RESEARCH ARTICLE

Analytical models for β-diversity and the power-law scaling of β-deviation

Dingliang Xing[1,2,3] (iD) | Fangliang He[1,2]

[1]Department of Renewable Resources, University of Alberta, Edmonton, AB, Canada

[2]ECNU-Alberta Joint Lab for Biodiversity Study, Tiantong Forest Ecosystem National Observation and Research Station, School of Ecological and Environmental Sciences, East China Normal University, Shanghai, China

[3]Institute of Eco-Chongming (IEC), Shanghai, China

Correspondence
Dingliang Xing
Email: dlxing@des.ecnu.edu.cn

Present address
Dingliang Xing, School of Ecological and Environmental Sciences, East China Normal University, 500 Dongchuan Road, Shanghai, 200241, China

Funding information
Natural Science and Engineering Research Council; East China Normal University

Handling Editor: Nick Isaac

Abstract

1. β-diversity is a primary biodiversity pattern for inferring community assembly. A randomized null model that generates a standardized β-deviation has been widely used for this purpose. However, the null model has been much debated and its application is limited to abundance data.

2. Here we derive analytical models for β-diversity to address the debate, clarify the interpretation and extend the application to occurrence data.

3. The analytical analyses show unambiguously that the standardized β-deviation is a quantification of the effect size of non-random spatial distribution of species on β-diversity for a given species abundance distribution. It robustly scales with sampling effort following a power law with exponent of 0.5. This scaling relationship offers a simple method for comparing β-diversity of communities of different sizes.

4. Assuming log-series distribution for the metacommunity species abundance distribution, our model allows for calculation of the standardized β-deviation using occurrence data plus a datum on the total abundance.

5. Our theoretical model justifies and generalizes the use of the β null model for inferring community assembly rules.

KEYWORDS

β-diversity, log-series distribution, maximum entropy, METE, null model, spatial aggregation, species abundance distribution, species spatial pattern

1 | INTRODUCTION

Comparing β-diversity, the variation in species composition among different regions is an important yet controversial topic in ecology (Bennett & Gilbert, 2016; Kraft et al., 2011; Qian et al., 2013; Ulrich et al., 2017). A randomization-based null model that was said to correct for the dependence of raw β-diversity on species pool (Kraft et al., 2011) has been widely adopted for this purpose (e.g. De Cáceres et al., 2012; Myers et al., 2013; Vannette & Fukami, 2017; Xing & He, 2019; Zhang et al., 2020). However, this randomization approach has been criticized on several aspects, including incorrect interpretations of the β-deviation metric and the dependence of the

metric on sampling effort and species pool (Bennett & Gilbert, 2016; Qian et al., 2013; Ulrich et al., 2017, 2018). Another limitation of the null model is that it requires individual count data and thus limits its application to situations where only abundances of individual species in each local community are available. Therefore, there is a strong need for a theoretical scrutiny of the null model to justify its use and also a need to extend its application to presence/absence data.

It has been established that the abundance of species and their spatial aggregation in a metacommunity are the two basic quantities that construct macroecological patterns of diversity, including β-diversity (He & Legendre, 2002; Morlon et al., 2008; Plotkin & Muller-Landau, 2002). All ecological processes, biotic and abiotic, are acting

through these two quantities to affect the observed macroecological patterns (He & Legendre, 2002). For instance, both dispersal limitation and environmental filtering can result in spatially aggregated distribution of species, which, in turn, contributes to high β-diversity. However, the reverse is not necessarily true: a high β-diversity does not have to result from aggregated distribution. This is because differences in species abundances could also alter values of observed β-diversity. Therefore, it is necessary to disentangle the effect of spatial aggregation from that of species pool for inferring contributions from different ecological processes on β-diversity.

Different from the original β-diversity, Kraft et al.'s null model (and the resultant standardized β-deviation, or β-deviation for short) was designed to preserve species abundance distribution (SAD) of metacommunity while randomizing the species-level spatial distribution. It is thus clear that the β-deviation is a measure of the effect of non-random spatial distribution of species on β-diversity for a given SAD. Although this link between β-deviation and intraspecific aggregation of species was acknowledged in previous studies (e.g. Crist et al., 2003; Myers et al., 2013), including Kraft et al. (2011), this interpretation of β-deviation has not been widely heeded in the literature (e.g. Qian et al., 2013). The original interpretation of β-deviation was 'a standard effect size of β-diversity deviations from a null model that corrects for γ dependence' (Kraft et al., 2011) but this misinterpretation attracted much criticism on the null model (Bennett & Gilbert, 2016; Qian et al., 2013; Ulrich et al., 2017). While these debates stimulated searches for alternative null models aiming to correct for the γ-dependence (Ulrich et al., 2018), the problem of the misinterpretation has not been solved but undermines the use of otherwise a promising and important approach in community ecology (Myers et al., 2013; Vannette & Fukami, 2017; Xing & He, 2019). Here by developing an analytical null model, we reinforce that β-deviation measures the effect of species spatial distribution on β-diversity.

A second problem of the β-deviation is that the metric is subject to the effect of sampling effort (Bennett & Gilbert, 2016). In a recent study, we demonstrated by simulation that β-deviation can effectively identify and correctly compare non-random β patterns across assemblages under the condition of constant sampling effort (see Appendix S2 in Xing & He, 2019). This suggests that there might be some scaling relationship between β-deviation and the sampling effort. In this study, we show that there is actually a surprisingly simple and robust scaling relationship between β-deviation and sampling effort. The discovery of this scaling relationship is remarkable. It analytically addresses the criticisms raised by Bennett and Gilbert (2016) on the sample effort dependence of β-deviation.

The last shortcoming of the currently used null model, inherited from the randomization procedure, is that it only deals with abundance data. Yet, in many studies, particularly at regional scales, presence/absence of species are the only data available (e.g. Buckley & Jetz, 2008; Gaston et al., 2007; McKnight et al., 2007). Therefore, there is a need to extend its application to presence/absence data so that it could be used in broader contexts such as delineating biogeographical regions (Kreft & Jetz, 2010).

Here, based on theories about species abundance distribution and spatial distribution of species, we develop an analytical null model to replace the randomization procedure and to address all the above problems. The analytical results reveal that Kraft et al.'s β-deviation (a) is a measure of the effect size of non-random spatial species distribution on β-diversity for a given SAD, (b) scales with the sampling effort following a power law with a constant exponent of 0.5 and (c) can be reasonably estimated based on presence/absence data, provided that a datum on the total abundance of the metacommunity is known (which is usually more readily available than abundance for each species). We test the performance and utility of our analytically derived models using several well-studied empirical datasets.

2 | DERIVATION OF ANALYTICAL MODELS FOR β-DIVERSITY

Consider a metacommunity consisting of M equal-sized (equal area) local communities and there are in total N individuals belonging to S species. The widely used proportional species turnover, denoted β_P following Tuomisto (2010), is defined as $\beta_P = 1 - \bar{\alpha} / S$, wher $\bar{\alpha}$ is the average number of species over the M local communities. Note here that we are not claiming that M or the area of a local community is an inherent property of the metacommunity. Instead, they are determined by the sampling design. Two other major multi-community measures of β-diversity, namely the M-community Sorensen- (β_S) and Jaccard-differentiation (β_J) (Chao & Chiu, 2016), are monotonically related with β_P: $\beta_S = \frac{1}{M-1} \frac{\beta_P}{1-\beta_P}$, $\beta_J = \frac{M}{M-1} \beta_P$. Hence, although we will only focus on β_P in this paper, our analytical models can be easily transformed to these other indices.

