A geometric approach to scaling individual distributions to macroecological patterns

Nao Takashina^{1*}, Buntarou Kusumoto², Yasuhiro Kubota³, Evan P. Economo¹

¹*Biodiversity and Biocomplexity Unit, Okinawa Institute of Science and Technology Graduate University, Onna-son, Okinawa 904-0495, Japan*

²*Center for Strategic Research Project, University of the Ryukyus, Nishihara, Okinawa 903-0213, Japan*

³*Faculty of Science, University of the Ryukyus, Nishihara, Okinawa, 903-0213, Japan*

Keywords: Endemic area relationship, relative species abundance, spatially explicit model, species area relationship

Abstract

Understanding macroecological patterns across scales is a central goal of ecology and a key need for conservation biology. Much research has focused on quantifying and understanding macroecological patterns such as the species-area relationship (SAR), the endemic-area relationship (EAR) and relative species abundance curve (RSA). Understanding how these aggregate patterns emerge from underlying spatial pattern at individual level, and how they relate to each other, has both basic and applied relevance. To address this challenge, we develop a novel spatially explicit geometric framework to understand multiple macroecological patterns, including the SAR, EAR, RSA, and their relationships, using theory of point processes. The geometric approach provides a theoretical framework to derive SAR, EAR, and RSA from species range distributions and the pattern of individual distribution patterns therein. From this model, various well-documented macroecological patterns are recovered, including the tri-phasic SAR on a log-log plot with its asymptotic slope, and various RSAs (e.g., Fisher's logseries and the Poisson lognormal distribution). Moreover, this approach can provide new insights such as a single equation describing the RSA at an arbitrary spatial scale, and explicit forms of the EAR with its asymptotic slope. The theory, which links spatial distributions of individuals and species with macroecological patterns, is ambiguous with regards to the mechanism(s) responsible for the statistical properties of individual distributions and species range sizes. However, our approach can be connected to mechanistic models that make such predictions about lower-level patterns and be used to scale them up to aggregate patterns, and therefore is applicable to many ecological questions. We demonstrate an application of the geometric model to scaling issue of beta diversity.

*Electronic address: nao.takashina@gmail.com; Corresponding author

1 Introduction

The problem of pattern and scale is central in ecology, and critical insights into conservation biology emerge as we observe different aspects of ecosystems across scales [1]. The species area relationship (SAR), endemic area relationship (EAR), and relative species abundance (RSA) are such examples characterizing macroecological and community patterns of ecosystem across scales. Each of these patterns has a long history and ample literatures (e.g., [2–6]), and there are still active discussions over fundamental patterns (e.g., [7, 8]). In particular, recent researches have focused on the quantitative investigation of scaling issues of macroecological patterns [9–12]. However, little is known about how individual-level spatial distributions scale up to aggregate patterns such as the SAR, EAR, and RSA (but see [11]).

A model linking individual distributions with aggregate patterns can provide a great applicability to other ecological questions in community ecology and conservation biology. For example, to investigate the scale-dependence of beta diversity [13], one needs species abundance information at finer scales and a consistent scale-change operation must be defined. These macroecological patterns are also critical for implementing effective ecosystem management, given accelerating loss of biodiversity worldwide [14, 15]. For example, the SAR and EAR provide information about how many species will be affected and be lost from the landscape when a certain region is degraded, respectively [16], and the RSA provides more detailed information on community structure such as composition of rare species. Practically, conservation decision making is spatially explicit provided optimal implementation strategy of protected areas with budget constraints [17, 18], and often high degree of spatial information, such as individual distributions and the number of individuals across multiple scales are required (e.g., [19, 20]).

Here, we propose a novel geometric approach to study macroecology that provides a bridge between spatial distribution patterns and emergent macroecological patterns. In this framework, the SAR, EAR, and RSA are derived from the geometry of species ranges and the spatial distribution of conspecific individuals therein. Conceptually similar approaches have been investigated in previous researches without individual distributions [9, 10, 21], and with explicitly integrating individual distributions [11, 22]. Allen and White [21] derived an upper bound of the SAR, by calculating an overlapping area between the sampling region and the randomly placed distribution range without considering individual distributions therein, leading to a biphasic SAR on a log-log plot. Although this model provides the well-known asymptotic slope of 1 at very large sampling scales, it did not capture the sampling process on smaller scales responsible for the triphasic curve. Plotkin et al. [22] applied the Thomas process, a point process model, to generate aggregated individual distribution patterns in tropical forest plots. They independently estimated a dispersal distance and density from spatial distribution of individual trees of each species, and generated superposition of multiple species in a forest. The fitted Thomas process model very precisely recovers SARs up to 50-ha. Grilli et al. [11] also applied a point process model to account for individual distributions. With aggregated communities generated by the Poisson cluster model, they derived the SAR, EAR, and RSA. In doing so, the model requires a neutral assumption and a specific input

of the RSA form in the whole system.

In our framework, the model uses a point field of communities that holds information of all individuals and species distributions. The point field is consistent with scale changes and, therefore, we can derive macroecological patterns across scales consistently without changing ecological parameters used. This property also provides a significant applicability to conservation practices, since the model provides macroecological information on arbitrary spatial scales. Particularly, we assume non-interacting species distribution ranges and individual distributions therein. In derivations of the SAR, EAR, and RSA, our model put a particular emphasis on geometric patterns that the sampling region, distribution ranges, and its individual distribution patterns therein generate. One of the advantages of this approach is that we can discuss in what conditions the model recovers emergent macroecological patterns, and, importantly, any preceding knowledge of these patterns are required unlike the previously developed model [11]. The model is also amenable to incorporate variations in species characteristics, and we can easily discuss the emergent patterns when species vary in ecological characteristics.

The paper contains three major parts. First, we develop a general theory based on our geometric approach to link the SAR, EAR, and RSA to spatial point patterns, without assuming specific spatial patterns. This provides a rather general framework, and any spatial patterns define species and individual distribution pattern can be applied. Second, we derive emergent macroecological patterns by assuming specific point patterns that phenomenologically describe geometric patterns of species and individual distributions therein. We show the model recovers many well-documented macroecological patterns including the tri-phasic SAR with its asymptotic slope 1 on a log-log plot. For RSAs, well-known RSAs such as negative binomial distribution, Fisher's logseries, and the Poisson lognormal are recovered at the small sampling scales. We demonstrate that at intermediate sampling scales, RSAs are described by combinations of two probability distribution functions, and it leads to a left-tailed or negatively-skewed bell shape curves on a logarithmic scale. Notably, these all RSAs are summarized by a single equation, and therefore easy to discuss transition patterns with sampling size. It is worth noting that the structure of these SAR and RSA patterns are often attributed to ecological and evolutionary mechanisms, but our geometric approach still can recover these patterns based on general assumptions about individual and range size distributions. Our model also provides new potential form of EARs with asymptotic slope 1 on a log-log plot that, unlike commonly used definition, directly counts the average number of enclosed species distributions. Third, to demonstrate potential applicability of the geometric model to ecological questions, we address the scale-dependence of beta diversity as an example.

2 General theory

We develop here a general theory of our geometric approach, an attempt to provide a general framework for calculations of the SAR, EAR, and RSA without assuming any specific geometric properties. More specifically, here we do not assume individual distribution patterns,

shapes of the sampling region and the distribution range. However, it is worth noting that the central assumptions of the theory are that we assume no interaction between species and, therefore, species distribution range of each species are chosen independently in the homogeneous environment. Once ecological properties and the sampling scheme are specified, the SAR, EAR, and RSA can be obtained via the general framework. We will discuss some specific situations in Results. Central parameters used throughout the paper are summarized in Table 1.

Table 1: Definition of parameters

Symbol	Parameter
S	Sampling region
$D, (D_i)$	Distribution range of species (i)
$\nu(A)$	Area of A
$\mathbf{x}_c^A$	Geometric center of A
l_A	Minimum distance between the center and boundary of A
L_A	Maximum distance between the center and boundary of A
X	Number of individuals
$N_i(A)$	Number of individuals of species i in A
$N_s(A)$	Number of species in A
λ_i	Intensity of species i
λ_s	Intensity of species
$\mu_i(A)$	Intensity measure of species i in A
$\mu_s(A)$	Intensity measure of species in A

2.1 Basic notation

2.1.1 Intensity and intensity measure

We introduce some basic quantities of ecosystem which will be used in the following discussions. Each of these are generally used in the field of point process, and we borrow some ideas in this paper. Interested readers may refer to literatures in the field (e.g., [23, 24]) or applications to ecological studies (e.g., [11, 19, 20, 22, 25]). Later, we will combine the general theory with point processes to examine specific individual distribution patterns such as random or clustering distributions.

The intensity, λ_i , and the intensity measure, μ_i , of species i are the central quantities of our model characterizing the ecological community: the average density of individuals of species i , and the average number of individuals of species i in region A , respectively. Given the area $\nu(A)$ and the number of individuals of species i , $N_i(A)$ in A , the intensity of species i in A is defined as [24]

$$\lambda_i := \frac{N_i(A)}{\nu(A)}. \quad (1)$$

The intensity measure of species in A is given by multiplying the intensity by area $\nu(A)$

$$\mu_i(A) := \lambda_i \nu(A). \quad (2)$$

The same quantities for species assembly are also defined within this framework. Let $N_s(A)$ be the number of species in A . Then, the intensity and intensity measure in A of the species assembly are defined as $\lambda_s := N_s(A)/\nu(A)$ and $\mu_s := \lambda_s \nu(A)$, respectively. Then the vector $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_s)$ holds the information of community abundance. Let us define the intensity measure of species, $\mu_s(A)$ as

$$\mu_s(A) := \int_A f(\mathbf{x}) d\mathbf{x}, \quad (3)$$

where, $f(\mathbf{x})$ is a point field holding information on probability of all individual locations in two-dimensional space $\mathbf{x} \in \mathbf{R}^2$. Let us assume the homogeneous environment, and then we can put $f(\mathbf{x}) = \lambda_s$, where λ_s is the intensity of species, and we obtain

$$\mu_s = \lambda_s \nu(A). \quad (4)$$

2.1.2 Sampling region and endemic region

In this paper, we refer to S as the *sampling region* which defines the spatial scales discussed. Similarly, we refer to D_i as the distribution range of species i . We often omit the subscript i and call the *distribution range of species* or simply *distribution range* if obvious. Throughout the analysis, the distribution range D is used as an endemic region of species where the individual distributions of the species are restricted within the region. For convenience, yet mathematically not rigorous, we use the symbols " $>$ " and " $<$ " regarding to relative sizes of S and D . For example, in our notations $S > D$ implies that we can always find a shift, without rotating, that satisfies $D \subset S$.

