Inferring multispecies distributional aggregation level from limited line transect-derived biodiversity data

Youhua Chen | Tsung-Jen Shen | Hoang Van Chung | Shengchao Shi | Jianping Jiang | Richard Condit | Stephen P. Hubbell

1CAS Key Laboratory of Mountain Ecological Restoration and Bioresource Utilization & Ecological Restoration and Biodiversity Conservation Key Laboratory of Sichuan Province, Chengdu Institute of Biology, Chinese Academy of Sciences, Chengdu, China; 2Institute of Statistics & Department of Applied Mathematics, National Chung Hsing University, Taichung, Taiwan; 3Field Museum of Natural History, Chicago, IL; 4Morton Arboretum, Lisle, IL; 5Smithsonian Tropical Research Institute, Apartado, Balboa, Panama and 6Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA

Abstract

1. Ecologists have generally concluded that species distributions are not random (e.g. aggregate), based on single-species studies that applied single-species-based statistical methods, like the negative binomial model. Although it is common knowledge that some specific species in an ecological community present aggregate distributions, this does not necessarily imply that the entire community presents an aggregate distribution. Studying community-level distributional aggregation patterns requires different statistical methods.

2. Herein, by utilizing a novel conspecific-encounter index derived from a multiple species Markov transition model that accounts for the non-independent sampling of consecutive individuals along line transects, we were able to show that tree assemblages in tropical forest ecosystems can present a strong signal of extensive distributional interspersion.

3. This interesting result is not unexpected, given the fact that neighbouring tree individuals in highly diverse tropical forests are usually of different species, resulting in strong niche packing or interspersed patterns. In contrast, for the amphibian assemblages surveyed from southwestern China and central-south Vietnam, the conspecific-encounter index was found to be consistently high, implying that amphibian communities tend to be highly aggregate in space.

4. Conclusively, using the conspecific-encounter index derived from the Markov non-independent sampling model, we provide a legible definition of community-level distributional aggregation as an interspersed or cluster-like distribution of different species. This definition is not idiosyncratic, as it is coincident with the definition of the contagion index used in landscape ecology. To this end, the model used in this paper establishes a framework explicitly linking community ecology and landscape ecology from a multi-object perspective.

KEYWORDS

biodiversity, distributional aggregation, distributional interspersion, landscape ecology, Markov chain, negative binomial model, neighbouring effect, sampling theory
1 | INTRODUCTION

Species distributions are not random in nature. Diverse distributional patterns are commonly found in field practice, including regular, aggregate and random distributions. Nearly all previous studies evaluating distributional patterns of species were conducted at the single-species level (He & Gaston, 2000; Zillio & He, 2010), largely ignoring the joint effects of multiple species and the associated joint distribution patterns. Studies of single-species distributional aggregation patterns imply that they are less likely to accurately evaluate how multiple species interact to cause joint distributional aggregation patterns from a whole-community perspective. This is because, community-level biodiversity patterns can present nonlinear and second-order interactions that might not be revealed using single-species metrics or indices (Chen, Shen, Condit, & Hubbell, 2018; Katabuchi et al., 2017). To this end, the development of statistical methods quantifying multispecies distributional aggregation pattern at the community level becomes urgently necessary, which is particularly true when the sampling and collection of field biodiversity data are limited.

Line transect is one of the most common strategies by ecologists to conduct biodiversity sampling in the field for its simplicity and effectiveness (Buckland, Borchers, Johnston, Henrys, & Marques, 2007; Burnham, Anderson, & Laake, 1980; Plumptre, 2000). It has been widely adopted by herpetologists for sampling amphibians and reptiles, ornithologists for watching birds, and botanists for recording trees (Buckland et al., 2007; Hsu, Kam, & Fellers, 2005; Lacki, Hummer, & Fitzgerald, 1994; Plumptre, 2000). Compared to other methods like quadrat-based sampling, the line transect method can be more time saving and labour inexpensive. Moreover, relevant statistical methods on estimating population size of species have been well-established for line transect sampling (Burnham et al., 1980). However, to date, no statistical methods have been built for inferring community-level diversity and distributional aggregation patterns based on the limited biodiversity data collected from line transects.

There are indeed some metrics available for quantifying spatial distributional aggregation of species in early ecological literature (Clark & Evans, 1954; Green, 1966; Lloyd, 1967; Morisita, 1959; O’Neill et al., 1988; Pielou, 1959). However, most of them are established at the single-species level, and many are calculated using the biodiversity data derived from quadrat-based sampling (Green, 1966; Lloyd, 1967; Morisita, 1959). Some metrics are developed based on sampling distances between neighbouring individuals (Clark & Evans, 1954; Mountford, 1961; Pielou, 1959). However, this may increase the sampling workload (or at least sampling time) as ecologists have to record the geographic coordinates for each individual. To this end, there is no biodiversity index available for quantifying multispecies distributional aggregation patterns in the context of sequential sampling of organisms without knowing sampling distances or coordinates along line transects.