From the definition and some simple rearrangements (see Appendix S1), it is easy to show that β_P is the expectation of the probability that a species from the metacommunity is absent from a local community and thus can be expressed as:

$$\beta_P = \sum_{n=1}^{\infty} \Pi(0|n, M) \Phi(n|S, N), \tag{1}$$

where $\Pi(0|n, M)$ is the probability that a species with n individuals in the metacommunity is absent from a randomly chosen local community, and $\Phi(n|S, N)$ is the metacommunity species abundance distribution (SAD), that is, the probability of a species having n individuals. This model provides a general framework to develop analytical solutions for Kraft et al.'s null model and the resultant β-deviation.

2.1 | Null model: β-diversity under random spatial distribution of species

Kraft et al.'s null model randomly shuffles species identities across the N individuals that comprise the M local communities while keeping the metacommunity SAD and each of the local community sizes

unchanged. By repeating the randomization procedure, a large number of times (e.g. 999) one can get the expectation (β_{null}) and corresponding variance ($Var_{\Pi}(\beta_{null})$) of β-diversity under the null model. Here the subscript Π is used to denote that the variation is due to species spatial distribution, in contrast to that due to SAD, which will be made clear below. This randomization is equivalent to assuming random spatial patterns of species. β-deviation is then defined as $\beta_{dev} = \frac{\beta_{obs} - \beta_{null}}{\sqrt{Var_{\Pi}(\beta_{null})}}$ (Kraft et al., 2011). This standardized metric describes the degree of departure from random distribution in spatial distribution of empirical species while maintaining the SAD $\Phi(n|S,N)$.

We now use the general formulation of β_P as defined by Equation 1 to derive the null β for random distribution of species in the metacommunity consisting of M local communities. The probability for a randomly distributed species with abundance n being absent from a local community is $\Pi(0|n,M) = \left(1 - \frac{1}{M}\right)^n$. What we need to determine is which SAD should be used in Equation 1. The empirical SAD was used in Kraft et al. (2011). It is well established in theory that SAD of metacommunity follows log-series distribution. This is a principal result of the neutral theory of ecology (Hubbell, 2001; Volkov et al., 2003) and is also predicted by the maximum entropy theory of ecology (Harte, 2011; Pueyo et al., 2007). In practice, log-series together with lognormal distribution have been repeatedly shown to be the two models fitting empirical data best (McGill et al., 2007) and recent meta-analyses even show log-series outperforms other models (Baldridge et al., 2016; White et al., 2012). In this study, we adopt the theoretical log-series SAD for metacommunity to derive β_{null} and its variance. We also tested the use of lognormal distribution by simulation but that does not change our main results about the scaling of β-deviation (see Appendix S2).

For a log-series SAD, $\Phi(n|S,N) = -\frac{1}{\ln(1-p)}\frac{p^n}{n}$, the only parameter $p = e^{-\lambda}$ is fully determined by the state variables N and S: $\frac{N}{S} = \frac{\sum_{n=1}^{N} e^{-\lambda n}}{\sum_{n=1}^{N} e^{-\lambda n}/n}$ (Harte, 2011). Based on this SAD model and the binomial model for random species spatial distribution (Barton & David, 1959; He & Reed, 2006), we derive the following analytical version of the null model (see Appendix S1 for mathematical details):

$$\beta_{null} = \frac{\ln\left(1 - p\left(1 - \frac{1}{M}\right)\right)}{\ln(1-p)} \approx \frac{\ln\left(\frac{M}{1+\lambda M}\right)}{\ln(1/\lambda)}, \quad (2a)$$

$$\begin{aligned} Var_{\Pi}(\beta_{null}) &= \frac{1}{SM\ln(1-p)}\left[\ln\left(\frac{M(1-p)+p}{M(1-p)+2p}\right)\right.\\ &\left.-M\ln\left(1 - \frac{1}{M(M(1-p)/p+2)}\right)\right]\\ &\approx \frac{1}{SM\ln(1/\lambda)}\left[\ln\left(\frac{2+\lambda M}{1+\lambda M}\right) - \frac{1}{2+\lambda M}\right]. \end{aligned} \quad (2b)$$

We note that the approximations are valid under conditions $M \gg 1$ and $\lambda \ll 1$ which are expected for real applications (Harte, 2011, p. 150). This is because in real applications the size of a local community is usually much smaller than the size of the metacommunity (M = metacommunity size/local-community size). However, we

would recommend using the exact formulas in real applications since they impose no computing challenge. However, the approximations offer analytical simplicity and we will use them to derive the scaling relationship between β-deviation and the sampling effort (see below).

From equation 2, it is clear that the expectation and variance of β_{null} under the spatial random distribution are fully determined by the parameter of the log-series SAD (p), the number of local communities (M), and the total number of species (S). The property that the parameter p is fully determined by S and N allows β_{null} to be parameterized using only the three state variables N, S, and M. This means we can compare β-diversity calculated from occurrence data with the null model as long as the metacommunity size N is known (or can be estimated). This offers a method of testing β-diversity in cases where abundances of individual species are not available. The resultant analytical β-deviation is obtained by substituting equation 2 into $\beta_{dev} = \frac{\beta_{obs} - \beta_{null}}{\sqrt{Var_{\Pi}(\beta_{null})}}$.

The variance in equation 2b is the expected conditional variance due to species spatial distribution, that is, $Var_{\Pi}(\beta_{null}) = E_{\Phi}(Var_{\Pi}(\beta|\Phi))$. No variation due to SAD arises because in Kraft et al.'s randomization approach the empirical SAD is used. However, in scientific inference, the empirical SAD ought to be considered as a sample of the underlying theoretical log-series SAD. It then becomes clear that the sampling procedure introduces additional variance in the expected β_{null}. According to the law of total variance, this second variance can be written as (Appendix S1):

$$\begin{aligned} Var_{\Phi}(\beta_{null}) &= Var_{\Phi}(E_{\Pi}(\beta|\Phi))\\ &= \frac{1}{S\ln(1-p)}\left[\ln\left(1 - p\left(1 - \frac{1}{M}\right)^2\right)\right.\\ &\left.-\frac{1}{\ln(1-p)}\ln^2\left(1 - p\left(1 - \frac{1}{M}\right)\right)\right], \end{aligned}$$

where the subscript Φ indicates that the variation is due to SAD. This leads naturally to a quantification of the variance of β-deviation due to the metacommunity SAD:

$$\begin{aligned} Var_{\Phi}(\beta_{dev}) &= \frac{Var_{\Phi}(\beta_{null})}{Var_{\Pi}(\beta_{null})}\\ &= \frac{M\left[\ln\left(1 - p\left(1 - \frac{1}{M}\right)^2\right) - \frac{1}{\ln(1-p)}\ln^2\left(1 - p\left(1 - \frac{1}{M}\right)\right)\right]}{\ln\left(\frac{M(1-p)+p}{M(1-p)+2p}\right) - M\ln\left(1 - \frac{1}{M(M(1-p)/p+2)}\right)}. \end{aligned} \quad (3)$$