2.2 Species area relationship

Here we derive a method to calculate the species area relationship (SAR). Often it exhibits the tri-phasic features on a log-log plot [3, 5], showing a fast increase on small scales, slowing down the increase on intermediate scales, and again shows a quick increase asymptotically toward to 1 for the slope on very large scales where sampling region exceeds correlation distance of biogeographic process with separate evolutionary histories [5, 6]. As the simplest example, let us first assume a neutral community where biological properties, such as intensity of each species, distribution range, and average dispersal distance are identical between species. However, as we will see below, it is straightforward to incorporate species variations. To calculate the species number in a given area, we use the nonzero probability $p^{nz}(\mathbf{x}_c^S, \mathbf{x})$ that a species at $\mathbf{x}$ with the geometric center of the distribution region $\mathbf{x}_c^D \in S_{out}$ provides at least one individual to the sampling region S with its geometric center $\mathbf{x}_c^S \in S$. S_{out} is the region that if the geometric center of the distribution range of a species D belongs to

Figure 1: Schematic representations of (a) the sampling region and distribution ranges, and typical geometric relationships used in derivations of the (b) SAR, (c) EAR, and (d) RSA. For each panel, S and D are the sampling region and distribution ranges, respectively. S and D have arbitrary shapes in reality, but they are shown by square (black) and ellipse (filled by color) with the geometric center, respectively as an example. S_{out} is the region described with dotted line, where if the geometric center of a distribution range $\mathbf{x}_c^D$ falls in this region, then the probability to overlap S and D is not zero: Namely $P(S \cap D) \neq 0$ provided $\mathbf{x}_c^D \in S_{out}$. S_{core} is the region described by grey line where provided $S > D$ ($S < D$) if the distribution center of a species $\mathbf{x}_c^D$ falls in this regions, then the entire distribution range D (S) is covered by S (D): Namely $P(D \subset S) = 1$ ($P(S \subset D) = 1$) provided $\mathbf{x}_c^D \in S_{core}$. For the derivation of the EAR, only the case $S > D$ is used, but both $S > D$ and $S < D$ are used for the derivation of the RSA. S_{core} is the closed region and typically a subset of the sampling region $S_{core} \subset S$, in which any points on this boundary, $\mathbf{x} \in \partial S_{core}$ holds the following relationship: when $S > D$, $\min\{\|\mathbf{x} - \mathbf{x}'\|; \mathbf{x} \in \partial S, \mathbf{x}' \in \partial S_{core}\} = \max\{\|\mathbf{x}_c^{D_i} - \mathbf{x}\|; \mathbf{x} \in \partial D\}$; when $S < D$, $\min\{\|\mathbf{x}_c^{D_i} - \mathbf{x}\|; \mathbf{x} \in \partial D\} = \max\{\|\mathbf{x} - \mathbf{x}'\|; \mathbf{x} \in \partial S, \mathbf{x}' \in \partial S_{core}\}$. Each region is determined by the minimum distance l_s , l_D and maximum distance L_s , L_D between the sampling center and its boundary. See further explanations in main text.

this region, $\mathbf{x}_c^D \in S_{out}$, the species potentially provides its individuals to the sampling region S (Fig. 1). The sampling region S is the subset of S_{out} , $S \subset S_{out}$. In this simplest example, the average number of species given sampling area $\nu(S)$ is described:

$$\begin{aligned} N_s(\nu(S)) &= \int_{S_{out}} f(\mathbf{x}) p^{nz}(\mathbf{x}_c^S, \mathbf{x}) d\mathbf{x}, \\ &= \lambda_s \nu^{nz}(S_{out}). \end{aligned} \quad (5)$$

Since the first differentiation of area is proportional to length, the final output of Eq. (5) represents the area of region S_{out} weighted by nonzero probability $p^{nz}(\mathbf{x}_c^S, \mathbf{x})$ over the region S_{out} . The SAR is obtained by calculating Eq. (5) across sampling areas $\nu(S)$. From Eq. (5), we can show that the slope of the SAR on a log-log plot approaches 1 at large scales. Firstly, the slope is calculated by

$$\frac{d \log N_s(\nu(S))}{d \log \nu(S)} = \frac{\nu(S)}{N_s(\nu(S))} \frac{d N_s(\nu(S))}{d \nu(S)}. \quad (6)$$

Secondly, when the sampling region becomes significantly larger than the distribution range $S \gg D$, S_{out} is nearly equivalent to S . In this limit, $p^{nz}(\mathbf{x}_c^S, \mathbf{x}) \simeq 1$ is satisfied, providing

$$\begin{aligned} N_s(\nu(S)) &\simeq \int_S f(\mathbf{x}) d\mathbf{x}, \\ &= \lambda_s \nu(S). \end{aligned} \quad (7)$$

The slope 1 is obtained by substituting this into Eq. (6). Due to the prerequisite relationship $S \gg D$, this asymptotic slope is achieved faster if the distribution ranges are smaller.

As noted above, it is easy to incorporate species variations by assuming that some biological properties follow certain probability distribution functions (pdfs). It is suffice to consider biological parameters that affect the non-zero probability $p^{nz}(\mathbf{x}_c^S, \mathbf{x})$ (e.g., distribution range and intensity of each species). We define such parameters by a vector $\mathbf{b} = (b_1, b_2, \dots, b_k) \in B$, where B is a biologically plausible parameter space. Here, let us assume the independence of biological parameters $\mathbf{b}$ from the distribution ranges D . D varies between species, and hence it generates a probability distribution of S_{out} , $p(S_{out})$. Therefore, the probability distribution functions of biological parameters and the distribution range are described by $p(\mathbf{b}, S_{out}) = p(\mathbf{b})p(S_{out})$. Now, Eq. (5) becomes

$$\begin{aligned} N_s(\nu(S)) &= \int_{\mathbf{b} \in B} p(\mathbf{b}) \int_{\mathbf{R}^2} p(S_{out}) \int_{S_{out}} f(\mathbf{x}) p_{\mathbf{b}}^{nz}(\mathbf{x}_c, \mathbf{x}) d\mathbf{x} dS_{out} d\mathbf{b}, \\ &= \lambda_s E_{\mathbf{b}} E_{S_{out}} [\nu^{nz}(S_{out})], \end{aligned} \quad (8)$$

where, $E_{S_{out}}$ and $E_{\mathbf{b}}$ are the average over the distribution ranges and biological parameters, respectively. As $\nu^{nz}(S_{out})$ is the average area to find at least one individual, $E_{\mathbf{b}} E_{S_{out}} [\nu^{nz}(S_{out})]$ is the such area averaged over the pdfs $p(\mathbf{b})$ and $p(S_{out})$. Eq. (5) is simply a special case of Eq. (8) by setting $p(\mathbf{b}) = p(S_{out}) = 1$ with choosing a set of parameters. We can make intuitive discussions about the effect of biological parameters from Eq. (8). For example, if the intensity of each species is high, it is likely to increase the area $\nu^{nz}(S_{out})$ and it increases the number of species in given area, and vice versa. In this situation, the same discussion about the slope of SAR curve Eq. (7) can be made as long as the sampling region is significantly larger than distribution range of each species, $S \gg D_i$ for all i .

2.3 Endemic area relationship

Next we discuss the derivation of the endemic area relationship (EAR). Despite its importance to conservation biology, understanding of the empirical shape across scales is largely deficient [10]. Here, we define the endemic species, with explicitly introducing distribution region with an arbitrary shape, as the species with its distribution range D is covered by the sampling region $D \subset S$. This definition is different from the commonly used definition (e.g., [10, 11, 16]) where the endemic species is defined, using the SAR, as the average number of species these population is completely covered by area A : $N_{en}(A) = N_s(A') - N_s(A' \setminus A)$, where $N_{en}(A)$ is the number of endemic species in area A , and A' is the system size. This

definition relies on the occurrence probability of individuals of a species, and hence its individual distribution patterns. Typically SAR curves show $N_s(A) > 0$ with $A > 0$, and therefore $N_{en}(A) > 0$ for any small A . On the other hand, our definition is based on an explicit form of the endemic region, and the number of the endemic species is 0 if the sampling region is smaller than distribution range of all species: $S < D_i$ for all i . In this framework, individual distribution patterns do not affect the EAR. This definition may be more straightforward to discuss an extinction of species, since habitable regions of all species are explicitly taken into account. However, as we will see below, the two definitions become equivalent in the large limit of A .

In our framework, the number of endemic species given area is calculated as follows

$$\begin{aligned} N_{es}(\nu(S)) &= \int_S f(\mathbf{x}) p^{en}(\mathbf{x}_c^S, \mathbf{x}) d\mathbf{x}, \\ &= \int_{S_{core}} f(\mathbf{x}) d\mathbf{x} + \int_{S \setminus S_{core}} f(\mathbf{x}) p^{en}(\mathbf{x}_c^S, \mathbf{x}) d\mathbf{x}, \\ &= \lambda_s(\nu(S_{core}) + \nu^{en}(S \setminus S_{core})), \end{aligned} \quad (9)$$

where, $p^{en}(\mathbf{x}_c^S, \mathbf{x}_c^{D_i})$ is the endemic probability that a distribution range with its geometric center $\mathbf{x}_c^{D_i} \in S \setminus S_{core}$ becomes subset of the sampling region with the geometric center $\mathbf{x}_c^S$, $D_i \subset S$. S_{core} is the subset of the sampling region S where if the distribution center of a species $\mathbf{x}_c^{D_i}$ falls in this regions, then the entire distribution range D_i is covered by S : $P(D_i \subset S) = 1$ provided that $\mathbf{x}_c^{D_i} \in S_{core}$ (Fig. 1). As in the case of the SAR, we can show the slope of the EAR on a log-log plot approach 1 when $S \gg D$. In this limit, S_{core} is nearly equivalent to S , and hence Eq. (9) becomes

$$\begin{aligned} N_{es}(\nu(S)) &\simeq \int_S f(\mathbf{x}) d\mathbf{x}, \\ &= \lambda_s \nu(S). \end{aligned} \quad (10)$$

This is equivalent to Eq. (7), and hence the slope approaches 1. As we noted, the same result can be obtained in another definitions using the SAR, since $N_{en}(A) \simeq N_s(A)$ when A approaches A' .

We now introduce the EAR in the situation where species have different distribution areas $\nu(D_i)$. This generates a probability distribution of the core region of sampling, $p(S_{core})$. By introducing the variation in distribution areas, Eq. (9) now becomes

$$\begin{aligned} N_{es}(\nu(S)) &= \int_{\mathbf{R}^2} p(S_{core}) \left(\int_{S_{core}} f(\mathbf{x}) d\mathbf{x} + \int_{S \setminus S_{core}} f(\mathbf{x}) p^{en}(\mathbf{x}_c^S, \mathbf{x}) d\mathbf{x} \right) dS_{core}, \\ &= \lambda_s E_{S_{core}}[\nu(S_{core}) + \nu^{en}(S \setminus S_{core})]. \end{aligned} \quad (11)$$

As in the case of the SAR, we obtain Eq. (9) by setting $p(S_{core}) = 1$ in Eq. (11). In addition, the same discussion about the slope of the EAR can be made in this situation, if the sampling region is significantly larger than the distribution ranges: $S \gg D_i$ for all i .

2.4 Relative species abundance

Here, we develop a method to derive the relative species abundance (RSA). RSA, or SAD, a histogram of species count, is a product of multiple ecological mechanisms such as species interactions and stochasticity [7]. For convenience, we use the notation $P(X = x)$ to describe the probability of finding a species with abundance x . Derivation of the RSA is more cumbersome than the SAR and EAR since we use all probabilities of $P(X = x)$. In our framework, the average abundance of species i is proportional to the area where the sampling region S and the distribution range of the species D_i are overlapped, $\nu(S \cap D_i)$. Each sampled species has different overlapped area ($\nu(S \cap D_i) \neq \nu(S \cap D_j)$, ($i \neq j$)), and its area changes as sampling and distribution scales as well as these shapes change. Therefore, to calculate the RSA across scales we need a pdf of $\nu(S \cap D)$, provided $\nu(S \cap D) \neq \phi$, which describes the variation of $\nu(S \cap D)$ between sampled species.