In this study, we introduce a multispecies Markov transition model to account for the non-independent sampling of subsequent individuals along line transects. On the basis of this Markov model, we develop a novel conspecific-encounter index that can allow ecologists, at the community level, to evaluate multispecies distributional aggregation patterns effectively based on the limited biodiversity data sampled from line transects. By conducting extensive numerical simulations and empirical tests using the proposed index, we proved that multispecies distributional patterns in highly diverse tropical forest communities are not as aggregate as inferred based on limited information that some specific species in the target communities present highly aggregate distributions. In other words, even though we may observe that some species are aggregate distributed, distributions of different species at the community level tend to overlap with each other, possibly implying a strong signal of dense niche packing or overlap (May, 1974; Rappoldt & Hogeweg, 1980).

2 | MATERIALS AND METHODS

2.1 | A Markov model for sequential sampling of individuals and the conspecific-encounter index \( \nu \)

We utilized a Markov sampling model (Solow, 2000) to establish a non-independent sampling framework for line transect sampling of individuals of species in field practice. To be specific, suppose the true relative abundances of \( S \) species in a metacommunity are given by \( p_i, i = 1, ..., S \) with \( \sum_{i=1}^{S} p_i = 1 \). Moreover, suppose ecologists consecutively sample \( m \) individuals one by one from the metacommunity, in which we record the sampling sequence as \( Z_k, k = 1, 2, ..., m \) (representing the species label of the \( k \)th sampled individual). Specifically, the underlying probability model of the sampling process is that the first individual is assumed to have been randomly sampled based on the species’ relative abundance, that is,

\[
P(Z_1 = i) = p_i, \tag{1}
\]

and subsequently sampled individuals follow the transition probabilities of a first-order Markov chain (Solow, 2000) as

\[
P(Z_k = j | Z_{k-1} = i) = \begin{cases} (1 - \lambda)p_i + \xi, & j = i \\ (1 - \lambda)p_j, & j \neq i \end{cases} \tag{2}
\]

The aim of Equation 2 is to model the relationship between any two adjacent individuals of the line transect sample. Specifically, given that the \( (k - 1) \)th individual belongs to species \( i \) (i.e. \( Z_{k-1} = i \)), the next individual observed (the \( k \)th individual) in the line transect sample can be either from species \( i \) (i.e. \( Z_k = i \)) with probability \((1 - \lambda)p_i + \xi\) or from another species \( j \) (i.e. \( Z_k = j \)) with probability \((1 - \lambda)p_j\), where species \( j \) can be any species except species \( i \). This model is called the first-order Markov probability model (Equation 2) because the sum of all the probabilities is equal to one (i.e. \( \sum_{j} (1 - \lambda)p_j + (1 - \lambda)p_i + \xi = 1 \)) and the \( k \)th state is only determined by the state at the \( (k - 1) \)th state.

Note that parameter \( \lambda \), with a value ranging from 0 and 1, is closely related to the measurement of the degree of non-independence of any two adjacent individuals which are of different species or the same species in the sampling model of Equations 1 and 2. This
is because, if \( x = 0 \), the sampling procedure actually leads to an independent sample of individuals (i.e. the species identity of the next individual sampled is fully determined by the relative abundance of species). In contrast, if \( x = 1 \), the sampling procedure will only sample individuals from a single species that is permitted in the metacommunity. A high \( x \) value implies that it is unlikely an individual from a different species will be encountered.

Given a line transect sample of sequential species labels \( Z_s \), we propose a new conspecific-encounter index \( \nu \) with the calculation formula of

\[
\nu = \frac{1}{m-1} \sum_{k=2}^{m} \sum_{i=1}^{s} I(Z_k = i, Z_{k-1} = i)
\]

(3)

to measure the non-independent level of the metacommunity. The denominator \( m - 1 \) in Equation 3 is the total number of adjacent pairs in the line transect sample while the indicator function, \( I(Z_k = i, Z_{k-1} = i) \), in the numerator of Equation 3 has a value of 1 to reflect the realistic situation that the kth and \((K-1)\)th adjacent individuals are conspecific as species \( i \) and a value of 0 otherwise. As a consequence, we develop the new conspecific-encounter index \( \nu \) as the ratio of the total number of conspecific pairs to that of all adjacent pairs in the line transect sample. A simple hypothetical example demonstrating the calculation of the new index \( \nu \) for quantifying the multispecies distributional aggregation is shown in Figure 1.