2.2 | β-diversity under aggregated spatial distribution of species

In reality, species almost always show aggregated instead of random spatial distribution due to a variety of ecological processes such as dispersal limitation, habitat filtering, and local competition. It has long been recognized that spatial aggregation of empirical species can be well modelled by the negative binomial distribution (NBD, Boswell & Patil, 1970; Pielou, 1977). Using the same log-series SAD model as before and the NBD for aggregated species spatial distribution, we can derive a prediction of β_P

(Barton & David, 1959; He & Reed, 2006; see Appendix S1 for the derivation):

$$\beta_{NBD} = \frac{-1}{\ln(1-p)} \sum_{n=1}^{\infty} \frac{p^n}{n} \left(1 + \frac{n}{Mk}\right)^{-k} \approx \frac{\ln\left(\frac{M}{1+\lambda M}\right) + C(k, \lambda M)}{\ln(1/\lambda)}, \quad (4a)$$

$$\text{Var}(\beta_{NBD}) = \left[\frac{1}{SM\ln(1-p)}\left(\ln\frac{1-pe^{-\frac{1}{M}}}{1-pe^{-\frac{2}{M}}} + \frac{pe^{-\frac{2}{M}}}{M\left(1-pe^{-\frac{2}{M}}\right)}\right)\right]$$
$$+ \left[\frac{-1}{S\ln(1-p)}\left(\sum_{n=1}^{\infty}\left(1+\frac{n}{Mk}\right)^{-2k}\frac{p^n}{n} - \frac{-1}{\ln(1-p)}\left(\sum_{n=1}^{\infty}\left(1+\frac{n}{Mk}\right)^{-k}\frac{p^n}{n}\right)^2\right)\right].$$
$$(4b)$$

where k (>0) is an aggregation parameter with smaller values indicating more aggregated distribution,

$$C(k, \lambda M) = \sum_{n=1}^{\infty}\left[\left(1+\frac{n}{Mk}\right)^{-k}\frac{e^{-\lambda n}}{n}\right] - \ln\left(\frac{M}{1+\lambda M}\right)$$
$$\approx \frac{\ln\left(\frac{1}{\lambda M}+1\right)}{1+(5.07-0.44\ln(\lambda M))k}$$

is a function fully determined by k and λM (i.e. a constant for given values of k and λM; see Appendix S1). Like equation 2, the approximation here is valid under conditions $M \gg 1$ and $\lambda \ll 1$. Note that the variance in equation 4b includes both the variation due to species spatial distribution (the first square bracket) and the variation due to SAD (the second square bracket). This is different from the null model (2), where only the variance due to random species spatial distribution occurs.

Model (4) provides a framework to predict β-diversity from the three state variables N, S, and M and the information about species spatial aggregation (k). Different from the null β of model (2), model (4) incorporates realistic spatial distribution of species and is expected to describe β-diversity of empirical communities (see Section 3 below for confirmation). The realism of model (4) allows us to explore various properties of β-diversity in relation to spatial distribution of species. As an example, below we derive a scaling relationship between β-deviation and sampling effort, to address a major criticism of the using of β-deviation. While model (4) is useful in many situations as stated above, we note that it is not needed if the purpose is to calculate and compare β-deviation, as is done by Kraft et al. (2011).

2.3 | Scaling of β-deviation with sampling effort

Substitute the approximate versions of equations 2 and 4a into Kraft et al.'s $\beta_{dev} = \frac{\beta_{obs} - \beta_{null}}{\sqrt{\text{Var}_\Pi(\beta_{null})}}$ (replace β_{obs} with β_{NBD}), we have:

$$\beta_{dev} = \frac{C(k, \lambda M)}{\sqrt{\ln\left(\frac{2+\lambda M}{1+\lambda M}\right) - \frac{1}{2+\lambda M}}}\sqrt{\frac{SM}{\ln\left(\frac{1}{\lambda}\right)}}. \quad (5a)$$

Until now all the derivations in the above assume that the metacommunity is fully sampled. That is, all the M local communities (that

comprise the metacommunity) are sampled. In real applications, we can only sample a fraction of a metacommunity and sampling intensity also changes from study to study. This sampling incompleteness can have consequences. As shown in Bennett and Gilbert (2016), β-deviation is subject to the effect of sampling effort. Considering Equation 5a, when only m (out of M) local communities are sampled, the parameter k will not change but both $S(m)$ and $\lambda(m)$ will change with m. Nevertheless, numerical analyses suggest that $\lambda(m)m$ and $S(m)/\ln(1/\lambda(m))$ are nearly constant if $M \gg 1$ (see also Harte, 2011, p. 243). This corresponds to a logarithmic sample-based species accumulation curve (Harte, 2011, p. 169). Hence, for a sample of m local communities, the complicated term

$$\frac{C(k, \lambda m)}{\sqrt{\ln\left(\frac{2+\lambda m}{1+\lambda m}\right) - \frac{1}{2+\lambda m}}}\sqrt{\frac{S}{\ln\left(\frac{1}{\lambda}\right)}}$$

in Equation 5a becomes a constant independent of m. The relationship between β-deviation and the sampling effort (m) turns out to be a simple power law with scaling exponent of 0.5:

$$\beta_{dev} \propto \sqrt{m}. \quad (5b)$$

3 | EMPIRICAL TEST OF THE MODELS

3.1 | Data

We used several well-studied datasets from the literature to evaluate the analytical results derived above. Alwyn Gentry's forest transect data (Phillips & Miller, 2002) were used to confirm the performance of our analytical models (2) and (4). This dataset has been previously analysed using the β-deviation (Kraft et al., 2011; Qian et al., 2013). The data consist of 226 plots distributed worldwide, of which 198 are 'standard' plots. Each plot is considered as a metacommunity (Kraft et al., 2011; Qian et al., 2013) that contains ten 2 m × 50 m transects (each transect is considered as a local community) where all plants with a stem diameter ≥2.5 cm were measured and identified to species or morphospecies. Following the previous studies, we used the 198 standard plots in this study. For each metacommunity, we calculated the observed β-diversity (β_{obs}) and the expected β-diversity (β_{null}), variance ($\text{Var}_\Pi(\beta_{null})$), and β-deviation (β_{dev}) for the spatial random hypothesis (equation 2). We also computed β_{dev} following the randomization approach of Kraft et al. (2011). We then plotted these two β_{dev} to evaluate consistency between our analytical null model and the original randomization approach. These two β_{dev} were also plotted versus absolute latitude for comparison. Our analytical models assumed log-series SAD for the metacommunities. To test this assumption and to evaluate potential consequences due to violations of this assumption, a Kolmogorov–Smirnov goodness-of-fit test was performed for each of the metacommunities. In statistics, the Kolmogorov–Smirnov test calculates a distance between the empirical distribution function of a sample (observed species abundances in our case) and the cumulative distribution function of the reference distribution (log-series

distribution in our case) and tests the null hypothesis that this distance is not statistically different from zero. Because log-series is a discrete distribution, we performed the test using the revised ks.test function from R package dgof (Arnold & Emerson, 2011).

