN) model is

$$\tilde{f}_n(S; D_b) = \frac{D_b \alpha^n}{\Gamma(n)} S^{nD_b-1} e^{-\alpha S^{D_b}} \quad (3)$$

where S is the normalized distance ($S = r/\bar{r}_n$, and $\bar{r}_n$ is the mean of r_n), n (≥ 1) is the rank of the NND, D_b is

the box-counting dimension, and $\alpha = [\Gamma(n + 1/D_b)/\Gamma(n)]^{D_b}$. Because the derivation of SN model is complicated, it cannot be easily extended to nonrandom point process on fractals. From a spatial ecology perspective, the SN model can be interpreted as follows: if individuals are randomly distributed in a fractal landscape, the probability distribution of the n th NND can be modeled by the SN model. The relationships among $f_n(r)$, $\tilde{f}_n(S; D_b)$, and other probability distribution models for species spatial distribution analysis are clearly illustrated in Fig. 1.

Data

The data set of tree species on BCI is one of the most robust, long-term ecology data sets. In 1981, a 1000×500 m rectangular plot was established. In this plot, all freestanding trees and shrubs ≥ 1 cm diameter at breast height have been enumerated, individually located on a reference map, and identified to species (He and Hubbell 2003). This study used data from the 2005 census, which included 368,124 plants belonging to 301 species.

Methods

First, the relationship between k and the spatial scale using NND method is studied. The parameter k is estimated using maximum likelihood estimation based on the probability model $f_n(r)$ in Eq. (1) for the first 15 NNDs ($n = 1, \dots, 15$). Then, the scaling patterns of k are categorized according to the $k(n) \sim n$ curves. To

verify these scaling patterns, a simple method is to use curve-fitting to predict the $k(n)$ of higher ranks ($n > 15$) based on the scaling patterns, and to compare the predicted and estimated values. However, for BCI tree species, we have an alternative way to verify the scaling patterns. Hubbell et al. (2008) found that for many of these species, the relationship between the mean of NND and the rank n followed a power law. Here, we denote $M[r_n]$ as the average of r_n . Then, the verification of the scaling patterns of $k(n)$ vs. n includes the following four steps:

1. Compute $M[r_n] = \sum r_n^{(i)}/N$ using the empirical data of the n th NND and estimate $k(n)$ for $n = 1, 2, \dots, 15$;
2. classify the scaling patterns of $k(n)$ vs. n , and use curve-fitting to predict $k(n)$ of higher ranks ($n = 16, 17, \dots, 20$), then substitute the predicted $k(n)$ into $E[r_n]$ ($n = 16, 17, \dots, 20$);
3. do a linear regression analysis of $\log(M[r_n])$ vs. $\log(n)$ for $n = 1, 2, \dots, 15$ and predict the mean of the n th NND (denoted as $\hat{M}[r_n]$, $n = 16, 17, \dots, 20$) (this step is referred to as Hubbell et al.'s method in this paper);
4. compare $E[r_n]$ and $\hat{M}[r_n]$ with $M[r_n]$ for $n = 16, 17, \dots, 20$, respectively, and use the relative root mean square error to evaluate the two estimation methods.

If $E[r_n]$ is closer to $M[r_n]$ than $\hat{M}[r_n]$, the nearest neighbor method performs better than Hubbell et al.'s method, meaning that the categorization of scaling patterns of k is satisfactory.

Fig. 1 Probability distribution models used for species spatial distribution analysis (*top row*). Discrete frequency distribution models of quadrat count data in a regular two-dimensional study area. The Poisson and negative binomial distribution models correspond to random and aggregated spatial distributions, respectively (*middle row*). Probability distribution models of nearest neighbor distance in a regular spatial distribution pattern, while $g_n(r)$

and $f_n(r)$ correspond to an aggregated one. r is the distance, and n is the rank of nearest neighbor distance (NND). The probability distribution models of spatial aggregation approach spatial randomness when the aggregation parameter k tends to infinity (*bottom row*). Probability distribution models of the nearest neighbor distance on a fractal landscape (occurrence maps). Randomness and self-similarity are two basic conditions of $\tilde{f}_n(S; D_b)$