Actually, the index \( \nu \) is an unbiased estimate of the probability of observing any two adjacent sampled individuals that belong to the same species in the metacommunity when using data from a line transect sample of \( m \) individuals. Note that this index has an immediate relationship with parameter \( x \) as \( E(\nu) = x + (1 - x) \sum_{i=1}^{p} p_i^2 \), but the proposed index \( \nu \) has a more immediate meaning for our research purpose than the latter. Additionally, as to the research aim in Solow (2000) of estimating the sample coverage that is crucial to species richness estimations, he suggested that \( \nu \) could serve as a good surrogate of \( x \) when applied to the Markov sampling model. In contrast, an application of index \( \nu \) to our topic is much more straightforward and important since it not only involves no parameter estimations (which usually makes the index estimation complicated, thus inflating variation) but also implicitly takes the parameters (\( x \) and species relative abundances, \( p_i \)) of the Markov model into account.

For constructing an interval estimation of the conspecific-encounter index, we also propose a variance estimator of the index as follows:

\[
\hat{\text{Var}}(\nu) = \frac{1}{m-1} \left\{ \nu(1-\nu) + \frac{2}{1-\nu} \left( 1 - \nu^2 \right) I\left( t \geq \nu^2 \right) \right\},
\]

(4)

where the sample size \( m \) requires a constraint as \( m \geq 3 \) and

\[
t = \frac{1}{m-2} \sum_{k=2}^{m} \sum_{i=1}^{s} I(Z_k = i, Z_{k-1} = i).
\]

Detailed derivation of Equation 4 is given in the Supporting Information. Using the result in Equation 4, a 95% confidence interval of the conspecific-encounter index is of the form as \( \nu \pm 1.96 \sqrt{\hat{\text{Var}}(\nu)} \).

We also evaluate the robustness of the conspecific-encounter index and the above variance estimator in case when the first-order Markov property is violated. Details are presented in the Supporting Information.

### 2.2 Numerical simulation

We utilized a Poisson cluster process to simulate spatial distributions of individuals for each of 100 hypothetical species over a 200 x 200 lattice (Chen, 2013, 2014; Plotkin et al., 2000). In brief, we created a random number of parental points for a species using a Poisson process with intensity \( \rho \), which is assumed to be the same for each species with a value of 3. Then, each parental point produces a random number of offspring points (treated as species individuals), drawn independently from any appropriate probability models (e.g. Poisson distribution). Here, we consider that the offspring numbers were drawn from a geometric distribution with parameter 0.01. The
In highly aggregate or less aggregate communities are visually presented and compared in Figure 2. This process may be more formally named as the Thomas cluster process (Baddeley & Turner, 2005; Thomas, 1949) with slight modifications (i.e. the density of offspring of each cluster follows a geometric distribution) or the Neyman-Scott cluster process (Baddeley, Rubak, & Turner, 2015).

In addition to the Thomas cluster process used here, for testing the general relationship between diverse spatial distribution patterns of species and the proposed conspecific-encounter index, we also simulated four other alternative clustering or regular spatial point models using R package spatstat (Baddeley & Turner, 2005), detailed introductions of which are presented in Supporting Information.

2.3 | Empirical tests

We utilized mapping data from a Barro Colorado Island (BCI) forest plot in Panama to illustrate the proposed conspecific-encounter index. The BCI forest plot has an area of 50 ha (1,000 m × 500 m) and was initially established by Stephen Hubbell and Robin Forster in 1980 (Condit, 1998; Condit, Hubbell, & Foster, 1996; Condit et al., 2002, 2012; Hubbell et al., 1999). In the present study, we only used 2005 census data. All trees with a diameter at breast height of ≥10 mm were used in this study.

We used line transects to conduct subsequent sampling of individuals in the BCI forest plot and the two simulated hypothetical communities for the following analyses. In detail, a line transect with a fixed width (e.g. 1 m here) is randomly placed in the community (Figure 3), and the first individual sampled within the line transect band is randomly selected at one end of the band. Excluding the preceding point, the next individual is selected as the one at a minimal distance with respect to the initial point, and so forth. Sampling is complete when the pre-designed sample size (m) is fulfilled. In our study, we designed small sample sizes of 50, 75, 100 and 125. Small sample sizes are commonly encountered in practical field sampling, especially in the sampling of animal individuals. In our study, 2,000 line transect bands were used as replicates for each combination of a specific sample size and either simulated or empirical datasets.

2.4 | Field application

To demonstrate the field application of the proposed index \( \nu \), we also conducted real-world line transect sampling of amphibians in Minya Konka from the central part of Sichuan Province, southwestern China and in national parks from the central-south part of Vietnam (Figure S1). We followed the same protocol when conducting line transect sampling. To be specific, when walking over each line transect, we took photos for each amphibian individual encountered sequentially and recorded the photography time accordingly. Organisms of some species can be easily identified in the field, while the others were identified in the laboratory through photos.

Minya Konka, with an average elevation of 6,000 m, is the highest mountain peak in the Hengduan mountainous range and Sichuan...
At the nighttime (from 7 p.m. to 12 p.m.). Locations of line transects were determined in prior at the daytime (from 6 a.m. to 9 a.m. in the morning or from 3 p.m. to 6 p.m. in the afternoon).