Roughly speaking, there exist two regimes to determine the maximum value of $\nu(S \cap D)$ among sampled species: $S > D$ and $S < D$, provided $l_s > L_d$ and $L_s < l_d$, respectively. Namely, the maximum overlapped area is either $\nu(S)$ or $\nu(D)$ depending on $S < D$ (Fig. 2a, b) or $S > D$ (Fig. 2c, d), respectively: $\max\{\nu(S \cap D)\} = \min\{\nu(S), \nu(D)\}$. This maximum

Figure 2: Schematic representation of typical geometric pattern in the derivation of the relative species abundance across sampling scales. The sampling region S and distribution range D are described by square and ellipse as an example. In each panel, different sampling scales are shown, but the size of distribution ranges remain the same. (a) and (d): when the sampling scale is sufficiently small or large, all sampled species have the same overlapped distribution range with the sampling region ($\nu(S)$ and $\nu(D)$, respectively). (b) and (c): in intermediate sampling scales, this overlapped region is different between sampled species.

value in the overlapped area occurs if the geometric center of species distribution $\mathbf{x}_c^D$ falls in S_{core} : $P(\nu(S \cap D) = \min\{\nu(S), \nu(D)\}) = 1$ provided $\mathbf{x}_c^D \in S_{core}$ (Fig. 1). As the assumption of random placement of species distribution center, the probability that the overlapped area

is its maximum value is proportional to the area of the S_{core} in S_{out} . Therefore, we have the following relationships:

$$\begin{cases} P(\nu(S \cap D) = \min\{\nu(S), \nu(D)\}) = \frac{\nu(S_{core})}{\nu(S_{out})}, \\ P(0 \leq \nu(S \cap D) < \min\{\nu(S), \nu(D)\}) = \frac{\nu(S_{out} \setminus S_{core})}{\nu(S_{out})}. \end{cases} \quad (12)$$

As implied in Fig. 2a and d, when $S \ll D$ or $S \gg D$ is satisfied, $P(\nu(S \cap D) = \min\{\nu(S), \nu(D)\}) \simeq 1$ is satisfied. The abundance of a species within overlapped area $\nu(S \cap D)$ shows variation, and its pdf corresponds to the relative species abundance in area $\nu(S \cap D)$. Let $P(x | \nu(S \cap D))$ be the probability of finding a species with x individuals in the sampling region S given the overlapped area $\nu(S \cap D)$. Then the RSA provided an arbitrary sampling area $\nu(S)$ is described as

$$\begin{aligned} P(x(\nu(S))) &= \frac{\nu(S_{core})}{\nu(S_{out})} P(x | \min\{\nu(S), \nu(D)\}) + \\ &\quad \frac{\nu(S_{out} \setminus S_{core})}{\nu(S_{out})} P(x | 0 \leq \nu(S \cap D) < \min\{\nu(S), \nu(D)\}). \end{aligned} \quad (13)$$

Especially, when $S \ll D$ or $S \gg D$ is satisfied, Eq. (13) is simplified as

$$P(x(\nu(S))) \simeq P(x | \min\{\nu(S), \nu(D)\}). \quad (14)$$

In line with the discussion above, it is straightforward to incorporate the species variations such as the distribution range D and biological parameters $\mathbf{b}$. By introducing pdfs these parameters follow: $p(D)$ and $p(\mathbf{b})$, assuming that the stochastic variable x is now a function of $\mathbf{b}$, $x(\mathbf{b})$, and averaging Eq. (13) by the pdfs, we obtain

$$\begin{aligned} P(x(\mathbf{b}) | \nu(S)) &= \int_{\mathbf{b} \in B} p(\mathbf{b}) \int_0^\infty p(D) \left[\frac{\nu(S_{core})}{\nu(S_{out})} P(x(\mathbf{b}) | \min\{\nu(S), \nu(D)\}) + \right. \\ &\quad \left. \frac{\nu(S_{out} \setminus S_{core})}{\nu(S_{out})} P(x(\mathbf{b}) | 0 \leq \nu(S \cap D) < \min\{\nu(S), \nu(D)\}) \right] dD d\mathbf{b}. \end{aligned} \quad (15)$$

3 Generating species distribution patters

To examine the general theory of the geometric approach developed above, we need to introduce a point field $f(\mathbf{x})$ defined above that holds information of individual distributions of all species. For this purpose, we make use of point processes [23, 24], a set of spatially explicit stochastic models that generate various point distribution patterns such as random and clustering patterns. One of the advantages of this models is that point processes are amenable to mathematical analysis and easy to implement numerical simulations. In addition, there are a number of applications to ecological studies [11, 19, 20, 22, 25], and models can provide consistent patterns with observed SARs or a population occupancy probability regardless of simple assumptions [11, 22, 25].

Here, we examine the theory developed using two individual distribution patterns: random and clustering distribution patterns, described by the homogeneous Poisson process and the Thomas process, respectively. The former is often used to develop a simplest possible model (e.g., [26]), or to see how the simple assumption deviates from more biologically reliable models (e.g., [16, 19, 22]). As we will see below, deviations of these two distributions in the SAR and RSA are relatively small to make qualitative discussions and these appear in small scales.

The homogeneous Poisson process and Thomas process are defined using the intensity and intensity measure of species i , defined above (Eqs. 1 and 2). Also, as mentioned above, individuals of species i is restricted within its distribution range D_i . For example, if species i is distributed randomly in region A , the probability to find x individuals within the region with area $\nu(A)$ follows the Poisson distribution with average $\mu_i(A) = \lambda_i \nu(A)$:

$$P(X = x) = \frac{\mu_i(A)^x}{x!} e^{-\mu_i(A)}. \quad (16)$$

This process generates the homogeneous Poisson process.

On the other hand, the Thomas process is described by the following three steps:

1. Parents of species i are randomly placed according to the homogeneous Poisson process with intensity λ_i^p .
2. Each parent of species i produces a random discrete number c_i of daughters, realized independently and identically.
3. The daughters are scattered around their parents independently with an isotropic bivariate Gaussian distribution with the variance σ_i^2 , and all the parents are removed in the realized point pattern.

The intensity of individuals of species i for the Thomas process is [24]

$$\lambda_{th,i} = \bar{c}_i \lambda_i^p, \quad (17)$$

where, $\bar{c}_i$ is the average number of daughters per parent. To guarantee a consistent number of total expected individuals given area between the two processes, we set λ_i^p and $\bar{c}_i$ as

$$\lambda_{th,i} = \bar{c}_i \lambda_i^p = \lambda_i. \quad (18)$$

We also assume that the number of daughters per parents c_i follows the Poisson distribution with the average number $\bar{c}_i$.

By superposing distributions of all species generated either by the homogeneous Poisson process or the Thomas process, we obtain individual distributions of all species in the whole ecosystem. With these specific individual distribution patterns, we can examine the SAR, EAR, and RSA across scales.

4 Emergent macroecological patterns of geometric model

By using the general theory developed, we discuss some specific situations where sampling region and the species distribution ranges are described by circles (Fig. 3a). Also, individual distributions therein are described the homogeneous Poisson process or Thomas process, showing random and clustering distribution patterns, respectively. The assumption of the shapes makes mathematical analysis rather simple and transparent as we can omit the effect of rotation of the species distribution ranges. As we will see below, the second term of the EAR (Eqs. 9, 11) disappears under this assumption. For convenience, all the analyses below are conducted in the polar coordinate. The schematic diagrams under this situation corresponding to Fig. 1 are shown in Fig. 3.

Let the intensity measure of species, μ_s , be a function $f(r, \theta)$ defined in the polar coordinate and we assume the homogeneous environment, then the intensity measure of species in the circle with a radius R is

$$\begin{aligned}\mu_s &= \int_0^{2\pi} \int_0^R f(r, \theta) r dr d\theta, \\ &= 2\pi \int_0^R f(r) r dr.\end{aligned}\tag{19}$$

If we define $\mu_s := \lambda_s \pi R^2$, where λ_s is the intensity of species, $f(r)$ is calculated as

$$f(r) = \lambda_s.\tag{20}$$

Figure 3: Schematic representations of (a) the sampling region and distribution ranges, and typical geometric relationships used in derivations of the (b) SAR, (c) EAR, and (d) RSA. S is the sampling region with radius r_s , D is the distribution range of a species with radius r_d . The distance between the centers of S and D is represented by r (in this example, $r = r_s + r_d$). See Fig. 1 for more explanations.

4.1 Species area relationship

Let us consider the situation where each species has its distribution area πr_d^2 with a radius r_d . We assume that centers of distribution ranges of species are randomly distributed across

space, and individual distributions therein show either random or clustering pattern. Then our sampling regime is introduced in such a way that we randomly place the sampling region with radius r_s and count the number of species across spatial scales by changing the size of sampling region. By doing so, we obtain the SAR (Fig. 3a). From Eqs (19, 20) and with the aid of Fig. 3b, we calculate the number of species N_s found within a sampling unit with area πr_s^2 , as

$$\begin{aligned} N_s(\pi r_s^2) &= \int_0^{2\pi} \int_0^{r_s+r_d} f(r, \theta) p^{nz}(r | r_s, r_d) dr d\theta, \\ &= 2\pi \lambda_s \int_0^{r_s+r_d} r p^{nz}(r | r_s, r_d) dr, \\ &= 2\pi \lambda_s E_{nz}[r], \end{aligned} \quad (21)$$

where, $p^{nz}(r | r_s, r_d)$ is the non-zero probability, provided r_s and r_d , that at least one individual of a specie is found given an r -distance separation of the centers of sampling and distribution range. Note that $2\pi \int_0^{r_s+r_d} r p^{nz}(r | r_s, r_d) dr$ in the second line gives the area corresponding to $\nu^{nz}(S_{out})$ in Eq. (5). Eq. (21) implies that the number of species within the sampling region S is proportional to the average distance that at least one individuals are found.

We can generalize Eq. (21) to a situation where each species has a different distribution range, therefore different r_d , and biological parameter $\mathbf{b}$, ($\mathbf{b} \in B$). Then r_d and $\mathbf{b}$ follow probability distribution functions, $p(r_d)$ and $p(\mathbf{b})$, respectively. In that situation, Eq. (21) becomes

$$\begin{aligned} N_s(\pi r_s^2) &= \int_{\mathbf{b} \in B} p(\mathbf{b}) \int_0^\infty p(r_d) \int_0^{2\pi} \int_0^{r_s+r_d} r p^{nz}(r | r_s, r_d) dr d\theta dr_d d\mathbf{b}, \\ &= 2\pi \lambda_s E_{\mathbf{b}} E_{r_d} E_{nz}[r]. \end{aligned} \quad (22)$$

Since this is a special case of the general theory developed above, the discussion about asymptotic slope (Eq. 7) holds.

Here, to show some numerical results in the situation where the distribution range and/or biological parameters differ between species, we examine a situation where the radius of distribution range r_d follows the exponential distribution with the parameter λ_e

$$f(r_d) = \lambda_e e^{-\lambda_e r_d}. \quad (23)$$

For the biological parameter, we assume that the intensity of each species λ_i follows the gamma distribution:

$$f(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\lambda\beta}, \quad (24)$$

where, α and β are shape and rate parameters of the gamma distribution respectively. We replace λ by λ_i or $\lambda_{th,i}$ depending on the underlying individual distribution process. In addition, we define $\bar{c}_i$ and λ_i^p as $\sqrt{\lambda_{th,i}}$ to satisfy $\lambda_i = \lambda_{th,i} = \bar{c}_i \lambda_i^p$. Fig. 4 shows numerically

calculated values of neutral (Eq. 21) and non-neutral (Eq. 22) situations, showing the triphasic feature on a log-log plot with different distribution ranges and individual distribution patterns. The third phase of the SAR appears around the average distribution area of species. As discussed above, the slope of each curve asymptotically approaches 1 at large scales. Also, if the distribution range is smaller, it approaches the asymptotic value faster as discussed above. Moreover, it shows that the difference between two individual distribution patterns are relatively small and slight deviations appear only on small sampling scales. Difference between the neutral situation and the non-neutral situation but with only the intensity λ_i are negligibly small when individual distributions is described by the homogeneous Poisson process with the radius of the distribution range $r_d = 10\text{km}$. We checked this holds true for the other curves, and also for different parameter sets of gamma distribution.

Figure 4: Species area relationship [SAR; (a)] and its slope (b) under the homogeneous Poisson process (red and green) and Thomas process (blue), obtained using Eqs. (21) and (22). The intensity λ_i varies between species according to the gamma distribution except for the neutral situation (star; labeled Poisson_N). The lines labeled $E[r_d]$ represents the situation where distribution ranges follows the exponential distribution. Otherwise, the distribution range is identical between species (line; $r_d=10\text{km}$ line, and dashed; $r_d=100\text{km}$). For the Thomas process, the parameters are $\lambda_s = 0.00064$ (50000 species/ $\pi \times 5.0 \times 10^8\text{km}^2$), $\lambda_{th,i} = 100$, $\bar{c}_i = 10$, $\lambda_i^p = 10$, $\sigma_p = 0.1$. For the gamma distribution, we used $\alpha = 10$ and $\beta = 0.1$ ($E[\lambda_i] = 100$), and for the exponential distribution $\lambda_e=0.1$ ($E[r_d]=10$) and 0.01 ($E[r_d]=10$) are used.