Second, we use the scale–area method and SN model to estimate the box-counting fractal dimension of tree species. To obtain the scale–area curve, a process of landscape subdivision is used (Harte et al. 1999; Hui and McGeoch 2007a). If the spatial structure of tree species is fractal, the scale–area curve in a log–log plot is linear and the slope (s) can be simply estimated via linear regression. Then the box-counting dimension can be computed as $D_b = 2 \times (1 - s)$. Meanwhile, with the SN model, the box-counting dimension D_b can be estimated using maximum likelihood. Although the maximum likelihood estimate of D_b cannot be derived analytically, an approximation can be obtained numerically. The rank of NND is also from 1 to 15. To evaluate the consistency of estimated fractal dimensions of tree species spatial distributions using the two methods, correlation coefficients are calculated.

Results

First, we select 113 tree species (with more than 50 individuals and DBH > 1 cm) and estimate the aggregation parameters ($n = 1, 2, \dots, 15$). We find that the relationship $k(n) \sim n$ can be classified into five categories: relatively stable, linearly increasing, gradually increasing, exponentially increasing, and exponentially decreasing (Tables 1, 2; Fig. 2). The first category corresponds to more or less scale-invariant spatial aggregation, and included only 14 species. The second, third, and fourth categories correspond to an increased aggregation parameter, meaning that the level of spatial aggregation decreases when sampling scale is increased. Most tree species belong to these three categories. The fifth type corresponds to a decreased aggregation parameter; however, only two species, *Bactris major* (Arecaceae) and *Inga peizifera* (Fabaceae), belong to this category (Table 1). The spatial distributions of these two tree species shared some features: a few highly-aggregated clusters and a few widely-distributed individuals (data not shown).

Table 1 summarizes the comparisons between the nearest neighbor method and Hubbell et al.'s method for the 113 BCI tree species. Obviously, for most tree species $E[r_n]$ is closer to $M[r_n]$ than $\hat{M}[r_n]$ ($n = 16, 17, \dots, 20$). In other words, the nearest neighbor method based on

$f_n(r)$ performs better than Hubbell et al.'s method in predicting the mean of NND of higher ranks. Moreover, this result verifies that the five scaling patterns of $k(n)$ vs. n summarized in this paper are reliable. In Fig. 2 and Table 2, five tree species are selected to illustrate the five types of scaling patterns of k .

When the scale–area method is used to estimate the box-counting dimension, we must first construct the scale–area curves. If self-similarity holds over all spatial scales, the scale–area curve of a tree species in a log–log plot will be almost linear. However, we find that self-similarity also breaks down (Supplementary materials). For some common tree species whose individuals are widely distributed in the regular two-dimensional study area, the slope of the scale–area curve in the log–log plot will be underestimated, and then the box-counting dimension will be overestimated. On the contrary, for some rare tree species whose individuals are not widely distributed but aggregated, the slope of scale–area curve in log–log plot will be overestimated, and then the box-counting dimension will be underestimated. In this paper, 85 species (with more than 100 individuals and less than 5000 individuals) are selected to illustrate the performance of the two methods in estimating the fractal dimension.

Fractal landscape and randomness are two basic conditions of the SN model. In the process of constructing scale–area curves, we see that the spatial structure of tree species is similar to that in fine-scale occurrence maps. Moreover, we find that the spatial distributions of tree species in occurrence maps are random, although they are considered to be aggregated in the regular two-dimensional study area (Supplementary materials). Next, we substitute the estimated box-counting dimensions using the scale–area method into Eq. (3) and compare the predicted probability distributions ($\hat{f}_n(S; D_b)$) with the empirical frequency distributions of the n th NND. For $n \leq 3$ and $n > 10$, the predicted and empirical probability density functions are usually not consistent. But for $4 \leq n \leq 10$, the predicted probability models fit the empirical frequency distribution well. In Fig. 3, we randomly select six tree species and illustrate the predicted and empirical probability distributions of the n th NND for $n = 6$. The fractal dimensions of all 85 tree species on BCI estimated using the scale–area and nearest neighbor methods