A total of 46 line transects were sampled with transect lengths ranging from 63 m to 1,200 m. Because the observed species richness and abundance in each line transect were not large (usually no more than three species and five individuals during the raining season), we combined the sampled data from all the line transects for each county (we use the surveyed data from three counties in this study: Kangding, Luding and Shimian) while retaining the sampling order of different line transects (based on the sampling time) and the subsequent sampling of amphibian individuals inside each line transect when applying the proposed conspecific-encounter index.

Besides, sequential line transect sampling of amphibians in the central-south part of Vietnam was conducted from 27 June to 8 July 2018. A total of 46 line transects were sampled with transect lengths ranging from 400 m to 2,500 m. Specimens were collected by hand at the night-time (from 7 p.m. to 12 p.m.). Locations of line transects were determined in prior at the daytime (from 6 a.m. to 9 a.m. in the morning or from 3 p.m. to 6 p.m. in the afternoon).

Because we combined the sampled biodiversity data derived from different line transects in the empirical tests, it is of importance to evaluate the potential effect of line transect combination on the estimation of the conspecific-encounter index. Details about the simulation procedure, results and discussion were presented in the Supporting Information. What we can confirm here was that the combination of multiple line transects had similar estimation of the index with respect to single line transects (Figures S4 and S5 in the Supporting Information), and accordingly, the validity of using combined multiple line transect data in the analyses was supported.

### TABLE 1

| Data         | Sample size (m) | 50   | 75   | 100  | 125  |
|--------------|-----------------|------|------|------|------|
| BCI          | Estimated \( \nu \) | 0.098| 0.097| 0.102| 0.103|
|              | Observed \( S \) | 25   | 32.5 | 38.6 | 44   |
| Less aggregate | Estimated \( \nu \) | 0.049| 0.048| 0.047| 0.047|
|              | Observed \( S \) | 26.4 | 35.3 | 42.6 | 49.1 |
| More aggregate | Estimated \( \nu \) | 0.466| 0.451| 0.447| 0.444|
|              | Observed \( S \) | 10.5 | 14.2 | 17.9 | 21.2 |

In the numerical tests, for the more aggregate hypothetical community in which distributions of species are highly aggregated and tend to present clustering patterns (\( \sigma^2 = 50 \)), the conspecific-encounter index, \( \nu \), was very high as expected (Figure 2a; Table 1). In contrast, for the less aggregated community in which species’ distributions are widespread (\( \sigma^2 = 500 \); Figure 2b), the conspecific-encounter index, \( \nu \), had very low values (less than and close to 0.05; Table 1). The distribution of \( \nu \) values over the replicates further confirmed these results, regardless of how the sample size varied (Figure 4). Furthermore, the mean and median of the fitted values were coincident and did not change when the sample size increased (Figure 4). The cases for alternative clustering processes (Matern and Cauchy cluster processes) presented similar results (Figure S1): the conspecific-encounter index \( \nu \) will decrease when increasing the scale parameter (increasing clustering degree). The regular processes (Strauss and Strauss hard processes) will have a reverse relationship (Figure S2). That is, increasing the inhibition strength (the resultant distribution of species will become more regular) will result in low conspecific-encounter index \( \nu \) (Figure S2).

For the empirical test using the BCI dataset, the value of \( \nu \) of the conspecific-encounter index was around 0.1 (Table 1). The distribution of \( \nu \) values over the replicates further confirmed such a result, regardless of how the sample size varied (Figure 4). Also, like the results for the hypothetical communities above, the mean and median of the fitted values were coincident and did not change when the sample size increased (Figure 4).
In all of the numerical and empirical tests (Table 1), one can see that values of \( \nu \) were not sensitive to changes in the sample size or the number of species identified from the sample. Additionally, fluctuations in the values of \( \nu \) over different line transects for the numerical and empirical tests consistently decreased with the sample size. From this observation, one can see that the index \( \nu \) possessed good statistical properties for measuring the degree of distributional interspersion at the community level, as it could be independent of the sample size and species richness.

The observed number of species for the less aggregated community increased much more quickly with the sample size than that for the more aggregated one (Table 1). Although we expected that more species would be observed when the sample size enlarged, the aggregate patterns of species individuals can vary with the increasing rate when the compared communities had the same species richness.

For the field application, it was observed that, in three national parks of central-south Vietnam and the Minya Konka of southwest China, the conspecific-encounter index \( \nu \) calculated from the surveyed amphibian data using the sequential line transect sampling was consistently high, being larger than or equal to 0.5 in most cases (Table 2).