With respect to the NBD model (equation 4), we tested its performance by plotting the predicted versus observed β-diversities. Beyond the three state variables (i.e. N, S and M) that are known for each metacommunity, model (4) requires an additional aggregation parameter k. We estimated it for each metacommunity by maximizing the log-likelihood function (He et al., 2002): $l = \sum_{i=1}^{S} [o_i \ln (E(o_i|n_i)) + (M - o_i) \ln (1 - E(o_i|n_i))]$ for the occupancy model $E(o_i|n_i) = M - M \left(1 + \frac{n_i}{Mk}\right)^{-k}$, where o_i is the number of local communities where a species with n_i individuals is present (i.e. the occupancy). The other notations are the same as in previous sections.

To test the prediction about the power-law scaling of β-deviation with sampling effort, we extracted data on β-deviation and sample size from the four different datasets reported in Bennett and Gilbert (2016). The four datasets are as follows: (a) plants in 605 1-m² plots collected from meadow patches in the Garry Oak Ecosystem of southern British Columbia and northern Washington State; (b) plants in 110 1-m² plots collected in an abandoned field at the Koffler Scientific Reserve in southern Ontario, Canada; (c) understory plants in 85 50-m² forest plots collected from Mount St. Hilaire, near Montreal, Canada and (d) diatoms in surficial sediments sampled from 492 North American lakes. More details about these datasets can be found in Bennett and Gilbert (2016) and the references therein. The extracted β-deviation and sample size were plotted on a log–log graph and were visually compared against the theoretical power law of Equation 5b.

3.2 | Test results

The empirical tests shown in Figure 1 confirm that our theoretical null model (equation 2) for the random species spatial distribution
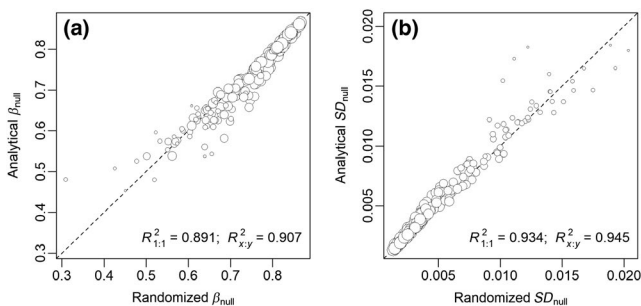

FIGURE 1 β-diversity of Gentry's 198 forest plots under the null model of random spatial distribution of species. (a) Relationship between the expected β-diversity calculated using equation 2a and the randomization approach of Kraft et al. (2011). (b) Relationship between standard deviation of β_{null} calculated using equation 2b and the randomization approach. Sizes of the points are proportional to the values of logarithm of the number of species. The dashed lines are 1:1 line. $R^2_{1:1}$ describes the goodness-of-fit of the 1:1 line to the data, while $R^2_{x:y}$ is the R^2 between x and y
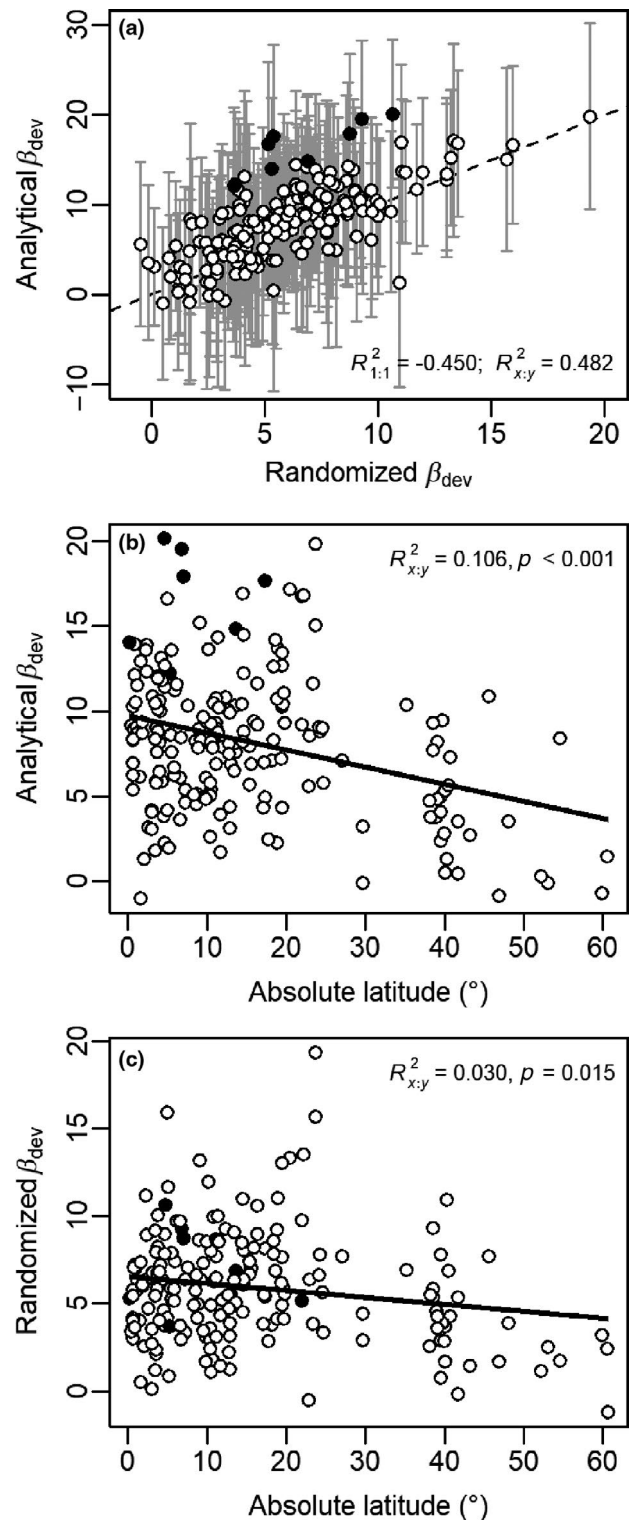

FIGURE 2 β-deviation for Gentry's 198 forest plots. (a) Relationship between β-deviation calculated using the analytical equation 2 and the randomization approach. Error bars represent 1.96 SD of β-deviation due to species abundance distribution (Equation 3). The dashed line is 1:1 line. $R^2_{1:1}$ describes how well the 1:1 line fits the data. (b) Relationship of β-deviation with latitude computed using the analytical solution, and (c) the randomization approach. The solid lines in (b, c) show ordinary least squares regression. $R^2_{x:y}$ is the R^2 between x and y. Filled points highlight the nine plots where the two β-deviations are significantly different

accurately implements the randomization procedure of Kraft et al. (2011). The analytical null model β-deviation has an appreciable but weaker correlation with the β-deviation computed from Kraft et al.'s randomization process ($R^2 > 0.47$, Figure 2a). Only nine out of Gentry's 198 metacommunities show significant difference between the two versions of β-deviation. Both methods are consistent in showing the decreasing latitudinal gradients of β-deviation, but the gradient is appreciably somewhat stronger using the analytical β-deviation (Figure 2b,c). The Kolmogorov–Smirnov test shows that the log-series describes the metacommunity SAD very well ($p > 0.05$) for 180 out of the 198 plots (Figure 3). All the nine 'outlier' plots identified in Figure 2 are among the few that have a poor fit of log-series ($p < 0.05$; Figure 3b).