Table 1 Five scaling patterns of $k(n)$ vs. n and their curve-fitting formulae

Type	Scaling pattern	Curve-fitting formula ^a	Total ^b	Improved ^c
1	Relative stable	$k(n) = b$	14	11
2	Linear increasing	$k(n) = a \times n + b$	51	45
3	Gradually increasing	$k(n) = a \times \log(n) + b$	28	25
4	Exponential increasing	$k(n) = a \times e^{b \times n} + c$	18	16
5	Exponential decreasing	$k(n) = a \times e^{b \times n} + c$	2	2

^a $k(n)$ is the aggregation parameter, n is the rank of nearest neighbor distance (NND), r_n is the n th NND, a , b and c are coefficients in the five curve-fitting formulae

^b The number of species belonging to each category of scaling pattern

^c The number of species with an improved estimation accuracy using $E[r_n]$ compared with Hubbell et al.'s method

Table 2 Results of linear regression and curve fitting for the five tree species presented in Fig. 2

Type	Species	Hubbell et al.'s method	Nearest neighbor method in this paper				
		$\text{Log}(M[r_n])$ vs. $\log(n)$	R^2	RRMSE	$k(n)$ vs. n	R^2	RRMSE
1	<i>Licania hypoleuca</i>	$\text{Log}(M[r_n]) = 0.5354 \times \log(n) + 3.106$	0.99	0.0499	$k(n) = 0.3146$	–	0.0174
2	<i>Eugenia galadonensis</i>	$\text{Log}(M[r_n]) = 0.566 \times \log(n) + 1.9598$	0.99	0.0159	$k(n) = 0.0425 \times n + 0.6309$	0.97	0.0029
3	<i>Miconia affinis</i>	$\text{Log}(M[r_n]) = 0.6388 \times \log(n) + 2.4375$	0.99	0.1371	$k(n) = 0.1582 \times \log(n) + 0.6309$	0.97	0.0204
4	<i>Psychotria graciliflora</i>	$\text{Log}(M[r_n]) = 0.7541 \times \log(n) + 1.7567$	0.98	0.3875	$k(n) = 0.005221 \times e^{0.0945 \times n} + 0.0002616$	0.94	0.0052
5	<i>Inga peizifera</i>	$\text{Log}(M[r_n]) = 0.4638 \times \log(n) + 3.0586$	0.99	0.054	$k(n) = 0.1098 \times e^{-0.3751 \times n} + 0.1038$	0.96	0.0292

$k(n)$ is the aggregation parameter, n is the rank of nearest neighbor distance (NND), r_n is the n th NND, $M[r_n]$ is the average of r_n . Regression and curve fitting use the first 15th nearest neighbor distances ($n = 1, \dots, 15$), and the relative root mean square error (RRMSE) is calculated for $n = 16, \dots, 20$

($n = 6$ and 10) are shown in Fig. 4. We observed that the estimates of fractal dimensions using the two methods are consistent (Pearson's $r = 0.8231$ for $n = 6$, and Pearson's $r = 0.8004$ for $n = 10$). For other values of rank n ($4 \leq n \leq 10$), similar results can be found (data not shown).

Discussion

In this study, we focused on two important characteristics of species spatial distributions: aggregation and self-similarity. However, we did not employ the commonly used methods based on quadrat sampling to detect spatial aggregation and self-similarity. Instead, we explored distance sampling, specifically, nearest neighbor methods. In spatial analysis, the nearest neighbor method was first applied to measure spatial aggregation (Clark and Evans 1954). Diggle (1983) also used the probability distribution of the NND to test spatial randomness. Nearest neighbor methods are very useful in spatial analysis and were once referred to as distance sampling methods (Buckland et al. 1993; Perry et al. 2006).