**TABLE 2** Field application of the conspecific-encounter index to the amphibian surveyed data in the Minya Konka of southwestern China (three counties: Kanding, Luding and Shimian) and the central-south part of Vietnam (three national parks: Bidoup Nui Ba, Chu Yang Sin and Phuoc Binh). The 95% confidence interval was calculated using Equation 4. Sequential order of organisms along each line transect was determined by photography time.

| Country  | Location               | Geographic coordinates | Estimated \( \nu \) | Lower bound of 95% CI | Upper bound of 95% CI |
|----------|------------------------|------------------------|-------------------|-----------------------|----------------------|
| Vietnam  | Bidoup Nui Ba National Park | N: 12°21′; E: 108°42′ | 0.522             | 0.461                 | 0.584                |
| Vietnam  | Chu Yang Sin National Park | N: 12°24′; E: 108°21′ | 0.492             | 0.316                 | 0.669                |
| Vietnam  | Phuoc Binh National Park  | N: 12°01′; E: 108°44′ | 0.715             | 0.622                 | 0.809                |
| China    | Kangding               | N: 29°59′; E: 101°57′ | 0.517             | 0.37                  | 0.664                |
| China    | Luding                 | N: 29°54′; E: 102°14′ | 0.588             | 0.194                 | 0.982                |
| China    | Shimian                | N: 29°13′; E: 102°21′ | 0.312             | 0.085                 | 0.54                 |
4 | DISCUSSION

Ecologists usually have the common sense that ecological communities should present aggregate distributions, especially in tropical forests (Plotkin et al., 2000; Umaña et al., 2017; Zillio & He, 2010). Such a belief is based on observations of highly aggregate distributions of some specific species found within a community (Shao et al., 2018; Zillio & He, 2010), in which each species is independently studied without considering interspecific interactions. Here, using line transect sampling along with a novel statistical model, we challenge this intuitive recognition and present a counterintuitive observation: ecological communities, at least for tropical tree species in BCI forest plot, might not be as aggregate in distribution as we expected (Table 1), even though the field application of the proposed index in amphibian communities of three national parks of central-south Vietnam and Minya Konka of southwestern China were consistently found to be high (Table 2). For the BCI forest plot, the distributions of different tree species tend to intersperse with each other. Such an observation on BCI tree community has been confirmed previously using a negative multinomial model (Chen et al., 2018). Why is this community-level evaluation of distributional aggregation in tree assemblages of BCI plot so surprising and contradictory to our traditional recognition? The reasons can be multifaceted.

A key reason is that if some species present wide-ranging distributions, other competing species may tend to be aggregate distributed so as to reduce interspecific interactions and avoid coexisting with other species. The consequence of biotic interactions is that the distributional pattern of each species differs within an ecological community: some species have very aggregate distributions, while others may have random or even regular distributions. As a result, the community-level joint distributional pattern of all species will very likely be random.

Another reason may be due to the fact that the relatively small extent of a studied local forest plot has relatively homogeneous environmental conditions, in which environmental covariates do not change remarkably (even though the responses of each specific species to the homogeneous environmental conditions may vary, that is, some species grow quickly and tend to actively disperse across the studied area, while other species may grow slowly with a fairly low dispersal ability). The joint response of all species distributed in a relatively homogeneous landscape, therefore, will present a somewhat random distributional pattern.

Moreover, when individuals of all species in an ecological community are superimposed on a map (e.g. 190,506 stems in a 50-ha rectangle for the BCI plot), the map will have very dense distributional points across all corners and locations. In such a case, it is easy to imagine that the entire community indeed presents a strong signal of distributional interspersion among species. Since the distributional points of individuals of all species occupy almost all locations of the map, when conducting subsequent sampling of individuals using the line transect method across the map, any two neighbouring individuals from different species are very likely to be observed. As a result, this makes the value of the conspecific encounter, $\nu$, very low. That is, it is very easy to encounter individuals of different species when travelling across the ecological community, implying that the distributions of different species within the community are interlaced over the space.

The above discussion is restricted to plant communities. However, for animal communities, at least for the amphibian assemblages surveyed in the field, individuals of the same species tend to group in space (Table 2). This observation implies that amphibian assemblages distribute aggregately, which can be understandable from the facts that this taxonomic group has limited dispersal capacity and several adult organisms can reproduce many tadpoles in the same habitat mosaic (Fei, 1999; Fei, Ye, & Jiang, 2012; Gonçalves, Honrado, Vicente, & Civantos, 2016; Semlitsch, 2008; Smith & Green, 2005). These drive the highly aggregate nature with respect to the community-level distribution pattern of amphibians.