The NBD β-diversity (equation 4) performs very well in predicting empirical β-diversity ($R^2 = 0.88$, Figure 4). Again, nine out

FIGURE 4 Relationship between predicted β-diversity using model (4) and observed β-diversity for Gentry's 198 forest plots. The grey dashed error bars represent 1.96 SD due to both species abundance distribution and species-level spatial distribution, with the black solid portions due to species-level spatial distribution only. The dashed line is the 1:1 line. Filled points represent the nine plots where the observed β-diversity differ significantly from the prediction. $R^2_{1:1}$ describes the goodness-of-fit of the 1:1 line to the data, while $R^2_{x:y}$ is the R^2 between the two β-diversity
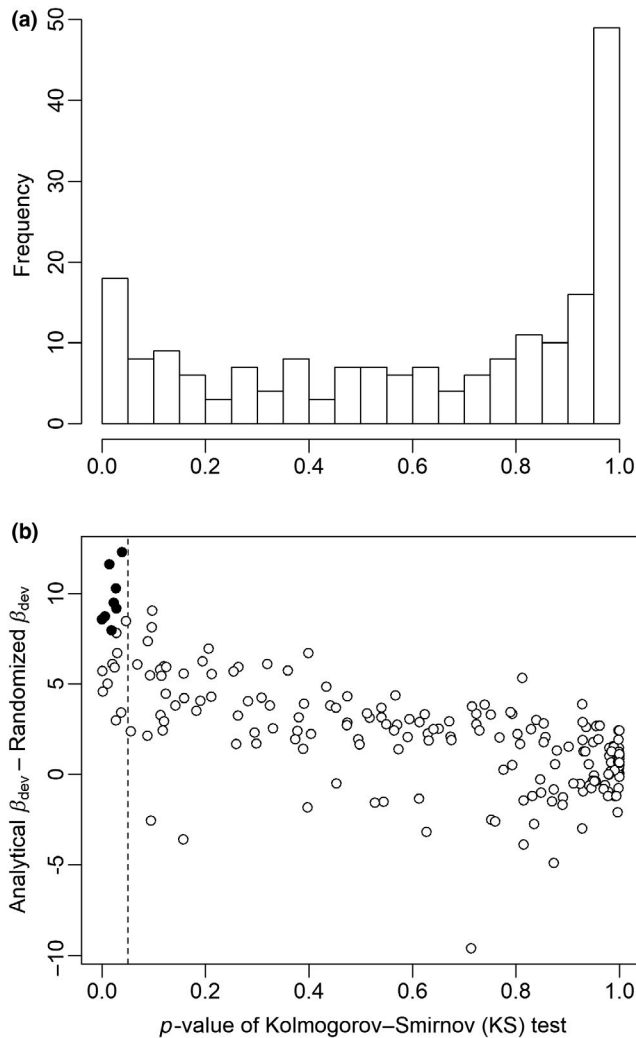

FIGURE 3 Goodness-of-fits of the log-series SAD to Gentry's data. (a) Histogram of p value of the Kolmogorov–Smirnov (KS) test for the log-series SAD model. (b) Relationship of the difference between the analytical and randomization β-deviation with p value of the KS test. Filled points in (b) are plots where the two β-deviations are significantly different (same as in Figure 2). The dashed vertical line indicates $p = 0.05$
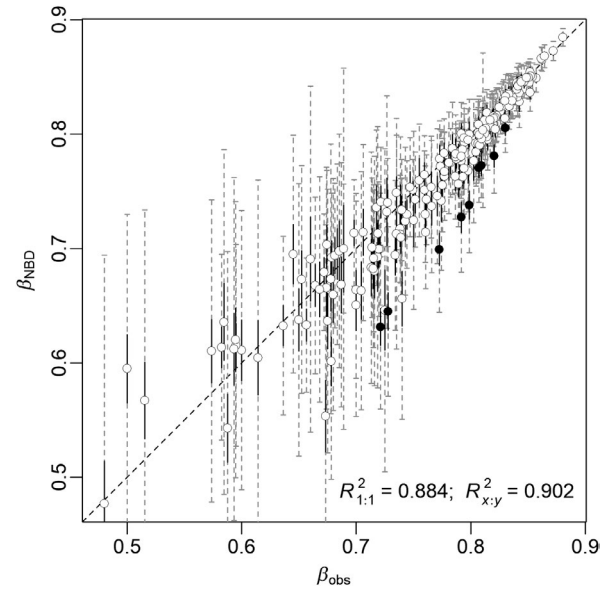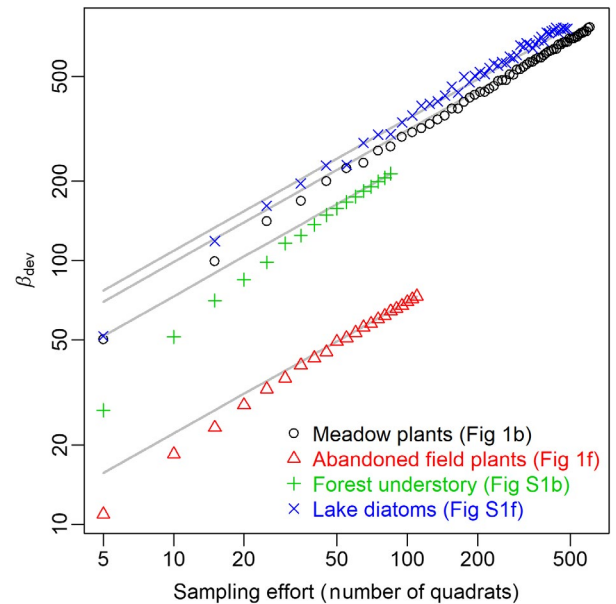

FIGURE 5 Relationships of β-deviation with sampling effort for the four different datasets from Bennett and Gilbert (2016). The data points are extracted from published figures in Bennett and Gilbert (2016) and their corresponding figure numbers are shown in the legend. The grey lines are the derived power-law with scaling exponent of 0.5 (Equation 5b). The positions of the power-law lines are determined by the data points with the largest sampling effort for each dataset, that is, each line goes through the most top-right data point in each corresponding dataset

of Gentry's 198 metacommunities show significant departure from the prediction (note seven of those nine are the same 'outlier' plots as in Figure 2). The empirical relationships between β-deviation and sampling effort employed by Bennett and Gilbert (2016) to argue against the use of β-deviation follow our theoretical power-law scaling Equation 5b very well, as long as the sample size is not too small (i.e. $m > 30$, Figure 5).