In our previous paper, the probability distribution model of the n th NND was presented (Gao 2013). With that model, we can estimate k in the NBD model, which is a widely-used probability distribution model for quadrat count data sampled from aggregated spatial distributions. Both theoretical and empirical research verified that spatial aggregation was scale-dependent (Conlisk et al. 2007; He and Gaston 2007; Hui et al. 2010). For BCI tree species, the relationship between k and spatial scale (quadrat size) was simply assumed to be a power law (Plotkin and Muller-Landau 2002; He and Hubbell 2003). In this paper, we studied the scaling patterns of k using the nearest neighbor method and identified five scaling patterns: relatively stable, linearly increasing, gradually increasing, exponentially increasing, and exponentially decreasing. We found that k did not always increase with spatial scale. The scaling patterns of k estimated using the nearest neighbor method were more diverse than those estimated using quadrat sampling in previous studies (Plotkin and Muller-Landau 2002; He and Hubbell 2003; Conlisk et al. 2007). Therefore, spatial aggregation should be treated carefully in studies of the relationship between abundance and occupancy (or range).

To study the relationship between abundance and range sizes of tree species at the community level, Hubbell et al. (2008) used an empirical result that was interpreted as a power-law relationship between the mean of the n th NND and the rank n . However, this power law is not universal for all individuals or species on BCI. In this paper, we also studied the relationship between the mean of the n th NND and the rank n , where the spatial aggregation was explicitly included, and verified the five scaling patterns of k . The results indicated that the nearest neighbor method performed better

Fig. 2 Scaling patterns and nearest neighbor distances (NNDs) (left column). Scaling patterns of $k(n)$ vs. n . The first 15 values (dots) are directly estimated from NNDs based on the probability distribution model $f_n(r)$, while the next five values (circles) are calculated using curve-fitting. r is the distance, and n is the rank of nearest neighbor distance (NND) (right column). Actual and predicted means of the n th NND. Solid dots represent the actual

values, $M[r_n]$; circles represent the predicted values, $\hat{M}[r_n]$, estimated using Hubbell et al.'s method and the solid lines are the fitted lines; crosses represent $E[r_n]$. The scaling patterns from top to bottom are: relatively stable, linearly increasing, gradually increasing, exponentially increasing, and exponentially decreasing. Example species are indicated

than Hubbell et al.'s method in predicting the mean of the NND of higher ranks (Table 1). This result also verified that the five scaling patterns of k identified in this study were satisfactory and reliable.

As a second category of methods to study the relationship between species abundance and range sizes, SPO models such as Kunin's LLM exploit the scale dependence of species range size by assuming a fractal (self-similar) landscape structure (Hui and McGeoch 2007a). The self-similar structure of the species distribution corresponds to a power-law form of a scale-area curve. However, self-similarity of a species distribution

usually breaks down at fine scales. In this study, the breakdown of self-similarity was also observed for tree species on BCI. For some common species with total abundance ≥ 5000 individuals (usually widely distributed), the box-counting dimensions are usually overestimated by the scale-area method. In contrast, for some rare species, the box-counting dimensions are usually underestimated. Thus, the fractal dimension cannot be estimated directly using the scale-area method.

In this paper, we used a nearest neighbor method to estimate fractal dimension of species spatial distribution. For many tree species on BCI (abundance ≥ 100 and $\leq$

Fig. 3 Empirical (histogram) and fitted (solid lines) probability distributions of the normalized n th nearest neighbor distance for six species. The fitted probability distribution model is estimated using $\tilde{f}_n(S, D_b)$, where S is the normalized distance and D_b is the fractal dimension. The Kolmogorov–Smirnov test was used to determine the goodness of fit, and the p values of the six species are 0.32, 0.45, 0.26, 0.16, 0.52, and 0.074, respectively

Fig. 4 Fractal dimensions estimated using the scale–area and nearest neighbor methods for 85 tree species (with > 100 individuals and < 5000 individuals) on Barro Colorado Island. **a** $n = 6$, Pearson's $r = 0.8231$; **b** $n = 10$, Pearson's $r = 0.8004$