So, what is the community-level distributional pattern? How should ecologists imagine and define it? In the present study, the introduction of a Markovian model accounted for the non-independent subsequent sampling of individuals and clearly demonstrated that the community-level distributional pattern was closely related to the interspersed distributions (or inversely, non-overlapping distributional clusters) of different species (Figure 1). That is, when the community is filled with species that tend to avoid each other by forming distinct spatial patches or clusters, the expected value of the conspecific-encounter index would be high, because it is very likely that two subsequent individuals from the same species will be observed. Therefore, the patch-like (or alternatively, non-interspersing) distributions of species in an ecological community will greatly contribute to the community-level distributional pattern (Figure 1). This suggests that the community-level distributional pattern quantifies the degree of spatial interspersion of different species. Correspondingly, community-level distributional aggregation quantifies the spatial non-overlapping clustering of species. This conclusion is not idiosyncratic in ecology. Actually, the definition of an aggregation index in landscape ecology (also called a contagion index) is also related to the patch-like distribution of different landscape classes (He, DeZonia, & Mladenoff, 2000; McGarigal & Marks, 1995; O’Neill et al., 1988): when different landscape classes have patch- or cluster-like distributions and tend to avoid each other, the aggregation index is expected to be high.

Finally and notably, the definition of community-level distributional aggregation or clustering discussed above makes sense with regard to the multiple species case. As an extreme case, if an ecological community has a single species and its distribution over the studied domain is very even or random, then computation of the conspecific-encounter index, $\nu$, could be as high as 1. This calculated value (implying that the distribution of the species is highly aggregate) is contradictory to the true situation (i.e. the actual distribution of the species is even or random as just mentioned). Therefore, we argue that the conspecific-encounter index, $\nu$, is meaningful only under multispecies situations (i.e.
species number ≥2). In contrast, a single-species distribution over a target area represents a single distributional cluster (regardless of whether its distribution is densely aggregate, even or random), with no evidence of distributional interspersion. Given that the assessment of interspersed distribution requires the number of compared species to be at least two so as to have a contrast over different species, the conspecific-encounter index υ is most applicable to measure multispecies distributional aggregation patterns at the community level.

The above discussion also implies that the interspersion-minimizing or patch-like distribution of species within an ecological community is an indicator that intraspecific biotic interactions are prevailing over interspecific biotic interactions. To this end, the conspecific-encounter index, υ, can also reflect the relative importance of interspecific versus intraspecific interactions in shaping the spatial community structure. Given that the empirical tropical dataset (BCI) we tested has a very low fitted υ value, interspecific competition might not be very strong in forest communities in the region (i.e. because of low competition, species prefer to have different species as neighbours). Thereby, this result provides indirect evidence showing that the mechanism of conspecific-negative density dependence (He & Duncan, 2000; Johnson, Beaulieu, Bever, & Clay, 2012; Volkov, Banavar, He, Hubbell, & Maritan, 2005) is indeed important in maintaining high biodiversity levels in highly diverse tropical communities.

Conclusively, the Markov model for fitting subsequent sampling of individuals along line transects used in the present study establishes an elegant link between community ecology and landscape ecology. By using the conspecific-encounter index to quantify correlated distributions of neighbouring individuals of species, it was found that community-level distributional patterns actually reflect interspersed or cluster-like distributions of different species (Figure 1).

ACKNOWLEDGEMENTS

The authors thank the editor Dr. Andres Baselga and two anonymous reviewers for their constructive comments that greatly help improve the manuscript; and the Center for Tropical Forest Science (CTFS) for generously providing the BCI plot data. This research was financially supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant Number: XDB31000000), the National Key Research and Development Program of China (2017YFC0505202) and the Hundred Talents Program of the Chinese Academy of Sciences (Y8C3041100). The BCI forest dynamics research project was founded by S.P. Hubbell and R.B. Foster and is now managed by R. Condit, S. Lao and R. Perez under the CTFS and the Smithsonian Tropical Research Institute in Panama. Numerous organizations provided funding, principally the U.S. National Science Foundation, and hundreds of field workers have contributed. T.J.S. was supported by the Taiwan Ministry of Science and Technology. H.V.C. thanked Dr. Tao Nguyen Thien for the support, and Anh Luong Mai and Yen Nguyen Thi for the assistance in the field works. The authors declare that there is no conflict of interest.

AUTHORS’ CONTRIBUTIONS

Y.C., T.-J.S., R.C. and S.P.H. conceived the initial ideas, contributed the research data and tools; Y.C. wrote a preliminary draft and conducted numerical simulations; T.-J.S. derived the model and analysed the data; Y.C. and J.P. designed the field survey protocol and examined the field data; H.V.C. and S.C. conducted the field sampling and handled the raw data. All authors contributed critically to the drafts and gave final approval for publication.

DATA ACCESSIBILITY

The dataset (BCI) we used in the paper is from the Center for Tropical Forest Science (CTFS). The terms and conditions prevent us from archiving these data in a publicly accessible repository. However, anyone can apply for using the BCI dataset of CTFS after filling out an application form which can be found here: http://ctfs.si.edu/webatlas/datasets/bci/bciiformtest_download.php. The Chinese and Vietnamese amphibian datasets we used in the section “Field application” are archived in and downloadable from an online repository (https://doi.org/10.5281/zenodo.2647164).