4.2 Endemic area relationship

When both the sapling region and endemic region are circle, the distribution range of an endemic species satisfies the condition shown in Fig. 3c: the sampling region must be larger than distribution range (i.e., $r_s > r_d$), and the distance between S and D must be small enough (i.e., $r \leq r_s - r_d$). Therefore, the number of endemic species N_{es} in area πr_s^2 is described as

$$\begin{aligned} N_{es}(\pi r_s^2) &= \int_0^{2\pi} \int_0^{r_s - r_d} f(r, \theta) dr d\theta, \\ &= \pi \lambda_s (r_s - r_d)^2. \end{aligned} \quad (25)$$

By applying Eq. (6) to calculate the slope of the EAR, it is easily shown provided $r_s > r_d$

$$\frac{d \log N_{es}(\pi r_s^2)}{d \log(\pi r_s^2)} = \frac{r_s}{(r_s - r_d)^2} \left(1 - \frac{r_d}{r_s} \right). \quad (26)$$

As discussed above, it approach 1 as the sampling region becomes significantly larger than the distribution range: $r_s \gg r_d$.

As above, we can generalize the EAR to incorporate variations in distribution ranges of species by introducing a probability distributions, $p(r_d)$. However, as noted above, biological parameters does not affect the EAR. For the sake of comparison with the case of the constant r_d , let us assume the average value satisfies $E[r_d] = r_d$. Eq. (25) now becomes

$$\begin{aligned} N_{es}(\pi r_s^2) &= \int_0^{2\pi} \int_0^{r_s} p(r_d) \int_0^{r_s - r_d} f(r, \theta) dr dr_d d\theta, \\ &= 2\pi \lambda_s \int_0^{r_s} p(r_d) \int_0^{r_s - r_d} r dr dr_d, \\ &= \pi \lambda_s \int_0^{r_s} p(r_d) (r_s - r_d)^2 dr_d. \end{aligned} \quad (27)$$

As an example, if we assume the distribution of r_d follows the exponential distribution $p(r_d) = \lambda_e e^{-\lambda_e r_d}$, the number of endemic species is

$$\begin{aligned} N_{es}(\pi r_s^2) &= \pi \lambda_s \int_0^{r_s} \lambda_e e^{-\lambda_e r_d} (r_s - r_d)^2 dr_d, \\ &= \frac{2\lambda_s \pi}{\lambda_e^2} \left(1 - r_s \lambda_e + \frac{(r_s \lambda_e)^2}{2} - e^{-r_s \lambda_e} \right), \\ &= \frac{2\lambda_s \pi}{\lambda_e^2} \sum_{n=3}^{\infty} (-1)^{n+1} \frac{(r_s \lambda_e)^n}{n!}. \end{aligned} \quad (28)$$

Since $dN_{es}/\pi r_s^2 = \lambda_s \sum_{n=2}^{\infty} (-1)^n \frac{(r_s \lambda_e)^{n-1}}{n!}$, we calculate the slope as

$$\frac{d \log N_{es}(\pi r_s^2)}{d \log(\pi r_s^2)} = -\frac{r_s \lambda_e}{2} \frac{1 - r_s \lambda_e - e^{-r_s \lambda_e}}{1 - r_s \lambda_e + \frac{(r_s \lambda_e)^2}{2} - e^{-r_s \lambda_e}}. \quad (29)$$

Fig. 5 shows the EAR calculated by Eqs. (25) and (27), and its slope defined by Eqs. (26) and (29). As discussed in General Theory, the EAR on a log-log plot shows asymptotic slope 1, equivalent to the SAR when the sampling region is significantly larger than the distribution ranges $S \gg D$. By the same discussion above, it occurs earlier when the distribution ranges are small. The slope of the EAR becomes infinitely large as soon as the sampling size becomes equivalent to single-sized distribution range. On the other hand, if the distribution range follows the exponential distribution, the slope approaches 1 from a finite value (Fig. 5b), and it also shows that, at small scales, the endemic species is proportional to the sampling area.

Figure 5: (a) Endemic area relationship (EAR) and (b) its slope. Color corresponds to distribution range size (red; $r_d = 10\text{km}$ and blue; $r_d = 100\text{km}$). Solid lines represent the situation where no variation in distribution range occurs (Eq. 25). Dotted lines correspond to the situation where the distribution range varies between species, and it follows the exponential distribution $p(r_d) \sim \exp(-\lambda_e r_d)$ (Eq. 28). To obtain the average value $E[r_d] = \{10, 100\}$, we set $\lambda_e = \{0.1, 0.01\}$. For other parameters, the same values are used as in Fig. 4.

4.3 Relative species abundance

To derive RSAs across sampling scales, we extensively use mixed probability distributions, especially mixed Poisson distributions [27]. The mixed Poisson distribution was first introduced in ecological study by Fisher et al. [2] in which the celebrated Fisher's logseries

was derived. In our geometric approach, mixed Poisson distributions appear by the nature of multiple stochasticity to determine the number of individuals observed. Namely, there exist intra-species (variations from the expected number of individuals) and inter-species variations (variations of the expected number of individuals itself) of individuals within an overlapped area between the sampling and distribution region $\nu(S \cap D)$. In addition, the overlapped area also varies between species, and the probability distribution of the area $P(\nu(S \cap D))$ depends on sampling scale as shown in Fig 2. Hence, to we need to resolve the scale effect to derive the RSA across scales.

Mixed Poisson distributions are known to produce a variety of pdfs [27], and it is hard to make an exhaustive list of potential RSAs. However, we can consistently obtain RSAs in arbitral sampling scales S by scaling up/down, once all the parameters are provided. Therefore, we will firstly focus to recover some well documented RSAs, when the sampling region is much smaller than the distribution range $S \ll D$, probably the most common situation in practice. And secondly, we will scale up the sampling region S , by keeping the same assumption, to see how RSA will change. In the first step, we will recover the negative binomial distribution [28], Fisher’s logseries [2], and Poisson lognormal distribution [29].

In the following analysis, we assume that individual distributions are random: it is described by the homogeneous Poisson process. However, as we will see below, the difference of RSAs between the homogeneous Poisson process and Thomas process are small to discuss qualitative features of the RSA, except for the small scales where the deviations of two processes appear in the SAR (Fig. 4).

4.3.1 Some basic properties of mixed Poisson distribution

Here, we introduce some basic properties of mixed Poisson distributions used in the following analysis. For a mixed Poisson distribution, the Poisson intensity itself also follows a probability distribution $g(\lambda)$, and the probability is described [27]

$$P(X = x) = \int_0^\infty \frac{\lambda^x}{x!} e^{-\lambda} g(\lambda) d\lambda. \quad (30)$$

As in [27], we denote this mixture as

$$f(x | \lambda) \underset{\lambda}{\wedge} g(\lambda). \quad (31)$$

Note that in our application of the homogeneous Poisson process where the probability variable is the number of individuals, the intensity λ in Eq. (30) is replaced by the intensity measure $\mu_i = \lambda_i \nu(A)$ where λ_i and/or $\nu(A)$ vary independently. To specify the probability variable, we use the following expressions:

$$f(x | \lambda_i \nu(A)) \underset{\lambda_i}{\wedge} g(\lambda_i), \quad (32)$$

$$f(x | \lambda_i \nu(A)) \underset{\lambda_i}{\wedge} g(\lambda_i) \underset{\nu(A)}{\wedge} h(\nu(A)), \quad (33)$$

where, the first expression is the case for only λ_i varies and the second expression is the case for both λ_i and $\nu(A)$ vary.

From the Proposition 1 and 2 in Appendix, the probabilities of the mixed Poisson distribution shown in Eqs. (32) and (33) are described

$$P(X = x) = \frac{1}{x!} \sum_{r=0}^{\infty} \nu(A)^{x+r} \frac{(-1)^r}{r!} \mu_{x+r}(\lambda_i), \quad (34)$$

$$P(X = x) = \frac{1}{x!} \sum_{r=0}^{\infty} \frac{(-1)^r}{r!} \mu_{x+r}(\lambda_i) \mu_{x+r}(\nu(A)), \quad (35)$$

respectively, where $\mu_r(x)$ is the r th moment of x about the origin. As we see below, $\nu(A) = \min\{\nu(S), \nu(D)\}$ in our application, and these are RSAs we explore.

4.3.2 Sampling region is much smaller than distribution range: $S \ll D$

In our framework, if species i is sampled, the distribution range of species i must be overlapped with the sampling region $P(S \cap D_i = \phi) \neq 0$. Provided by $P(S \cap D_i = \phi) \neq 0$, and if the sampling region is much smaller than the distribution range, the sampling region is completely covered by the distribution range $P(\nu(S \cap D_i) = \nu(S)) = 1$ as in Fig. 2a. In that situation, we can neglect the variations of the overlapped region between species, and consider only the abundance variations of inter- and intra-species within the the sampling region S . Namely, this is the situation described by Eq. (32) and the RSA under this condition is obtained by calculating Eq. (34).

Negative binomial (Poisson-gamma) distribution and Fisher's logseries

It is well known that the mixing the Poisson distribution with the gamma distribution produces the negative binomial distribution [27]. By taking certain limits of the negative binomial distribution, Fisher et al. [2] obtained the Fisher's logseries. Here, we follow the same idea to derive these two distributions. Firstly, by Eq. (30), mixing the Poisson distribution with the gamma distribution Eq. (24) gives the following probability function

$$P(X = x) = \frac{\beta^\alpha}{x! \Gamma(\alpha)} \int_0^\infty \lambda^{x+\alpha-1} e^{-\lambda(1+\beta)} d\lambda, \quad (36)$$

In our situation, the mixture is described by Eq. (32), not by Eq. (31), and the probability of finding a species with x individuals are described

$$\begin{aligned} P(X = x) &= \int_0^\infty \frac{(\lambda_i \nu(S))^x}{x!} e^{-\lambda_i \nu(S)} \frac{\beta'^\alpha}{\Gamma(\alpha)} \lambda_i^{\alpha-1} e^{-\lambda_i \beta'} d\lambda_i, \\ &= \frac{\beta'^\alpha \nu(S)^x}{x! \Gamma(\alpha)} \int_0^\infty \lambda_i^{x+\alpha-1} e^{-\lambda_i(\beta' + \nu(S))} d\lambda_i, \end{aligned} \quad (37)$$

where, we set $\beta' = \beta \nu(S)$ and changing the variable as $\nu(S) \lambda_i = \lambda'_i$, we recover the form of Eq. (36). Eq. (36) is well known to produce the negative binomial distribution

$$P(X = x) = \binom{x + \alpha - 1}{x} p^x (1 - p)^\alpha, \quad (38)$$

where, in our case $p = \nu(S)/(\nu(S) + \beta)$.

To obtain Fisher's logseries from Eq. (38), we need to further assume the sampling effect: the number s of individuals are sampled. By taking the Fisher's limit [30], we obtain:

$$P(X = x) = \lim_{\substack{s \rightarrow \infty, \alpha \rightarrow 0 \\ s\alpha \rightarrow \gamma}} s \binom{x + \alpha - 1}{x} p^x (1 - p)^\alpha = \frac{\gamma p^x}{x}. \quad (39)$$

Poisson-lognormal distribution

Poisson-lognormal distribution is obtained by mixing the Poisson distribution with the log-normal distribution $f(\lambda) = 1/(\lambda\sqrt{2\pi\sigma^2})e^{-(\log \lambda - \mu)^2/2\sigma^2}$ [29]:

$$P(X = x) = \frac{1}{x!\sqrt{2\pi\sigma^2}} \int_0^\infty \lambda_i^{x-1} e^{-\lambda_i - \frac{(\log \lambda_i - \mu)^2}{2\sigma^2}} d\lambda_i \quad (40)$$

In our situation, the mixture is as in Eq. (32), and it is described

$$\begin{aligned} P(X = x) &= \int_0^\infty \frac{(\lambda_i \nu(S))^x}{x!} e^{-\lambda_i \nu(S)} \frac{1}{\lambda_i \sqrt{2\pi\sigma^2}} e^{-\frac{(\log \lambda_i - \mu)^2}{2\sigma^2}} d\lambda_i, \\ &= \frac{\nu(S)}{x!\sqrt{2\pi\sigma^2}} \int_0^\infty (\lambda_i \nu(S))^{x-1} e^{-\lambda_i \nu(S) - \frac{(\log \lambda_i - \mu)^2}{2\sigma^2}} d\lambda_i, \end{aligned} \quad (41)$$

where, by setting $\mu = \mu' - \log \nu(S)$ and $\nu(S)\lambda_i = \lambda'_i$, we recover the form of Eq. (40). Therefore, if the parameter λ_i follows the lognormal distribution and sampling region is much smaller than the distribution range, we expect to observe an RSA curve that follows the Poisson-lognormal distribution.