5000 individuals), the estimates of fractal dimension using these two methods are consistent. The nearest neighbor method is based on the model $\tilde{f}_n(S; D_b)$, which is the probability distribution model of the n th NND of randomly distributed points on a fractal landscape. In

other words, the fractal dimension estimated using the nearest neighbor method is actually that of the fractal landscape. We also detected spatial randomness in occurrence maps (occupied cells), although the spatial distribution patterns of these tree species are aggregated

in the regular two-dimensional study area. As we know, species distributions result from different ecological processes and forces on multiple scales (Dungan et al. 2002; Wiegand and Moloney 2004; Fortin and Dale 2005). If we can identify the fractal landscape and determine the environmental differences between the fractal landscape and other areas, we could reveal the roles of environmental heterogeneity and ecological processes in shaping the spatial structure of plant species in nature. However, we are not sure whether the occurrence map is the fractal landscape, and further research is needed.

In practice, when the probability distribution model $\tilde{f}_n(S; D_b)$ is used, we have many choices of rank n . Although whether there is one universal and optimal value of n has not been verified, we conclude based on this study that the rank n should be neither too large nor too small. For tree species on BCI, the rank n is generally from 4 to 10. The closer the individuals are, the more strongly they interact. Therefore, the real distances from a focal individual to its nearest few ($n \leq 3$) neighbors are usually larger than the theoretical values, causing the empirical frequency distribution of the n th NND to be right-skewed. In contrast, a large rank n will reduce the sampling dependence and increase the burden of field surveys. Moreover, because the spatial structure of species distributions is not perfectly self-similar (Halley et al. 2004) and fractal dimension is also a scale-dependent measure in this context (Hui et al. 2010), we must be cautious using fractal properties in modeling species spatial distributions.

Acknowledgments This work was supported by the National Natural Science Foundation of China (Nos. 31000197, 41001360) and the Knowledge Innovation Project of the China Academy of Sciences (No. KZCX2-EW-QN209). The authors are grateful to the Center for Tropical Forest Science for providing the BCI data set and to Dr. Guochun Shen for preparing the data. Help from the editors and all anonymous reviewers of the original manuscript is also acknowledged.