ORCID

Tsung-Jen Shen https://orcid.org/0000-0002-1742-9730

REFERENCES

Baddeley, A., Rubak, E., & Turner, R. (2015). Spatial point patterns: Methodology and applications with R. Boca Raton, FL: Chapman and Hall/CRC. https://doi.org/10.1201/b19708

Baddeley, A., & Turner, R. (2005). spatstat: An R package for analyzing spatial point patterns. Journal of Statistical Software, 12, 1–42.

Buckland, S., Borchers, D., Johnston, A., Henrys, P., & Marques, T. (2007). Line transect methods for plant surveys. Biometrics, 63, 989–998. https://doi.org/10.1111/j.1541-0420.2007.00798.x

Burnham, K., Anderson, D., & Laake, J. (1980). Estimation of density from line transect sampling of biological populations. Wildlife Monographs, 72, 3–202.

Chen, Y. (2013). Species-area relationship is overestimated using distributional range maps. Theoretical Biology Forum, 106, 17–21.

Chen, Y. (2014). Estimation of species extinction: What are the consequences when total species number is unknown? Theory in Biosciences, 133, 175–178. https://doi.org/10.1007/s12064-014-0202-2

Chen, Y., Shen, T., Condit, R., & Hubbell, S. (2018). Community-level species’ correlated distribution can be scale-independent and related to the evenness of abundance. Ecology, 99, 2787–2800. https://doi.org/10.1002/ecy.2544

Clark, P., & Evans, F. (1954). Distance to nearest neighbor as a measure of spatial relationships in populations. Ecology, 35, 445–453. https://doi.org/10.2307/1931034

Condit, R. (1998). Tropical forest census plots. Berlin, Germany and Georgetown, Texas: Springer-Verlag and R. G. Landes Company. https://doi.org/10.1007/978-3-662-03664-8
Condit, R., Hubbell, S., & Foster, R. (1996). Changes in a tropical forest with a shifting climate: Results from a 50-ha permanent census plot in Panama. *Journal of Tropical Ecology, 12*, 231–256. https://doi.org/10.1017/S0266467400009433

Condit, R., Lao, S., Perez, R., Dolins, S., Foster, R., & Hubbell, S. (2012). Barro colorado forest census plot data, 2012 version. Center for Tropical Forest Science Databases. https://doi.org/10.5479/data.BF02524740

Condit, Richard, Pitman, N., Leigh, E. G., Chave, J., Terborgh, J., Foster, R. B., ... Hubbell, S. P. (2002). Beta-diversity in tropical forest trees. *Science, 295*(5555), 666–669. https://doi.org/10.1126/science.1066854

Fei, L. (1999). Atlas of amphibians of China. Zhengzhou, China: Henan Science and Technology Press.

Fei, L., Ye, C., & Jiang, J. (2012). Colored atlas of Chinese amphibians and their distributions. Chengdu, China: Sichuan Science & Technology Press.

Gonçalves, J., Honrado, J. P., Vicente, J. R., & Civantos, E. (2016). A model-based framework for assessing the vulnerability of low dispersal vertebrates to landscape fragmentation under environmental change. *Ecological Complexity, 28*, 174–186. https://doi.org/10.1016/j.ecocom.2016.05.003

Green, R. (1966). Measurement of non-randomness in spatial distributions. *Research in Population Ecology, 8*, 1–7. https://doi.org/10.1007/BF02524740

He, H., DeZonia, B., & Mladenoff, D. (2000). An aggregation index (AI) to quantify spatial patterns of landscapes. *Landscape Ecology, 15*, 591–601. https://doi.org/10.1023/A:1008102521322

He, F., & Duncan, R. (2000). Density-dependence effects on tree survival in an old-growth Douglas fir forest. *Journal of Ecology, 88*, 676–688. https://doi.org/10.1046/j.1365-2745.2000.00482.x

He, F., & Gaston, K. (2000). Estimating species abundance from occurrence. *The American Naturalist, 156*, 553–559. https://doi.org/10.1086/303403

Hsu, M., Kam, Y., & Fellers, G. (2005). Effectiveness of amphibian monitoring techniques in a Taiwanese subtropical forest. *Herpetological Journal, 15*, 73–79.