4 | DISCUSSION

In this paper, we have developed analytical β-diversity for random and aggregated spatial distribution. The random β-diversity provides a null model for standardizing empirical β-diversity which would otherwise have to be implemented by randomization (Kraft et al., 2011). The simulation and empirical tests show that the analytical β-diversity models perform very well in comparison to the original randomization process of Kraft et al. (2011) (Figures 1 and 2). The practical difference between the theoretical null model and the original randomization approach can be simply explained by the use of different SADs, which is one of the two differences between our null model and the original randomization approach. The second difference is that we do not preserve the abundance in each local community. Previous study has demonstrated that this second difference is neglectable (see figure 3a in Xu et al., 2015). Our null model uses a theoretical distribution (could be log-series, or lognormal; see Appendix S2) while the randomization approach uses the empirical SAD that could be subject to sample size and sampling variation. The variation arising from sampling SAD is quantified by Equation 3. Recognizing the existence of this variation is important because in real applications almost all empirical SADs are based on sampling data and themselves are not true metacommunities. This sampling problem is one of the major criticisms of Kraft et al.'s (2011) original analyses of Gentry's data that the empirical SAD does not represent true metacommunity (Tuomisto & Ruokolainen, 2012). From this point of view, our analytical β-diversity based on the theoretically justified log-series SAD of metacommunity provides a desirable solution. That said, it could happen that the SAD of a real metacommunity does not follow log-series distribution. In this case, our beta-deviation is biased. Fortunately, beta-deviation in general is quite robust to the variation in SAD as we tested by assuming lognormal SAD distribution (Appendix S2).

Unlike the randomization process, our β-diversity models do not require abundances of individual species but only the data on metacommunity size (N) and total richness (S) (for null β-diversity equation 2) or N and S plus spatial distribution parameter k (for the NBD β-diversity equation 4). More significantly, these analytical results reveal the dependence of β-diversity on N, S, species spatial pattern and sampling effort and thus offer unambiguous interpretation on β-deviation which is otherwise not obvious and has been a subject of debate (Bennett & Gilbert, 2016; Kraft et al., 2011; Qian et al., 2013; Ulrich et al., 2017). In the following discussion, we will address and clarify the controversies surrounding the use of the randomized

β-deviation. Our discussion will make it clear that β-deviation is an important and solid diversity measure for testing community assembly mechanisms and the much-criticized sampling effect on β-deviation (Bennett & Gilbert, 2016) can be easily dealt with using the simple scaling law of equation (5).

The misinterpretation of β-deviation as 'a standard effect size of β-diversity deviations from a null model that corrects for γ dependence' (Kraft et al., 2011) has raised much concern about the null model because β-deviation is not independent of γ diversity (Bennett & Gilbert, 2016; Qian et al., 2013; Ulrich et al., 2017, 2018). The fact that the null β-diversity of equation 2 is derived from the assumption of random spatial distribution of species and the consistence between our analytical model and the randomization approach (Figures 1 and 2) clearly indicate that the null model of Kraft et al. (2011) is for randomizing spatial distribution of species given a metacommunity with fixed SAD. Hence, the β-deviation is a metric measuring the effect size of non-random species spatial distribution on β-diversity for a given SAD. It is not for correcting for the effect of γ diversity as originally interpreted, neither should it be interpreted as correcting for SAD. The reason that information of SAD is included in β-deviation is because the randomization procedure (and the analytical model (2)) is implemented under a given metacommunity SAD. Because of that, if metacommunity SAD changes, beta-deviation will also change. Indeed, Equation 5a shows that the magnitude of β-deviation is collectively dependent on aggregation parameter k, metacommunity size N, total richness S and the number of local communities M. When compared among regions, due to the difference in SAD (or N and S), communities with the same k could have different β-deviation, and vice versa (see Figure S3 of Appendix S3).

There are a number of alternative approaches used in the literature that are closely related with the null model proposed here, but they all suffer from similar problems. Qian et al. (2013) and Xu et al. (2015) analysed the raw (i.e. $\beta_{obs} - \beta_{null}$), instead of the standardized, β-deviation. From our models (2) and (4), it is clear that the raw β-deviation is also dependent on γ, thus not serving as a correction for the γ-dependence. A more fundamental problem with the raw β-deviation is that it is not a measure of effect size and thus is of little use for comparing the effect of species aggregation on β-diversity. Ulrich et al. (2018) proposed a 'fixed-fixed' null model that preserves both row and column sums of the community matrix to correct for the γ-dependence. However, the 'fixed-fixed' null model does not have an analytical solution and its ecological interpretation is not clear.

Another major criticism on β-deviation is that it is sampling effort dependent, meaning that different sample sizes from the same system can lead to drastically different β-deviation (Bennett & Gilbert, 2016). This problem makes the method not useful for comparing different communities unless sample size is standardized (Xing & He, 2019; Zhang et al., 2020). The scaling relationship between β-deviation and sampling effort (Equation 5b) derived in this study thaws this criticism and offers a simple solution. To compare β-deviation with different sample size (m), one only needs to standardize the effect of sample size by dividing the β-deviation

by $\sqrt{m}$. This analytical result is a generalization of the simulation study of Xing and He (2019) (see Appendix S2 there). It is remarkable because the simulation study only justifies comparisons of β-deviation under constant sampling effort. It does not provide the standardization method revealed by the analytical study presented here. The standardization method revealed here is critical for future comparison study or meta-analyses of published results from different studies. However, it is important to note that in applications the scaling relationship does not behave as well for small sample sizes (e.g. $m < 30$). This may reflect the fact that small sample size leads to underestimation of the difference in β-diversity of multiple communities, as shown by M-community Jaccard β (Marion et al., 2017), which differs from β_{P} only by a factor of $M/(M - 1)$. To our knowledge, no correction for the small sample effect on these M-community β measures is currently available. Hence at this stage we can only recommend sampling more local communities to calculate M-community β and β-deviation. The small sample effect warrants further research. We also note that both β and β-deviation are dependent on the sampling grain and extent. This study deals with neither of these two aspects of scale effect. In application, the same grain size and similar spatial extent are required when β or β-deviation is compared among different regions.

The null model of β-diversity and the resultant β-deviation provide a promising approach for inferring community assembly rules. The derivation of the analytical version of the null model in this study clarifies the interpretation of the model and addresses the problem that the β-deviation is subject to the effect of sampling effort. Beyond these apparent benefits, our analytical model also extends the application of β-deviation to occurrence data. The requirement for species abundance data by the original randomization null model can seriously limit its application in systems of large spatial extent. Data on abundance of individual species are no longer needed for the theoretical model derived here as long as the size of the metacommunity is known. This affords the use of the theoretical model to systems of large extent and can take advantage of the occurrence data available for many taxa (e.g. birds, amphibians, mammals, and plants; Gaston et al., 2007; McKnight et al., 2007; Buckley & Jetz, 2008; Maitner et al., 2018). Coupling with these data, our analytical model offers an approach to explore the global pattern and underlying drivers of β-diversity.