Figure 6: An example form of the overlapped area $\nu(S \cap D)$ given an r -distance separation of geometric centers between the sampling region and the distribution range of the species (a), and the cumulative distribution functions of overlapped area (b).

4.3.3 General situation

With the result of the analysis above, we explore more general situations where the sampling region does not necessarily satisfy $S \ll D$. In that situation, the probability function of finding a species with x individuals has the form of Eq. (13), in which as in Eq. (14), the above-mentioned situation $S \ll D$ (and $S \gg D$) is treated as a special situation. The weighting terms Eq. (12) in this example is easily calculated with the aid of Fig. 3:

$$\begin{cases} \frac{\nu(S_{core})}{\nu(S_{out})} = \frac{\max\{r_d - r_s, r_s - r_d\}}{r_d + r_s}, \\ \frac{\nu(S_{out} \setminus S_{core})}{\nu(S_{out})} = \frac{\min\{2r_s, 2r_d\}}{r_d + r_s}, \end{cases} \quad (42)$$

where, one can easily see that the second weighting term disappears when $S \ll D$ and $S \gg D$, namely $r_s \ll r_d$ and $r_s \gg r_d$, respectively. The first probability function of Eq. (13) corresponds to the situation where we can neglect the variation of overlapped region, and this situation was already discussed above. Namely, it is described by Eq. (34). The second probability of Eq. (13) corresponds to the situation where the overlapped area varies between species. As we discussed above, the mixed probability distribution has the relationship of Eq. (33), and the probability of finding a species with x individuals is described by Eq. (35).

For simplicity, we focus here on the case where the λ_i follows the gamma distribution as in the first example above (Poisson-gamma), since this case is mathematically more amenable than the case of Poisson-lognormal, and therefore it allows us to discuss results in a mathematically more transparent way. This sole assumption, however, produces a variety of RSAs including similar forms to the negatively skewed lognormal distribution [6], or a left-skewed bell shape curve on a logarithmic scale, resembling the Poisson-lognormal distribution.

To make use of Eq. (35), we need the r th moment of a probability distribution of the overlapped area $\nu(S \cap D)$. Fig. 6a shows an example of the relationship of the overlapped area $\nu(S \cap D)$ given an r -distance separation of geometric centers between the sampling region and the distribution range of species. Since r is randomly given provided that $\nu(S \cap D) \neq \phi$, $\nu(S \cap D)$ is a probability variable. We can therefore treat the graph (Fig. 6b), obtained by rotating Fig. 6a and normalizing to 1, as a cumulative distribution function of $P(\nu(S \cap D))$, and infer the pdf. Especially, since this function has two different phases, we decompose it into two parts. The region described with δ corresponds to the first probability function in Eq. (13) in which the probability is described by a delta distribution and has a peak at $\min\{\nu(S), \nu(D)\}$. The rest part corresponds to the second probability function in Eq. (13), and it has a similar shape to an arcsin function. Because we require the moment of a probability distribution function, $P(\nu(S \cap D))$, it may be a reasonable to approximate it by an existing probability distribution. Therefore, we apply the arcsin distribution, which is a special case of the beta distribution [31]

$$f(x) = \frac{1}{\pi \sqrt{x(\min\{\nu(S), \nu(D)\} - x)}}, \quad x \in [0, \min\{\nu(S), \nu(D)\}] \quad (43)$$

where, the support is defined as $x \in [0, \min\{\nu(S), \nu(D)\}]$. The r th moment of Eq. (43) about the origin is, by the Proposition 3 in Appendix, described by $\mu_r^{arc} = \frac{\nu(S)^r}{\pi} B(\frac{1}{2}, \frac{1}{2} + r)$

where, $B(x, y) = \int_0^1 u^{x-1}(1-u)^{y-1}du$ is the beta function. By substituting this equation, together with the r th moment of the gamma distribution Eq. (24), $\mu_r^{gam} = \frac{\Gamma(r+\alpha)}{\Gamma(\alpha)} \frac{1}{\beta^r}$, into Eq. (35), we obtain the following probability function (see Appendix for the detailed derivation):

$$P(X = x) = \gamma \binom{x + \alpha - 1}{x} p^x (1-p)^\alpha {}_2F_1\left(\frac{1}{2} + x, x + \alpha; 1 + x; -\frac{\min\{\nu(S), \nu(D)\}}{\beta}\right), \quad (44)$$

where, γ is the coefficient $\gamma = \Gamma(1/2+x)(\sqrt{\pi}\Gamma(x+1))^{-1}(1-p)^{-\alpha-x}$ and this part is simplified for $x \geq 1$ as $\gamma = x^{-1}B(1/2, x)^{-1}(1-p)^{-\alpha-x}$, and p is now becomes $p = \min\{\nu(S), \nu(D)\}/(\min\{\nu(S), \nu(D)\} + \beta)$. The second to forth factors correspond to the negative binomial distribution the form corresponds to Eq. (38), and ${}_2F_1$ is the Gauss hypergeometric function. We call this distribution the Poisson-gamma-arcsine distribution. From, Eqs. (38), (42), and (44), we derive the general form as in Eq. (13):

$$P(X = x) = \binom{x + \alpha - 1}{x} p^x (1-p)^\alpha \times \left(\frac{\max\{r_d - r_s, r_s - r_d\}}{r_d + r_s} + \gamma \frac{\min\{2r_s, 2r_d\}}{r_d + r_s} {}_2F_1\left(\frac{1}{2} + x, x + \alpha; 1 + x; -\frac{\min\{\nu(S), \nu(D)\}}{\beta}\right) \right), \quad (45)$$

where, the probability distribution is a weighting sum of the negative binomial distribution and the Poisson-gamma-arcsine distribution. As already discussed above, $S \ll D$ ($S \gg D$) is the special case of Eq. (45). That is, by taking $r_s \ll r_d$ ($r_s \gg r_d$), Eq. (45) becomes

$$P(X = x) \simeq \binom{x + \alpha - 1}{x} p^x (1-p)^\alpha. \quad (46)$$

Namely, the RSAs follow the same probability distribution when the area of sampling region and distribution range are significantly different. It is worth emphasizing that to derive Eq.(46), the term approximated by arcsine distribution disappears. Since all the parameters except for $\nu(S)$ are consistent across spatial scales, the difference between these two cases is only in $p = \min\{\nu(S), \nu(D)\}/(\min\{\nu(S), \nu(D)\} + \beta)$. It is expected that p is close to 1 if $\nu(D) = \min\{\nu(S), \nu(D)\}$. The same discussion can be made when the intensity follows the Poisson-lognormal distribution Eq. (40), and this is discussed in Appendix (see Poisson-lognormal-arcsine distribution), but with mathematically less tractable and computationally intensive form.

Examples of RSAs derived by Eq. (45) across scales (the radius of the sampling region is $r_s = 0.05, 0.1, 0.5, 10, 100$ and 1000km , respectively) are shown in Fig. 7, associated with numerically obtained RSAs provided that the underlying individual distributions are the homogeneous Poisson or Thomas processes. Since numerical calculations with large distribution ranges are computationally expensive, we set $r_d = 10\text{km}$. However, this is suffice to check our analysis. Especially, we examined the zero-truncated form of Eq. (45), since in practice we do not observe the event $X = 0$. To do this, we multiply each distribution by $(1 - P(X = 0))^{-1}$, where $P(X = 0)$ is $(1-p)^\alpha$ for Eq. (38) and ${}_2F_1(\frac{1}{2}, \alpha; 1; -\frac{\min\{\nu(S), \nu(D)\}}{\beta})$ for Eq. (44), respectively.

In general, the homogeneous Poisson process and Thomas processes show qualitatively similar curves except for Fig. 7b where deviations between two processes in the SAR appear (Fig. 4). As expected, when the spatial scales of sampling and distribution ranges are significantly different (i.e., Figs. 7a and b; $S \ll D$; the weighting term is $(r_d - r_s)/(r_d + r_s) < 0.02$) the analytical results show good agreement with numerical results, as Eq. (45) approaches the non-approximated RSA (i.e., Poisson-gamma distribution; Eq. 46). Outside this region, the effect of approximation appears (Figs. 7c-f), showing deviations from the numerical results especially for small x , but it still describe qualitative aspects of each RSA. The tail on small x (Fig. 7f) disappears in the large limit of the sampling region (Eq. 46) Fig. 8, showing a left-skewed bell shape on a logarithmic scale.

Figure 7: Relative species abundance across sampling regions. The radius of the sampling region in each panel (a)-(f) is $r_s = 0.05, 0.1, 0.5, 10, 100$ and 1000km , respectively, and these values are chosen so as to make transitions of the RSA tractable. Each panel shows three different curves regarding to the situations where underlying individual distributions are the homogeneous Poisson process and its theoretical form (Eq. 45), and Thomas process. The theoretical form agrees well with its fully simulated values when the spatial scales of sampling and distribution range are significantly different (i.e., $S \ll D$ and $S \gg D$), in which Eq. (45) approaches to Eq. (46) where no approximation is made. Outside these regions, the theoretical curve still captures qualitative aspects of the numerical results, but with different degrees. See the main text for further explanations.

Figure 8: Relative species abundance in the large limit of the sampling region (Eq. 46) with four different parameter sets (α, β) of the gamma distribution.

5 Spatial scaling of β diversity: An application of geometric approach

β diversity is an important concept in community ecology and conservation biology that describes variations of species compositions between multiple assemblies across spatial scales [13, 32]. The spatial variations inherently include the scaling effect of the size of concerned region and its subregions, and its effect and relationship between other macroecological properties is often of interest of community ecologists [13]. However, data is often not sufficiently enough to unveil the relationship empirically, and as far as we know, there is no theoretical attempt to explore its property (but Barton et al. [13] provided conceptual discussions). Here, to demonstrate potential uses of the geometric model developed, we apply the model to discuss this issue. For this purpose, our analysis is restricted to two scenarios: same parameter values are used as the results of the SARs $E[r_d] = 10^1 \text{km}^2$ Poisson/Thomas shown in Fig. 4. In the following, we briefly summarize scaling issues of β diversity and the diversity index used in the analysis.

5.1 Spatial grain and spatial extent

To quantify β diversity it requires us to define two spatial scales; *spatial extent* and *spatial grain* [13] Fig. 9. Spatial extent is the scope of our observation, and spatial grain is the unit of sampling within the extent. Once we define the spatial extent, spatial grain, and point patterns, we obtain the matrix P : provided that there are s species in the spatial extent, and arbitrarily dividing the community into n assemblies, the occurrence probability of species i

in community j , p_{ij} ($\sum_i p_{ij} = 1$), creates the following matrix

$$P = \begin{pmatrix} p_{11} & \cdots & p_{1n} \\ \vdots & \ddots & \vdots \\ p_{s1} & \cdots & p_{sn} \end{pmatrix}. \quad (47)$$

Changing the size of spatial extent and/or spatial grain changes the size of the matrix P , since a smaller spatial extent may hold a fewer number of species, and a fine spatial grain increases the number of sampling patches. Therefore, if we change either one or both of these scales, the matrix P is changed into another matrix P' . We describe this operation as $P \mapsto P'$. For example, if Eq. (47) is the $s \times n$ matrix with equal-sized patches and n is an even number, and if its spatial grain is doubled, the operation $P \mapsto P'$ gives a $s \times n/2$ matrix.

In practice, each patch has a different significance on a diversity index, and this is described by the weight vector $\mathbf{w} = (w_1, w_2, \dots, w_n)$. We can also define the same operation for the weight vector associated with a scale change.

Figure 9: Schematic figure of spatial extent and spatial grain. Colors indicate the abundance of all the species within the spatial extent 128km×128km (left) and 256km×256km (right). Spatial grain is 1km×1km (bottom) and 4km×4km (top).