Hubbell, S., Foster, R., O’Brien, S., Harms, K., Condit, R., Wechsler, B., ... de Lao, S. (1999). Light gap disturbances, recruitment limitation, and tree diversity in a neotropical forest. *Science, 283*, 554–557. https://doi.org/10.1126/science.283.5401.554

Johnston, D., Beaulieu, W., Bever, J., & Clay, K. (2012). Conspecific negative density dependence and forest diversity. *Science, 336*, 904. https://doi.org/10.1126/science.1220269

Katabuchi, M., Wright, S. J., Swenson, N. G., Feeley, K. J., Condit, R., Hubbell, S. P., & Davies, S. J. (2017). Contrasting outcomes of species- and community-level analyses of the temporal consistency of functional composition, *Ecology, 98*(9), 2273–2280. https://doi.org/10.1002/ecy.1952

Lacki, M. J., Hummer, J. W., & Fitzgerald, J. L. (1994). Application of line transects for estimating population density of the endangered copperbell water snake in Southern Indiana. *Journal of Herpetology, 28*(2), 241–245. https://doi.org/10.2307/1564627

Lloyd, M. (1967). Mean crowding. *Journal of Animal Ecology, 36*, 1–30. https://doi.org/10.2307/3012

May, R. (1974). On the theory of niche overlap. *Theoretical Population Biology, 5*, 297–332. https://doi.org/10.1016/0040-5809(74)90055-0

McGarigal, K., & Marks, B. (1995). Fragstats: Spatial pattern analysis program for quantifying landscape structure. Portland, OR: U.S. Department of Agriculture, Forest Service. Pacific Northwest Research Station. https://doi.org/10.2737/PNW-GTR-351

Morisita, M. (1959). Measuring of the dispersion and analysis of distributional patterns. *Memoires of the Faculty of Science, Kyushu University, Series E. Biology, 2*, 215–235.

Mountford, M. (1961). On E.C. Pielou’s index of nonrandomness. *Journal of Ecology, 49*, 271–275. https://doi.org/10.2307/2257261

O’Neill, R., Krummel, J., Gardner, R., Sugihara, G., Jackson, B., DeAngelis, D., ... Graham, R. (1988). Indices of landscape pattern. *Landscape Ecology, 1*, 153–162. https://doi.org/10.1007/BF00162741

Pielou, E. (1959). The use of point-to-plant distances in the study of the pattern of plant populations. *Journal of Ecology, 47*, 607–613. https://doi.org/10.2307/2257293

Plotkin, J., Potts, M., Leslie, N., Manokaran, N., Lafrance, J., & Ashton, P. (2000). Species-area curves, spatial aggregation, and habitat specialization in tropical forests. *Journal of Theoretical Biology, 207*, 81–99. https://doi.org/10.1006/jtbi.2000.2158

Plumptre, A. (2000). Monitoring mammal populations with line transect techniques in African forests. *Journal of Applied Ecology, 37*, 356–368. https://doi.org/10.1046/j.1365-2664.2000.00499.x

Rappoldt, C., & Hogeweg, P. (1980). Niche packing and number of species. *The American Naturalist, 116*(4), 480–492. https://doi.org/10.1086/283643

Semptisch, R. D. (2008). Differentiating migration and dispersal processes for pond-breeding amphibians. *Journal of Wildlife Management, 72*(1), 260–267. https://doi.org/10.2193/2007-082

Shao, X., Brown, C., Worthy, S. J., Liu, L., Cao, M., Li, Q., ... Swenson, N. G. (2018). Intra-specific relatedness, spatial clustering and reduced demographic performance in tropical rainforest trees. *Ecology Letters, 21*(8), 1174–1181. https://doi.org/10.1111/ele.13086

Smith, M. A., & Green, D. M. (2005). Dispersal and the metapopulation paradigm in amphibian ecology and conservation: Are all amphibian populations metapopulations? *Ecography, 28*(1), 110–128. https://doi.org/10.1111/j.0906-7590.2005.04042.x

Solow, A. (2000). The effect of dependence on estimating sample coverage. *Environmetrics, 11*, 245–249. https://doi.org/10.1002/(ISSN)1099-095X

Thomas, M. (1949). A generalization of Poisson’s binomial limit for use in ecology. *Biometrika, 36*, 18–25. https://doi.org/10.1093/biomet/36.1.16

Umaña, M. N., Mi, X., Cao, M., Enquist, B. J., Hao, Z., Howe, R., ... Swenson, N. G. (2017). The role of functional uniqueness and spatial aggregation in explaining rarity in trees. *Global Ecology and Biogeography, 26*(7), 777–786. https://doi.org/10.1111/gbei.12583

Volkov, I., Banavar, J., He, F., Hubbell, S., & Maritan, A. (2005). Density dependence explains tree species abundance and diversity in tropical forests. *Nature, 438*, 658–661. https://doi.org/10.1038/nature04030

Zilio, T., & He, F. (2010). Modeling spatial aggregation of finite populations. *Ecology, 91*(12), 3698–3706. https://doi.org/10.1890/09-2233.1

**SUPPORTING INFORMATION**

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**How to cite this article:** Chen Y, Shen T-J, Van Chung H, et al. Inferring multispecies distributional aggregation level from limited line transect-derived biodiversity data. *Methods Ecol Evol*, 2019;10:1015–1023. https://doi.org/10.1111/2041-210X.13197