References

- Buckland ST, Anderson DR, Burnham KP, Laake JL (1993) Distance sampling: estimating abundance of biological populations. Chapman and Hall, London
- Clark PJ, Evans FC (1954) Distance to nearest neighbor as a measure of spatial relationships in populations. *Ecology* 35:445–453
- Conlisk E, Conlisk J, Harte J (2007) The impossibility of estimating a negative binomial clustering parameter from presence absence data: a comment on He and Gaston. *Am Nat* 170:651–654
- Diggle PJ (1983) Statistical analysis of spatial point patterns. Academic Press Inc., New York
- Dungan JL, Perry JN, Dale MRT, Legendre P, Citron-Pousty S, Fortin MJ, Jakomulska A, Miriti M, Rosenberg MS (2002) A balanced view of scale in spatial statistical analysis. *Ecography* 25:626–640
- Falconer K (1990) Fractal geometry: mathematical foundations and applications. Wiley, Chichester
- Fortin MJ, Dale M (2005) Spatial analysis: a guide for ecologists. Cambridge University Press, Cambridge
- Gao M (2013) Detecting spatial aggregation from distance sampling: a probability distribution model of nearest neighbor distance. *Ecol Res* 28:397–405
- Gaston KJ (1994) Rarity. Chapman and Hall, London
- Gaston KJ, Blackburn TM (2000) Patterns and processes in macroecology. Blackwell Science, Oxford
- Gatrell AC, Bailey TC, Diggle PJ, Rowlingson BS (1996) Spatial point pattern analysis and its application in geographical epidemiology. *Trans Inst Brit Geogr* 21:256–274
- Halley JM, Hartley S, Kallimanis AS, Kunin WE, Lennon JJ, Sgardelis SP (2004) Uses and abuses of fractal methodology in ecology. *Ecol Lett* 7:254
- Harte J, Kinzig A, Green JL (1999) Self-similarity in the distribution and abundance of species. *Science* 284:334–336
- Harte J, Blackburn T, Ostling A (2001) Self-similarity and the relationship between abundance and range size. *Am Nat* 157:374–386
- He F, Gaston KJ (2000) Estimating species abundance from occurrence. *Am Nat* 156:553–559
- He F, Gaston KJ (2007) Estimating abundance from occurrence: an underdetermined problem. *Am Nat* 170:655–659
- He F, Hubbell SP (2003) Percolation theory for the distribution and abundance of species. *Phys Rev Lett* 91:198103
- Hubbell SP, He F, Condit R, Borda-de-Água L, Kellner J, ter Steege H (2008) How many tree species are there in the Amazon and how many of them go extinct? *Proc Natl Acad Sci USA* 105:11498–11504
- Hui C (2009) On the scaling patterns of species spatial distribution and association. *J Theor Biol* 261:481–487
- Hui C, McGeoch MA (2007a) A self-similarity model for the occupancy frequency distribution. *Theor Popul Biol* 71:61–70
- Hui C, McGeoch MA (2007b) Modeling species distributions by breaking the assumption of self-similarity. *Oikos* 116:2097–2107
- Hui C, McGeoch MA (2008) Does the self-similarity species distribution model lead to unrealistic prediction. *Ecology* 89:2946–2952
- Hui C, McGeoch MA, Warren M (2006) A spatially explicit approach to estimating species occupancy and spatial correlation. *J Anim Ecol* 75:140–147
- Hui C, McGeoch MA, Beyers B, Le Roux PC, Greve M, Chown SL (2009) Extrapolating population size from the occupancy–abundance relationship and the scaling pattern of occupancy. *Ecol Appl* 19:2038–2048
- Hui C, Veldtman R, McGeoch MA (2010) Measures, perceptions and scaling patterns of aggregated species distributions. *Ecography* 33:95–102
- Kunin WE (1998) Extrapolating species abundance across spatial scales. *Science* 281:1513–1515
- Legendre P, Fortin MJ (1989) Spatial pattern and ecological analysis. *Vegetatio* 80:107–138
- Mladenoff DJ, White MA, Pastor J, Crow TR (1993) Comparing spatial pattern in unaltered old-growth and disturbed forest landscapes. *Ecol Appl* 3:294–306
- Ostling A, Harte J, Green J (2000) Self-similarity and clustering in the spatial distribution of species. *Science* 290:671
- Perry JN, Liebhold AM, Rosenberg MS et al (2002) Illustrations and guidelines for selecting statistical methods for quantifying spatial patterns in ecological data. *Ecography* 25:578–600
- Perry GLW, Miller BP, Enright NJ (2006) A comparison of methods for the statistical analysis of spatial point patterns in plant ecology. *Plant Ecol* 187:59–82
- Pielou EC (1960) A single mechanism to account for regular, random and aggregated populations. *J Ecol* 48:575–584
- Plotkin JB, Muller-Landau HC (2002) Sampling the species composition of a landscape. *Ecology* 83:3344–3356
- Sakhr J, Nieminen JM (2006) Spacing distributions for point processes on a regular fractal. *Phys Rev E* 73:036201
- Stoyan D, Penttinen A (2000) Recent applications of point process methods in forestry statistics. *Stat Sci* 15:61–78

- Thompson HR (1956) Distribution of distance to n-th nearest neighbor in a population of randomly distributed individuals. *Ecology* 37:391–394
- Tilman D, Kareiva P (1997) *Spatial ecology: the role of space in population dynamics and interspecific interactions*. Princeton University Press, Princeton
- Wiegand T, Moloney KA (2004) Rings, circles, and null models for point pattern analysis in ecology. *Oikos* 104:209–229
- Zillio T, He F (2010) Modeling spatial aggregation of finite populations. *Ecology* 91:3698–3706