5.2 Diversity indices

Statistical discussions over its reliable definitions given a huge number of different definitions [33] and bridge between different definitions has been actively discussed [32, 34–37]. Here we adapt the definition of Jost [32]. Jost [32] showed that when weights in diversity indices are unequal, meaningful diversity index is only the Shannon measure since this only satisfies five requisite conditions for diversity indices. The Shannon measure is described by

$${}^1D_\alpha = \exp\left\{-w_1 \sum_i^s p_{i1} \log(p_{i1}) - w_2 \sum_i^s p_{i2} \log(p_{i2}) \cdots - w_n \sum_i^s p_{in} \log(p_{in})\right\}, \quad (48a)$$

$${}^1D_\gamma = \exp\left\{-\sum_i^s (w_1 p_{i1} + w_2 p_{i2} + \cdots + w_n p_{in}) \log(w_1 p_{i1} + w_2 p_{i2} + \cdots + w_n p_{in})\right\}, \quad (48b)$$

$${}^1D_\beta = \frac{{}^1D_\gamma}{{}^1D_\alpha}, \quad (48c)$$

In order to discuss the effect of a scale change of the spatial grain consistently, we further need a condition for independence of a choice of weights on the gamma diversity Eq. (48b). That is to say, scale changes of the spatial grain, provided a consistent spatial extent, should not affect the gamma diversity. It is easy to see this requisite is satisfied if we define the weight vector by population abundance of each patch, $w_j = \sum_i N_{ij} / \sum_{i,j} N_{ij}$, where N_{ij} is the abundance of species i in patch j . Furthermore, this is the only choice to hold this condition (Theorem 1).

Since β diversity depends on the number of spatial grains, we require normalize the β diversity onto $[0, 1]$ to make a meaningful comparison of spatial variations [32]. When all the weighting terms are not identical, the regional homogeneity measure $(1/{}^1D_\beta - 1/{}^1D_w)/(1 - 1/{}^1D_w)$ [32] may be used for the normalized measure of β diversity, where 1D_w is the Shannon measure of weighting term ${}^1D_w = \exp(-\sum_j w_j \log(w_j))$. We use its complement

$$1 - \frac{1/{}^1D_\beta - 1/{}^1D_w}{1 - 1/{}^1D_w}, \quad (49)$$

as a relative inhomogeneity measure. This measures is 0 if all the communities are identical and 1 if all the communities are distinct.

5.3 Normalized β diversity across spatial extent and grains

To see how the normalized β diversity (inhomogeneity measure) changes, we compute Eq. (49) across spatial extents and spatial grains. We choose the range of spatial extent to cover three different phases in the SAR (Fig. 4) ($2^{-6}\text{km}^2 - 2^{16}\text{km}^2$), and we divided the spatial extent into equal-sized subregion with spatial grain size $2^{-8}\text{km}^2 - 2^{s-2}\text{km}^2$, where s determines the size of spatial extent, so that each scenario has minimum 4 subregions. To compute inhomogeneity measure across spatial extent and spatial grains, we used only one realization of point patterns by taking the following three steps: (i) define point patterns in

the maximum spatial extent (e.g., defined on the plane $[0, 2^8] \times [0, 2^8]$); (ii) define a spatial extent (e.g., on the plane $[0, 2^{-1}] \times [0, 2^{-1}]$); and (iii) calculate the inhomogeneity measure for each spatial grain with area $(2^{-8}, 2^{-6}, 2^{-4})$. In the step (ii), we started from the minimum spatial extent 2^{-8}km^2 , and repeated the step (ii) and (iii) until spatial extent reached the maximum extent 2^{16}km^2 . We eliminate the situation where no individual exists when the minimum spatial grain size 2^{-8}km^2 is applied to make sure each scenario contains at least one individual. When all individuals situated in one subregion, we define the inhomogeneity measure as 0.

Fig. 10 shows the numerically calculated inhomogeneity measure averaged over 500 simulation trials, associated with the SAR curves of underlying processes. Overall, the underlying processes under concern does not cause prominent qualitative effect. The top two figures show heat map of the inhomogeneity measure under the (a) Poisson and (b) Thomas processes. In both panels, the inhomogeneity measure shows higher value as the spatial extent increases. Under this operation, the effect of applying different spatial grain becomes small, especially the SAR is in the third phase (the bottom two figures with fixed spatial grains 2^{-8}km^2 , 2^{-6}km^2). Conversely, the inhomogeneity measure shows small value when spatial extent underlies in the left-side of the second phase of of the SAR, and its spatial grain is large.

6 Discussion

We develop a novel framework to derive macroecological patterns across scales by explicitly linking the distribution of individuals within ranges, the size of ranges, and the spatial extent of the sampling region. The model phenomenologically describes species and individual distributions and in the derivations the SAR, EAR, and RSA, no preceding assumptions about these relationships are required. Rather, the model requires a single set of parameters to generate macroecological patterns across scales. Although the model does not explicitly assume specific biological mechanisms such as population or community dynamics, dispersal, and speciation, it still recovers several well-known macroecological patterns including the tri-phasic SAR and Fisher's logseries [2], the Poisson gamma distribution (negative binomial distribution) [28], Poisson lognormal distribution [29], and similar form to a negatively skewed lognormal distribution [6] as RSAs.

This finding does not imply that the biological mechanisms shaping biodiversity pattern are irrelevant or uninteresting. Rather, our theory demonstrates the minimum assumptions are sufficient to recover such ubiquitous ecological patterns by linking pattern in individuals and species to aggregate macroecological patterns. The latter can be useful as mechanistic theory development as well, as once a mechanistic model has predicted intraspecific point patterns and/or geographic range geometry, our theory can link them to macroecological patterns. That said, it is clear the general forms of commonly-studied macroecological patterns investigated here are general features of biodiversity pattern emerging from the most basic assumptions, and are not indicative of specific ecological processes to the exclusion of others.

Figure 10: Inhomogeneity measure averaged over 500 simulation trials, associated with the SAR curves of underlying (a) Poisson processes and (b) Thomas process. Top two figures show heat map of the inhomogeneity measure across spatial extent and spatial grain. Bottom two figures show some slices of the top figures, where fixed spatial grains (2^{-8}km^2 , 2^{-6}km^2) are applied while spatial extent varies. Same parameter values are used as the results of the SARs $E[r_d] = 10^1\text{km}^2$ Poisson/Thomas shown in Fig. 4.

We showed that the tri-phasic SAR with and its asymptotic slope on a log-log plot are ubiquitous patterns in the geometric model, emerging all the scenarios examined (Fig. 4a) that encompasses random and cluster individual distribution patterns, and neutral and non-neutral ecological parameters. It can be discussed in line with the findings of Grilli et al. [11] that the tri-phasic SAR is the outcome associated with two bending points in the SAR produced by local and large sampling area that scales the species number provided by a simple geometric realization. Especially, we showed that the third phase in the SAR appears around the average area of the species distribution ranges. This may correspond to the biological interpretations that sampling area exceeds correlation distance of biogeographic process [5, 6]. Notably, we found a negligible difference in the SAR between the neutral situation and a quasi-neutral situation where only the individual intensity (λ_i) varies between species. However, this variation is necessary to derive well-documented RSAs discussed above. This fact suggests that only small effect occurs on non-zero probability that is used

in the definition of the SAR (Eq. 8) given underlying ecological assumptions. In addition, we demonstrated that the spatial scaling of beta diversity can be discussed with the tri-phasic features of the SAR curve. Typically the normalized beta diversity (inhomogeneity measure) increases with the spatial extent, and it shows largest value in the third phase of the SAR where the spatial extent exceeds the correlation length of biogeographic process.

In our framework, the EAR is calculated by explicitly taking distribution ranges into account. Under the definition, the EAR on a log-log plot has the same asymptotic slope 1 as the SAR. This result cannot be directly compared to previous studies [10, 11, 16] where the endemic species is calculated based on the probability to find a species given region where the probability asymptotically approaches 0 as the sampling region approaches 0. On the other hand we apply more straightforward definition: the expected number of species distribution range enclosed completely by the sampling region. Therefore, our definition gives 0 endemic species unless the scale of sampling region exceeds the scale of species distribution range. Asymptotic behavior at small sampling scales can also occur in our definition by assuming a pdf for the species distribution ranges where rather small distribution ranges allow to exist. Although this latter assumption may not be realistic but may be used to discuss about the regional EAR as discussed in He and Hubbell [6]. Difference between the two definitions appear especially in small sampling scales. In fact, Grilli et al. [11] showed that the slope of the EAR converges to the slope of SAR at large scales that agrees with our discussion at large sampling scales.

Importantly, we provide a single equation describing the RSA with an arbitrary sampling scale (Eqs. 13 and 45). The equation is composed of two mixed-probability distributions where their weights are determined by geometries and scales of the sampling region and species distribution range. Specifically, one of the distribution accounts effects of intra- and inter- species variations in an equal-sized region, and the other distribution accounts, on top of these variations, variations of sampled distribution range sizes. In the latter distribution the effect of sampling size is considered: a species that sampled only partially shows smaller abundance than the case that the species is fully sampled. This latter variation causes left-tailed RSAs as some species are sampled only small regions and others are fully sampled (Fig. 7c-f). Once the sampled region becomes sufficiently large enough, the species sampled only partially becomes negligible and the left tail is vanished and it results in lognormal like distribution (Fig. 8) provided each species has sufficiently large abundance. Lognormal or lognormal-like has been claimed as the potential RSA (SAD) at global scales in previous works [38–40]. Specifically, Eq. (45) is the weighted sum of the Poisson-gamma distribution and Poisson-gamma-arcsine distribution. Notably, this single equation does not provide a single pdf corresponding to the RSA but generates multiple pdfs. This property may provide a new insight into a long standing discussion in ecological community about a fundamental RSA pattern of ecosystems, where a number of attempts has tried to fit a single pdf to RSAs across scales (e.g., [8, 28]). In addition, Eqs. (13 and 45) predict that in the two limits of spatial scales $S \ll D$ and $S \gg D$, namely the scales of sampling region and distribution ranges are significantly different, the underlying pdfs are equivalent, but with different parameter values. In practice, RSAs in plant communities are typically discussed

in relatively small scales (i.e., < 50 -ha; e.g. [22, 41]), and this case may be the situation of $S \ll D$, where our results provide Fisher’s logseries, Negative binomial distribution, and Poisson lognormal distribution, depending on the biological assumptions.

Another notable finding is that the prominent variations of different individual distribution patterns (random and cluster) occur only a limited spatial scales in the SAR and RSA. For the SAR, it occurs around the transition points between the first and second phases of the tri-phasic curve (Fig. 4), shows agreement with the results obtained by [22], in which the authors examined the random-placement model and Thomas process to fit an observed SAR patterns up to 50-ha in tropical forests. No, if any, variations between the random and clustered model suggests that there exist spatial scales where some specific biological processes such as the dispersal distance and size of clusters dominate community pattern formations, since these create significantly different geometric patterns within the scope of small scales. However, once the sampling area becomes large enough, e.g., significantly larger than the area dispersal kernel covers, each component of a cluster plays the same role as randomly placed individuals (Note that each parent location placed randomly). Intuitively speaking, if the two geometric patterns generated by the homogeneous Poisson and Thomas process are observed by the scope of such a large scale, it is not easy to say which geometric pattern is which, as long as the number of individuals are the same and the number is large enough to create a certain geometric pattern. This property suggest that random individual distributions can be used to explore the macroecological and community structures with larger spatial scales, and it makes analysis significantly accessible.

In this paper, we developed a theoretical framework to describe macroecological and community patterns across multiple scales with an example for a potential application to a question in community ecology. Other promising applications are for biodiversity conservation, where spatially integrated approaches such as spatial design of reserve networks have been widely discussed [42, 43], and our framework would provide a theoretical exercise of ecosystem management. We minimized the assumptions of the model for its parsimony and keeping analytical tractability. Nonetheless, based on the general theory developed to derive the SAR, EAR, and RSA in an arbitrary situation, our framework can be flexibly extended and examined various ecological assumptions, at least numerically; for example, point processes provide a framework to discuss a heterogeneous environment (e.g., [24]). Rather, one can use any model to generate a point field $f(\mathbf{x})$, and examine the macroecological patterns by using the general theory. Such experiments would provide further insights into macroecological patterns and applicability of the model.

Acknowledgements

We would like to thank Yoh Iwasa for a constructive discussion. NT was supported by Grant-in-Aid for the Japan Society for the Promotion of Science (JSPS) Fellows. Financial support was provided by JSPS (No. 15K14607 to YK and No. 17K15180 to EPE) and Program for Advancing Strategic International Networks to Accelerate the Circulation of Talented Researchers, JSPS (to YK). EPE and NT were additionally supported by subsidy

funding to OIST.