When applied to occurrence data, our analytical model needs four inputs to calculate β-deviation: total abundance (N), total richness (S), number of local communities (M) and the observed beta diversity (β_{obs}). S, M and β_{obs} can be directly obtained from the occurrence data. N needs to be determined separately. Although not trivial, we argue that estimating N is much easier than obtaining abundances for each species. This is because not like S, N is additive when small areas are merged to form a larger one and thus scales linearly with area (He et al., 2002; Preston, 1962). An example of how to map the global distribution of N for trees is shown by Crowther et al. (2015). In this kind of applications, S and

N are parameters of the metacommunity and the uncertainty in their estimation could have consequence on the calculation of β-deviation. Fortunately, simulation shows that the errors of β-deviation caused by the variation in S and/or N are on average much smaller than the variation in S and/or N themselves (Figure S4 in Appendix S3). The relative error in β-deviation tends to be larger for metacommunities with more random species spatial patterns (i.e. when β-deviation is small).

One potential limitation of this study is the assumption of log-series for metacommunity SAD. In situations where the true SAD does not follow log-series, our model could elevate type I error (i.e. the β-deviation tends to reject a true hypothesis that communities are randomly assembled in space). That is, deviation of the true SAD from log-series will weaken the accuracy of predictions from our models (Figure 3). As for the power-law scaling of β-deviation with sampling effort, both our simulation (Appendix S2) and the empirical data from various sources (Figure 5) show that it is robust to variations in SAD. The ecological interpretation of β-deviation revealed by our analytical analyses does not suffer from the SAD assumption either. Hence, among the three major implications of our analytical results, only the application in estimating β-deviation from occurrence data could potentially be affected by the assumption. Because log-series for metacommunity SAD is theoretically justified (Harte, 2011; Hubbell, 2001; Pueyo et al., 2007) and well supported in numerous empirical tests (Baldridge et al., 2016; White et al., 2012), our theoretical models are of strong practical significance, as exemplified by the majority of Gentry's forest plots (Figures 1–4). However, erring on the side of caution, we suggest avoiding using our model if the SAD of the study system is far from log-series.

AUTHORS' CONTRIBUTIONS

D.X. and F.H. conceived the study; D.X. derived the analytical results and analysed the data; D.X. and F.H. wrote the paper.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/2041-210X.13531.

DATA AVAILABILITY STATEMENT

Data and codes supporting the results can be accessed at https://github.com/DLXING/beta_analytical and at Zenodo https://doi.org/10.5281/zenodo.4270201 (Xing & He, 2020).

ORCID

Dingliang Xing 🆔 https://orcid.org/0000-0001-9133-6674

REFERENCES

Arnold, T. B., & Emerson, J. W. (2011). Nonparametric goodness-of-fit tests for discrete null distributions. *R Journal*, *3*(2), 34–39. https://doi.org/10.32614/RJ-2011-016

Baldridge, E., Harris, D. J., Xiao, X., & White, E. P. (2016). An extensive comparison of species-abundance distribution models. *PeerJ*, *4*, e2823. https://doi.org/10.7717/peerj.2823

Barton, D. E., & David, F. N. (1959). Contagious occupancy. *Journal of the Royal Statistical Society. Series B (Methodological)*, *21*(1), 120–133. https://doi.org/10.1111/j.2517-6161.1959.tb00320.x

Bennett, J. R., & Gilbert, B. (2016). Contrasting beta diversity among regions: How do classical and multivariate approaches compare? *Global Ecology and Biogeography*, *25*(3), 368–377. https://doi.org/10.1111/geb.12413

Boswell, M. T., & Patil, G. P. (1970). Chance mechanisms generating the negative binomial distributions. In G. P. Patil (Ed.), *Random counts in models and structures* (pp. 3–22). Pennsylvania State University Press.

Buckley, L. B., & Jetz, W. (2008). Linking global turnover of species and environments. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(46), 17836–17841. https://doi.org/10.1073/pnas.0803524105

Chao, A., & Chiu, C.-H. (2016). Bridging the variance and diversity decomposition approaches to beta diversity via similarity and differentiation measures. *Methods in Ecology and Evolution*, *7*(8), 919–928. https://doi.org/10.1111/2041-210X.12551

Crist, T. O., Veech, J. A., Gering, J. C., & Summerville, K. S. (2003). Partitioning species diversity across landscapes and regions: A hierarchical analysis of α, β, and γ diversity. *The American Naturalist*, *162*(6), 734–743. https://doi.org/10.1086/378901

Crowther, T. W., Glick, H. B., Covey, K. R., Bettigole, C., Maynard, D. S., Thomas, S. M., Smith, J. R., Hintler, G., Duguid, M. C., Amatulli, G., Tuanmu, M.-N., Jetz, W., Salas, C., Stam, C., Piotto, D., Tavani, R., Green, S., Bruce, G., Williams, S. J., … Bradford, M. A. (2015). Mapping tree density at a global scale. *Nature*, *525*(7568), 201–205. https://doi.org/10.1038/nature14967

De Cáceres, M., Legendre, P., Valencia, R., Cao, M., Chang, L.-W., Chuyong, G., Condit, R., Hao, Z., Hsieh, C.-F., Hubbell, S., Kenfack, D., Ma, K., Mi, X., Supardi Noor, M. N., Kassim, A. R., Ren, H., Su, S.-H., Sun, I.-F., Thomas, D., … He, F. (2012). The variation of tree beta diversity across a global network of forest plots. *Global Ecology and Biogeography*, *21*(12), 1191–1202. https://doi.org/10.1111/j.1466-8238.2012.00770.x

Gaston, K. J., Davies, R. G., Orme, C. D. L., Olson, V. A., Thomas, G. H., Ding, T.-S., Rasmussen, P. C., Lennon, J. J., Bennett, P. M., Owens, I. P. F., & Blackburn, T. M. (2007). Spatial turnover in the global avifauna. *Proceedings of the Royal Society B: Biological Sciences*, *274*(1618), 1567–1574. https://doi.org/10.1098/rspb.2007.0236

Harte, J. (2011). *Maximum entropy and ecology: A theory of abundance, distribution, and energetics*. Oxford University Press.

He, F., Gaston, K., & Wu, J. (2002). On species occupancy-abundance models. *Écoscience*, *9*(1), 119–126. https://doi.org/10.1080/11956860.2002.11682698

He, F., LaFrankie, J. V., & Song, B. (2002). Scale dependence of tree abundance and richness in a tropical rain forest. *Malaysia. Landscape Ecology*, *17*(6), 559–568.

He, F., & Legendre, P. (2002). Species diversity patterns derived from species–area models. *Ecology*, *83*(5), 1185–1198.

He, F., & Reed, W. (2006). Downscaling abundance from the distribution of species: occupancy theory and applications. In J. Wu, K. B. Jones, H. Li, & O. L. Loucks (Eds.), *Scaling and uncertainty analysis in ecology* (pp. 89–108). Springer, Netherlands. https://doi.org/10.1007/1-4020-4663-4_5

Hubbell, S. P. (2001). *The unified neutral theory of biodiversity and biogeography*. Princeton University Press.