References

- [1] S. A. Levin, The problem of pattern and scale in ecology: the Robert H. MacArthur award lecture, *Ecology* 73 (6) (1992) 1943–1967.
- [2] R. A. Fisher, A. S. Corbet, C. B. Williams, [The Relation Between the Number of Species and the Number of Individuals in a Random Sample of an Animal Population](#), *J. Anim. Ecol.* 12 (1) (1943) 42. doi:10.2307/1411.
URL <http://www.jstor.org/stable/1411?origin=crossref>
- [3] F. Preston, Time and space and the variation of species, *Ecology* 41 (4) (1960) 611–627.
- [4] R. M. May, Patterns of Species Abundance and Diversity, in: M. J. Cody, J. M. Diamond (Eds.), *Ecol. Evol. Communities*, The Belknap Press of Harvard Univ. Press, Cambridge, MA and London, 1975, pp. 81–120.
- [5] M. Rosenzweig, *Species diversity in space and time*, Cambridge University Press, Cambridge, UK, 1995.
- [6] S. P. Hubbell, *The Unified Neutral Theory of Biodiversity and Biogeography*, Princeton University Press, Princeton, NJ, 2001.
- [7] B. J. McGill, R. S. Etienne, J. S. Gray, D. Alonso, M. J. Anderson, H. K. Benecha, M. Dornelas, B. J. Enquist, J. L. Green, F. He, A. H. Hurlbert, A. E. Magurran, P. A. Marquet, B. A. Maurer, A. Ostling, C. U. Soykan, K. I. Ugland, E. P. White, [Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework](#), *Ecol. Lett.* 10 (10) (2007) 995–1015.
URL <http://doi.wiley.com/10.1111/j.1461-0248.2007.01094.x>
- [8] E. Baldrige, D. J. Harris, X. Xiao, E. P. White, [An extensive comparison of species-abundance distribution models](#), *PeerJ* 4 (2016) e2823. doi:10.7717/peerj.2823.
URL <https://peerj.com/articles/2823>
- [9] D. Storch, A. L. Šizling, J. Reif, J. Polechová, E. Šizlingová, K. J. Gaston, The quest for a null model for macroecological patterns: Geometry of species distributions at multiple spatial scales (2008). doi:10.1111/j.1461-0248.2008.01206.x.
- [10] D. Storch, P. Keil, W. Jetz, [Universal speciesarea and endemicsarea relationships at continental scales](#), *Nature* 488 (7409) (2012) 78–81. doi:10.1038/nature11226.
URL <http://www.nature.com/doifinder/10.1038/nature11226>
- [11] J. Grilli, S. Azaele, J. R. Banavar, A. Maritan, Spatial aggregation and the species-area relationship across scales, *J. Theor. Biol.* 313 (2012) 87–97. doi:10.1016/j.jtbi.2012.07.030.

- [12] S. Azaele, A. Maritan, S. J. Cornell, S. Suweis, J. R. Banavar, D. Gabriel, W. E. Kunin, Towards a unified descriptive theory for spatial ecology: Predicting biodiversity patterns across spatial scales, *Methods Ecol. Evol.* 6 (3) (2015) 324–332. doi:[10.1111/2041-210X.12319](https://doi.org/10.1111/2041-210X.12319).
- [13] P. S. Barton, S. A. Cunningham, A. D. Manning, H. Gibb, D. B. Lindenmayer, R. K. Didham, The spatial scaling of beta diversity, *Glob. Ecol. Biogeogr.* 22 (6) (2013) 639–647. doi:[10.1111/geb.12031](https://doi.org/10.1111/geb.12031).
- [14] S. L. Pimm, C. N. Jenkins, R. Abell, T. M. Brooks, J. L. Gittleman, L. N. Joppa, P. H. Raven, C. M. Roberts, J. O. Sexton, [The biodiversity of species and their rates of extinction, distribution, and protection](#), *Science* 344 (6187) (2014) 1246752–1246752. doi:[10.1126/science.1246752](https://doi.org/10.1126/science.1246752).
URL <http://www.sciencemag.org/cgi/doi/10.1126/science.1246752>
- [15] J. E. Watson, D. F. Shanahan, M. Di Marco, J. Allan, W. F. Laurance, E. W. Sanderson, B. Mackey, O. Venter, Catastrophic Declines in Wilderness Areas Undermine Global Environment Targets, *Curr. Biol.* 26 (21) (2016) 2929–2934. doi:[10.1016/j.cub.2016.08.049](https://doi.org/10.1016/j.cub.2016.08.049).
- [16] F. He, S. P. Hubbell, Species-area relationships always overestimate extinction rates from habitat loss, *Nature* 473 (7347) (2011) 368–371.
- [17] C. R. Margules, R. L. Pressey, Systematic conservation planning., *Nature* 405 (6783) (2000) 243–53. doi:[10.1038/35012251](https://doi.org/10.1038/35012251).
- [18] N. Takashina, M. Baskett, Exploring the effect of the spatial scale of fishery management, *J. Theor. Biol.* 390. doi:[10.1016/j.jtbi.2015.11.005](https://doi.org/10.1016/j.jtbi.2015.11.005).
- [19] N. Takashina, M. Beger, B. Kusumoto, S. Rathnayake, H. Possingham, A theory for ecological survey methods to map individual distributions, *Theor. Ecol.* doi:<https://doi.org/10.1007/s12080-017-0359-7>.
- [20] N. Takashina, M. Beger, B. Kusumoto, H. Possingham, Population assessment to map the distribution of individuals, *bioRxiv* doi:<https://doi.org/10.1101/131037>.
- [21] A. P. Allen, E. P. White, Effects of range size on species-area relationships, *Evol. Ecol. Res.* 5 (4) (2003) 493–499. doi:[10.1073/pnas.0500424102](https://doi.org/10.1073/pnas.0500424102).
- [22] J. B. Plotkin, M. D. Potts, N. Leslie, N. Manokaran, J. Lafrankie, P. S. Ashton, Species-area curves, spatial aggregation, and habitat specialization in tropical forests., *J. Theor. Biol.* 207 (1) (2000) 81–99. doi:[10.1006/jtbi.2000.2158](https://doi.org/10.1006/jtbi.2000.2158).
- [23] J. Illian, A. Penttinen, H. Stoyan, D. Stoyan, *Statistical Analysis and Modelling of Spatial Point Patterns*, John Wiley & Sons, Ltd, Chichester, UK., 2008. doi:[10.1002/9780470725160](https://doi.org/10.1002/9780470725160).

- [24] S. N. Chiu, D. Stoyan, W. S. Kendall, J. Mecke, Stochastic Geometry and Its Applications, John Wiley & Sons, New York, 2013.
- [25] S. Azaele, S. J. Cornell, W. E. Kunin, Downscaling species occupancy from coarse spatial scales, *Ecol. Appl.* 22 (3) (2012) 1004–1014. doi:[10.1890/11-0536.1](https://doi.org/10.1890/11-0536.1).
- [26] B. D. Coleman, On random placement and species-area relations, *Math. Biosci.* 54 (3-4) (1981) 191–215. doi:[10.1016/0025-5564\(81\)90086-9](https://doi.org/10.1016/0025-5564(81)90086-9).
- [27] D. Karlis, E. Xekalaki, Mixed poisson distributions, *Int. Stat. Rev.* 73 (1) (2005) 35–58. doi:[10.1111/j.1751-5823.2005.tb00250.x](https://doi.org/10.1111/j.1751-5823.2005.tb00250.x).
- [28] S. R. Connolly, M. A. MacNeil, M. J. Caley, N. Knowlton, E. Cripps, M. Hisano, L. M. Thibaut, B. D. Bhattacharya, L. Benedetti-Cecchi, R. E. Brainard, A. Brandt, F. Bulleri, K. E. Ellingsen, S. Kaiser, I. Kroncke, K. Linse, E. Maggi, T. D. O'Hara, L. Plaisance, G. C. B. Poore, S. K. Sarkar, K. K. Satpathy, U. Schuckel, A. Williams, R. S. Wilson, Commonness and rarity in the marine biosphere, *Proc. Natl. Acad. Sci.* 111 (23) (2014) 8524–8529. doi:[10.1073/pnas.1406664111](https://doi.org/10.1073/pnas.1406664111).
URL <http://www.pnas.org/cgi/doi/10.1073/pnas.1406664111>
- [29] M. G. Bulmer, On Fitting the Poisson Lognormal Distribution to Species-Abundance Data, *Biometrics* 30 (1) (1974) 101. doi:[10.2307/2529621](https://doi.org/10.2307/2529621).
URL <http://www.jstor.org/stable/2529621?origin=crossref>
- [30] B. Dennis, G. P. Patil, Applications in ecology, in: E. L. Crow, K. Shimizu (Eds.), Lognormal Distrib. theory Appl., Markel Dekker, Inc, New York, USA, 1988, pp. 303–330.
- [31] Arcsine distribution. Encyclopedia of Mathematics. URL: http://www.encyclopediaofmath.org/index.php?title=Arcsine_distribution&oldid=33530
Data Accessed: Nov 26 2017.
URL <http://www.encyclopediaofmath.org/index.php?title=Arcsine{distribution}&oldid=33530>
- [32] L. Jost, Partitioning diversity into independent alpha and beta components, *Ecology* 88 (10) (2007) 2427–2439. doi:[10.1890/06-1736.1](https://doi.org/10.1890/06-1736.1).
- [33] M. J. Anderson, T. O. Crist, J. M. Chase, M. Vellend, B. D. Inouye, A. L. Freestone, N. J. Sanders, H. V. Cornell, L. S. Comita, K. F. Davies, S. P. Harrison, N. J. Kraft, J. C. Stegen, N. G. Swenson, Navigating the multiple meanings of β diversity: A roadmap for the practicing ecologist, *Ecol. Lett.* 14 (1) (2011) 19–28. doi:[10.1111/j.1461-0248.2010.01552.x](https://doi.org/10.1111/j.1461-0248.2010.01552.x).
- [34] L. Jost, Entropy and diversity, *Oikos* 113 (2) (2006) 363–375. doi:[10.1111/j.2006.0030-1299.14714.x](https://doi.org/10.1111/j.2006.0030-1299.14714.x).

- [35] A. Chao, C. H. Chiu, T. C. Hsieh, B. D. Inouye, Proposing a resolution to debates on diversity partitioning, *Ecology* 93 (9) (2012) 2037–2051. [doi:10.1890/11-1817.1](https://doi.org/10.1890/11-1817.1).
- [36] C. H. Chiu, L. Jost, A. Chao, Phylogenetic beta diversity, similarity, and differentiation measures based on Hill numbers, *Ecol. Monogr.* 84 (1) (2014) 21–44. [doi:10.1890/12-0960.1](https://doi.org/10.1890/12-0960.1).
- [37] A. Chao, C. H. Chiu, Bridging the variance and diversity decomposition approaches to beta diversity via similarity and differentiation measures, *Methods Ecol. Evol.* 7 (8) (2016) 919–928. [doi:10.1111/2041-210X.12551](https://doi.org/10.1111/2041-210X.12551).
- [38] R. D. Gregory, Abundance patterns of European breeding birds, *Ecography* (Cop.). 23 (2) (2000) 201–208. [doi:10.1111/j.1600-0587.2000.tb00276.x](https://doi.org/10.1111/j.1600-0587.2000.tb00276.x).
- [39] B. J. McGill, Does Mother Nature really prefer rare species or are log-left-skewed SADs a sampling artefact?, *Ecol. Lett.* 6 (8) (2003) 766–773. [doi:10.1046/j.1461-0248.2003.00491.x](https://doi.org/10.1046/j.1461-0248.2003.00491.x).
- [40] J. Rosindell, S. J. Cornell, S. P. Hubbell, R. S. Etienne, Protracted speciation revitalizes the neutral theory of biodiversity, *Ecol. Lett.* 13 (6) (2010) 716–727. [doi:10.1111/j.1461-0248.2010.01463.x](https://doi.org/10.1111/j.1461-0248.2010.01463.x).
- [41] R. Condit, P. S. Ashton, P. Baker, S. Bunyavejchewin, S. Gunatilleke, N. Gunatilleke, S. P. Hubbell, R. B. Foster, A. Itoh, J. V. LaFrankie, H. S. Lee, E. Losos, N. Manokaran, R. Sukumar, T. Yamakura, Spatial patterns in the distribution of tropical tree species, *Science* 288 (1982) (2000) 1414–1418. [doi:10.1126/science.288.5470.1414](https://doi.org/10.1126/science.288.5470.1414).
- [42] H. Possingham, I. Ball, S. Andelman, Mathematical methods for identifying representative reserve networks, in: S. Ferson, M. Burgman (Eds.), *Quant. methods Conserv. Biol.*, Springer-Verlag New York, New York, USA, 2000, pp. 291–305.
- [43] M. Beger, J. McGowan, E. A. Treml, A. L. Green, A. T. White, N. H. Wolff, C. J. Klein, P. J. Mumby, H. P. Possingham, Integrating regional conservation priorities for multiple objectives into national policy, *Nat. Commun.* 6. [doi:10.1038/ncomms9208](https://doi.org/10.1038/ncomms9208).