Kraft, N. J. B., Comita, L. S., Chase, J. M., Sanders, N. J., Swenson, N. G., Crist, T. O., Stegen, J. C., Vellend, M., Boyle, B., Anderson, M. J., Cornell, H. V., Davies, K. F., Freestone, A. L., Inouye, B. D., Harrison, S. P., & Myers, J. A. (2011). Disentangling the drivers of β diversity along latitudinal and elevational gradients. *Science*, *333*(6050), 1755–1758. https://doi.org/10.1126/science.1208584

Kreft, H., & Jetz, W. (2010). A framework for delineating biogeographical regions based on species distributions. *Journal of Biogeography*, *37*(11), 2029–2053. https://doi.org/10.1111/j.1365-2699.2010.02375.x

Maitner, B. S., Boyle, B., Casler, N., Condit, R., Donoghue, J., Durán, S. M., Guaderrama, D., Hinchliff, C. E., Jørgensen, P. M., Kraft, N. J. B., McGill, B., Merow, C., Morueta-Holme, N., Peet, R. K., Sandel, B., Schildhauer, M., Smith, S. A., Svenning, J.-C., Thiers, B., … Enquist, B. J. (2018). The bien r package: A tool to access the Botanical Information and Ecology Network (BIEN) database. *Methods in Ecology and Evolution*, *9*(2), 373–379. https://doi.org/10.1111/2041-210X.12861

Marion, Z. H., Fordyce, J. A., & Fitzpatrick, B. M. (2017). Pairwise beta diversity resolves an underappreciated source of confusion in calculating species turnover. *Ecology*, *98*(4), 933–939. https://doi.org/10.1002/ecy.1753

McGill, B. J., Etienne, R. S., Gray, J. S., Alonso, D., Anderson, M. J., Benecha, H. K., Dornelas, M., Enquist, B. J., Green, J. L., He, F., Hurlbert, A. H., Magurran, A. E., Marquet, P. A., Maurer, B. A., Ostling, A., Soykan, C. U., Ugland, K. I., & White, E. P. (2007). Species abundance distributions: Moving beyond single prediction theories to integration within an ecological framework. *Ecology Letters*, *10*(10), 995–1015. https://doi.org/10.1111/j.1461-0248.2007.01094.x

McKnight, M. W., White, P. S., McDonald, R. I., Lamoreux, J. F., Sechrest, W., Ridgely, R. S., & Stuart, S. N. (2007). Putting beta-diversity on the map: Broad-scale congruence and coincidence in the extremes. *PLoS Biology*, *5*(10), e272. https://doi.org/10.1371/journal.pbio.0050272

Morlon, H., Chuyong, G., Condit, R., Hubbell, S., Kenfack, D., Thomas, D., Valencia, R., & Green, J. L. (2008). A general framework for the distance–decay of similarity in ecological communities. *Ecology Letters*, *11*(9), 904–917. https://doi.org/10.1111/j.1461-0248.2008.01202.x

Myers, J. A., Chase, J. M., Jiménez, I., Jørgensen, P. M., Araujo-Murakami, A., Paniagua-Zambrana, N., & Seidel, R. (2013). Beta-diversity in temperate and tropical forests reflects dissimilar mechanisms of community assembly. *Ecology Letters*, *16*(2), 151–157. https://doi.org/10.1111/ele.12021

Phillips, O., & Miller, J. S. (2002). *Global patterns of plant diversity: Alwyn H. Gentry's forest transect data set* (Vol. 89). Missouri Botanical Garden Press.

Pielou, E. C. (1977). *Mathematical ecology*. John Wiley & Sons, Ltd. Retrieved from http://193.190.8.15/dpm/handle/0/511

Plotkin, J. B., & Muller-Landau, H. C. (2002). Sampling the species composition of a landscape. *Ecology*, *83*(12), 3344–3356.

Preston, F. W. (1962). The canonical distribution of commonness and rarity: Part I. *Ecology*, *43*(2), 185–215. https://doi.org/10.2307/1931976

Pueyo, S., He, F., & Zillio, T. (2007). The maximum entropy formalism and the idiosyncratic theory of biodiversity. *Ecology Letters*, *10*(11), 1017–1028. https://doi.org/10.1111/j.1461-0248.2007.01096.x

Qian, H., Chen, S., Mao, L., & Ouyang, Z. (2013). Drivers of β-diversity along latitudinal gradients revisited. *Global Ecology and Biogeography*, *22*(6), 659–670. https://doi.org/10.1111/geb.12020

Tuomisto, H. (2010). A diversity of beta diversities: Straightening up a concept gone awry. Part 1. Defining beta diversity as a function of alpha and gamma diversity. *Ecography*, *33*(1), 2–22. https://doi.org/10.1111/j.1600-0587.2009.05880.x

Tuomisto, H., & Ruokolainen, K. (2012). Comment on 'Disentangling the drivers of β diversity along latitudinal and elevational gradients'. *Science*, *335*(6076), 1573. https://doi.org/10.1126/science.1216393

Ulrich, W., Baselga, A., Kusumoto, B., Shiono, T., Tuomisto, H., & Kubota, Y. (2017). The tangled link between β- and γ-diversity: A Narcissus effect weakens statistical inferences in null model analyses of diversity patterns. *Global Ecology and Biogeography*, *26*(1), 1–5. https://doi.org/10.1111/geb.12527

Ulrich, W., Kubota, Y., Kusumoto, B., Baselga, A., Tuomisto, H., & Gotelli, N. J. (2018). Species richness correlates of raw and standardized co-occurrence metrics. *Global Ecology and Biogeography*, *27*(4), 395–399. https://doi.org/10.1111/geb.12711

Vannette, R. L., & Fukami, T. (2017). Dispersal enhances beta diversity in nectar microbes. *Ecology Letters*, *20*(7), 901–910. https://doi.org/10.1111/ele.12787

Volkov, I., Banavar, J. R., Hubbell, S. P., & Maritan, A. (2003). Neutral theory and relative species abundance in ecology. *Nature*, *424*(6952), 1035–1037. https://doi.org/10.1038/nature01883

White, E. P., Thibault, K. M., & Xiao, X. (2012). Characterizing species abundance distributions across taxa and ecosystems using a simple maximum entropy model. *Ecology*, *93*(8), 1772–1778. https://doi.org/10.1890/11-2177.1

Xing, D., & He, F. (2019). Environmental filtering explains a U-shape latitudinal pattern in regional β-deviation for eastern North American trees. *Ecology Letters*, *22*(2), 284–291. https://doi.org/10.1111/ele.13188

Xing, D., & He, F. (2020). Data and code from 'Analytical models for β-diversity and the power-law scaling of β-deviation'. *Zenodo*, https://doi.org/10.5281/zenodo.4270202

Xu, W., Chen, G., Liu, C., & Ma, K. (2015). Latitudinal differences in species abundance distributions, rather than spatial aggregation, explain beta-diversity along latitudinal gradients. *Global Ecology and Biogeography*, *24*(10), 1170–1180. https://doi.org/10.1111/geb.12331

Zhang, C., He, F., Zhang, Z., Zhao, X., & von Gadow, K. (2020). Latitudinal gradients and ecological drivers of β-diversity vary across spatial scales in a temperate forest region. *Global Ecology and Biogeography*, *29*(7), 1257–1264. https://doi.org/10.1111/geb.13101

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Xing D, He F. Analytical models for β-diversity and the power-law scaling of β-deviation. *Methods Ecol Evol.* 2020;00:1–10. https://doi.org/10.1111/2041-210X.13531