Appendix

Here we describe some propositions which are used in the main text. In the main text, the description $\nu(A)$ used below is replaced by $\min\{\nu(S), \nu(D)\}$.

Proposition 1. Provided that the moments of the mixing distribution in a mixed Poisson model exist, the probability function of the mixture distribution can be written as

$$P(X = x) = \frac{1}{x!} \sum_{r=0}^{\infty} \nu(A)^{x+r} \frac{(-1)^r}{r!} \mu_{x+r}(\lambda), \quad (\text{A.1})$$

where, $\mu_r(\lambda)$ are the r th moment of λ about the origin.

Proof. The proof is straightforward from the definition and similar result is found in [27]:

$$\begin{aligned}
 P(X = x) &= \int_0^\infty \frac{(\lambda \nu(A))^x}{x!} e^{-\lambda \nu(A)} g(\lambda) d\lambda, \\
 &= \frac{1}{x!} \int_0^\infty \left(\sum_{r=0}^\infty (-1)^r \frac{\lambda^r \nu(A)^r}{r!} \right) \lambda^x \nu(A)^x g(\lambda) d\lambda, \\
 &= \frac{1}{x!} \sum_{r=0}^\infty \nu(A)^{x+r} \frac{(-1)^r}{r!} \int_0^\infty \lambda^{x+r} g(\lambda) d\lambda, \\
 &= \frac{1}{x!} \sum_{r=0}^\infty \nu(A)^{x+r} \frac{(-1)^r}{r!} \mu_{x+r}(\lambda).
 \end{aligned}$$

□

Proposition 2. Provided that the moments of the mixing distribution in a mixed Poisson model exist, the probability function of the mixture distribution can be written as

$$P(X = x) = \frac{1}{x!} \sum_{r=0}^\infty \frac{(-1)^r}{r!} \mu_{x+r}(\lambda) \mu_{x+r}(\nu(A)), \quad (\text{A.2})$$

where, $\mu_{x+r}(\lambda)$ and $\mu_{x+r}(\nu(A))$ are the r th moment of λ and $\nu(A)$ about the origin, respectively.

Proof. The proof is just an extension of the proposition 1.

$$\begin{aligned}
 P(X = x) &= \int_0^\infty \int_0^\infty \frac{(\lambda \nu(A))^x}{x!} e^{-\lambda \nu(A)} g(\lambda) h(\nu(A)) d\lambda d\nu(A), \\
 &= \frac{1}{x!} \int_0^\infty \int_0^\infty \sum_{r=0}^\infty (-1)^r \left(\frac{\lambda^r \nu(A)^r}{r!} \right) \lambda^x g(\lambda) \nu(A)^x h(\nu(A)) d\lambda d\nu(A), \\
 &= \frac{1}{x!} \sum_{r=0}^\infty \frac{(-1)^r}{r!} \int_0^\infty \lambda^{x+r} g(\lambda) d\lambda \int_0^\infty \nu(A)^{x+r} h(\nu(A)) d\nu(A), \\
 &= \frac{1}{x!} \sum_{r=0}^\infty \frac{(-1)^r}{r!} \mu_{x+r}(\lambda) \mu_{x+r}(\nu(A)).
 \end{aligned}$$

□

Proposition 3. The r th moment of the arcsine distribution with support $x \in [0, \nu(A)]$

$$f(x) = \frac{1}{\pi \sqrt{x(\nu(A) - x)}}, \quad (\text{A.3})$$

is described by

$$\mu_r^{arc} = \frac{\nu(A)^r}{\pi} B\left(\frac{1}{2} + r, \frac{1}{2}\right), \quad (\text{A.4})$$

where, $B(x, y) = \int_0^1 u^{x-1}(1-u)^{y-1} du$ is the beta function.

Proof. Using the substitution $w = x/\nu(A)$,

$$\begin{aligned} \mu_r &= \frac{1}{\pi} \int_0^{\nu(A)} x^r x^{-\frac{1}{2}} (\nu(A) - x)^{-\frac{1}{2}} dx, \\ &= \frac{\nu(A)^r}{\pi} \int_0^1 w^{r+\frac{1}{2}-1} (1-w)^{\frac{1}{2}-1} dw, \\ &= \frac{\nu(A)^r}{\pi} B\left(\frac{1}{2} + r, \frac{1}{2}\right). \end{aligned}$$

□

Theorem 1. *The gamma diversity defined by Eq. (48b) that is independent of the choice of the number of patches or its size is uniquely determined, and it is when the weight vector is proportional to the population abundance of each patch.*

Proof. Let us assume there exists a vector $\mathbf{w}' = (w'_1, w'_2, \dots, w'_n)$ that satisfies ${}^1D_\gamma^{\mathbf{w}} = {}^1D_\gamma^{\mathbf{w}'}$, where superscripts $\mathbf{w}$ and $\mathbf{w}'$ indicate weight vector used. By the assumption, we have $\sum_i^s (w_1 p_{i1} + w_2 p_{i2} + \dots + w_n p_{in}) \log(w_1 p_{i1} + w_2 p_{i2} + \dots + w_n p_{in}) = \sum_i^s (w'_1 p_{i1} + w'_2 p_{i2} + \dots + w'_n p_{in}) \log(w'_1 p_{i1} + w'_2 p_{i2} + \dots + w'_n p_{in})$ and arranging this expression, it becomes $\sum_{i,j} p_{ij} \{w_j \log(w_1 p_{i1} + w_2 p_{i2} + \dots + w_n p_{in}) - w'_j \log(w'_1 p_{i1} + w'_2 p_{i2} + \dots + w'_n p_{in})\} = 0$. Since p_{ij} may or may not be zero we must have $w_j \log(w_1 p_{i1} + w_2 p_{i2} + \dots + w_n p_{in}) = w'_j \log(w'_1 p_{i1} + w'_2 p_{i2} + \dots + w'_n p_{in})$ for all i and j . This is clearly $w_j = w'_j$. □

Derivation of the Poisson-gamma-arcsine distribution (Eq. 44)

Now, using Proposition 3 and the r th moment of the gamma distribution $f(\lambda) = \beta^\alpha / \Gamma(\alpha) \lambda^{\alpha-1} e^{-\lambda\beta}$

$$\mu_r^{gam} = \frac{\Gamma(r+\alpha)}{\Gamma(\alpha)} \frac{1}{\beta^r}, \quad (\text{A.5})$$

we calculate the probability of mixing distribution Eq. (A.2) with properties of the beta function $B(x, y) = \Gamma(x)\Gamma(y)/\Gamma(x+y)$ and the Gamma function $\Gamma(1/2) = \sqrt{\pi}$:

$$\begin{aligned} P(X=x) &= \frac{1}{x!} \sum_{r=0}^{\infty} \frac{(-1)^r}{r!} \frac{\Gamma(r+x+\alpha)}{\Gamma(\alpha)} \frac{1}{\beta^{x+r}} \frac{\nu(A)^{r+x}}{\pi} B\left(\frac{1}{2} + r + x, \frac{1}{2}\right), \\ &= \frac{1}{x!} \left(\frac{\nu(A)}{\beta}\right)^x \frac{1}{\sqrt{\pi}\Gamma(\alpha)} \sum_{r=0}^{\infty} \frac{(-\nu(A)/\beta)^r}{r!} \frac{\Gamma(r+x+\alpha)\Gamma(\frac{1}{2} + r + x)}{\Gamma(1+r+x)}, \\ &= \frac{1}{x!} \left(\frac{\nu(A)}{\beta}\right)^x \frac{1}{\sqrt{\pi}\Gamma(\alpha)} \frac{\Gamma(x+\alpha)\Gamma(\frac{1}{2} + x)}{\Gamma(1+x)} {}_2F_1\left(\frac{1}{2} + x, x + \alpha; 1 + x; -\nu(A)/\beta\right), \\ &= a \binom{x+\alpha-1}{x} p^x (1-p)^\alpha {}_2F_1\left(\frac{1}{2} + x, x + \alpha; 1 + x; -\nu(A)/\beta\right), \end{aligned} \quad (\text{A.6})$$

where, a is $\Gamma(1/2+x)(\sqrt{\pi}\Gamma(x+1))^{-1}(1-p)^{-\alpha-x}$ and is simplified for $x \geq 1$ as $x^{-1}B(1/2, x)^{-1}(1-p)^{-\alpha-x}$, the second to forth factors correspond to the negative binomial distribution with $p = \nu(A)/(\nu(A) + \beta)$, and ${}_2F_1$ is the Gauss hypergeometric function. α and β are the shape and rate parameters of the gamma distribution. Note the negative binomial distribution here is equivalent to Eq. (38).

Poisson-lognormal-arcsine distribution

By substituting the r th moment about the origin of the arcsine distribution Eq. (A.4), and the Poisson-lognormal distribution $\mu_r^{pl} = \exp(r\mu + r^2\sigma^2/2)$ [29] into Eq. (35), we obtain the RSA of an arbitrary form

$$P(X = x) = \frac{\min\{\nu(S), \nu(D)\}^x}{x!\sqrt{\pi}} \sum_{r=0}^{\infty} \frac{(-\min\{\nu(S), \nu(D)\})^r}{r!} \frac{\Gamma(\frac{1}{2} + r + x)}{\Gamma(1 + r + x)} e^{(x+r)\mu + \frac{(x+r)^2}{2}\sigma^2} \quad (\text{A.7})$$

where as in the case of the Poisson-gamma-arcsine distribution, we call this pdf the Poisson-lognormal-arcsine distribution. Using Eqs. (40), (42), and (A.7), we obtain the full form of the RSA across scales

$$P(X = x) = \frac{\max\{r_d - r_s, r_s - r_d\}}{r_d + r_s} \frac{1}{x!\sqrt{2\pi\sigma^2}} \int_0^{\infty} \lambda_i^{x-1} e^{-\lambda_i - \frac{(\log \lambda_i - \mu)^2}{2\sigma^2}} d\lambda_i + \quad (\text{A.8})$$

$$\frac{\min\{2r_s, 2r_d\}}{r_d + r_s} \frac{\min\{\nu(S), \nu(D)\}^x}{x!\sqrt{\pi}} \sum_{r=0}^{\infty} \frac{(-\min\{\nu(S), \nu(D)\})^r}{r!} \frac{\Gamma(\frac{1}{2} + r + x)}{\Gamma(1 + r + x)} e^{(x+r)\mu + \frac{(x+r)^2}{2}\sigma^2}.$$

As above, Eq. (A.8) becomes the Poisson-lognormal distribution in the limits of $S \ll D$ and $S \gg D$:

$$P(X = x) = \frac{1}{x!\sqrt{2\pi\sigma^2}} \int_0^{\infty} \lambda_i'^{x-1} e^{-\lambda_i' - \frac{(\log \lambda_i' - \mu')^2}{2\sigma^2}} d\lambda_i', \quad (\text{A.9})$$

where, by setting $\mu = \mu' - \log(\min\{\nu(S), \nu(D)\})$ and $\min\{\nu(S), \nu(D)\}\lambda_i = \lambda_i'$, we recover the form of Eq. (40). Therefore, if the parameter λ_i follows the Log-normal distribution and sampling region is much smaller than the distribution range, we expect to observe an RSA curve that follows Poisson-lognormal distribution.