Article

Effective discrimination of spatial distribution patterns of rare species: Methodological reflections, theoretical framework, and empirical strategies beyond the binary dilemma of Aggregated *vs*. Random

WenJun Zhang
School of Life Sciences, Sun Yat-sen University, Guangzhou 510275, China
E-mail: wjzhang@iaees.org, zhwj@mail.sysu.edu.cn

Received 2 March 2026; *Accepted 25 June 2026*; *Published online 12 July 2026*; *Published 1 March 2027*

Abstract
The accurate identification of spatial distribution patterns for rare species constitutes a fundamental challenge in community ecology and conservation biology. Rare species typically exhibit stronger spatial aggregation than common species, yet their low abundance renders standard aggregation indices unreliable, creating a persistent difficulty in distinguishing true ecological signals from statistical artifacts. This study systematically examines the methodological limitations that produce the core dilemma: the binary classification of patterns as aggregated or random fails to account for the profound uncertainty inherent in small-sample inference. I develop an innovative three-dimensional analytical framework for the effective discrimination of spatial patterns in rare species, integrating multidimensional rarity typology, scale continuity of pattern, and explicit quantification of estimation uncertainty. Central to this framework is the concept of the spatial pattern continuum, which replaces discrete pattern categories, and the introduction of pattern identifiability as a measurable property derived from the confidence intervals of aggregation indices. I formulate statistical decision rules based on whether the confidence interval of an estimated aggregation index fully exceeds or encompasses critical threshold values, and define a statistically credible pattern only when such criteria are met. Furthermore, I derive the theoretical concept of minimum distinguishable sample size, offering a direct quantitative link between sampling effort and the resolvability of spatial patterns. A comprehensive analytical methodology is proposed, combining bootstrap and Bayesian uncertainty estimation, multiscale pattern fingerprinting, and a five-step decision workflow. By moving beyond the traditional aggregated versus random dichotomy and embracing a rigorous treatment of uncertainty, this work establishes a new standard for the inference of spatial processes governing rare species distributions and provides practical guidance for conservation monitoring design.

Keywords rare species; spatial distribution pattern; aggregation index; scale dependence; estimation uncertainty; effective discrimination.

Computational Ecology and Software
ISSN 2220-721X
URL: http://www.iaees.org/publications/journals/ces/online-version.asp
RSS: http://www.iaees.org/publications/journals/ces/rss.xml
E-mail: ces@iaees.org
Editor-in-Chief: WenJun Zhang
Publisher: International Academy of Ecology and Environmental Sciences

1 Introduction

1.1 Research Background and Problem Statement

Rare species form a dominant fraction of species in virtually all biological communities and contribute disproportionately to the functional diversity and resilience of ecosystems (Lyons et al., 2005; Gaston, 1994). Understanding the spatial distribution of these species is critical, because spatial pattern reflects the integrated outcome of key ecological processes such as dispersal limitation, environmental filtering, and biotic interactions (Shang et al., 1991; Zhang and Shang, 1992; Condit et al., 2000; Seidler and Plotkin, 2006; Zhang, 1992, 1993, 2007a-b; Liu and Zhang, 2011). Numerous empirical studies have demonstrated that rare species tend to be more spatially aggregated than common species (Plotkin et al., 2000; He and Legendre, 2002). However, the very property that defines rarity, namely a small number of individuals at local or regional scales, creates a severe methodological impasse. Standard indices for quantifying aggregation, such as the variance to mean ratio, the negative binomial parameter k, Morisita's index, or summary functions from spatial point pattern analysis, lose statistical power and precision when sample sizes are small (Taylor, 1961; Zhang and Shang, 1992; Perry, 1995; Dale and Fortin, 2005; Zhang, 1992, 1993, 2007a). Consequently, the distinction between an intrinsically aggregated pattern and a random pattern generated by stochastic demographic noise becomes uncertain. The central research question of this paper is: how can we effectively discriminate the true spatial distribution pattern of a rare species from statistical artifacts, given the inherent constraints of low abundance?

1.2 Research Aims and Significance

This study aims to construct a coherent theoretical and methodological framework for the effective discrimination of spatial distribution patterns in rare species. The theoretical significance lies in refining our understanding of community assembly mechanisms: if the spatial pattern of rare species cannot be reliably identified, inferences about the relative roles of deterministic niche processes versus stochastic dispersal and ecological drift remain questionable (Hubbell, 2001; Chase and Myers, 2011). The methodological significance is the development of a systematic analytical workflow that explicitly integrates estimation uncertainty into pattern discrimination, moving beyond the prevailing practice of reporting only point estimates of aggregation indices. The applied significance is a direct contribution to conservation biology, providing clear guidelines for the sampling design of rare species monitoring programs and for communicating the reliability of species distribution predictions to managers (MacKenzie and Royle, 2005; Jeliazkov et al., 2022).

1.3 Structure of Present Study

The paper is organized as follows. Section 2 provides a comprehensive literature review on the definition of rarity, theoretical foundations of spatial pattern formation, analytical methods, and unresolved challenges. Section 3 presents a methodological critique of the current binary classification paradigm and identifies deep theoretical limitations. Section 4 introduces the novel theoretical framework based on the spatial pattern continuum, multidimensional rarity, scale continuity, and estimation uncertainty. Section 5 details the corresponding analytical methodology and decision workflow. Section 6 outlines a validation strategy via simulation and empirical case studies. Section 7 discusses the ecological and conservation implications, limitations, and future directions, followed by conclusions in Section 8.

2 Theoretical Foundations and Methodological Evolution in the Study of Rare Species' Spatial Patterns

2.1 Conceptualization and Typology of Rarity

Rarity has been defined through multiple, nonmutually exclusive dimensions. Gaston (1994) distinguished three primary axes of rarity: geographic range size, habitat specificity, and local population density.

Rabinowitz (1981) combined these dimensions to propose a classification of seven forms of rarity, ranging from species that are geographically widespread, habitat generalist, and locally abundant, to those that are restricted in range, habitat specialists, and locally sparse. This typology has profound implications for spatial pattern analysis because different forms of rarity are likely to arise from distinct ecological mechanisms and thus exhibit distinct spatial signatures (Kunin and Gaston, 1993). Hartley and Kunin (2003) demonstrated that the degree of rarity and the associated extinction risk are scale dependent; a species that appears rare at a fine sampling grain may be common at a broader extent, highlighting the importance of scale in both defining and measuring rarity.

2.2 Theoretical Basis of Spatial Distribution Patterns

The classical typology of spatial patterns comprises three fundamental types: aggregated, random, and uniform (Pielou, 1977; Shang et al., 1991; Zhang and Shang, 1992; Dale, 1999; Condit et al., 2000; Seidler and Plotkin, 2006; Zhang, 1992, 1993, 2007a; Liu and Zhang, 2011). Random patterns are generated when individual positions are independent of each other and of environmental gradients, described by a homogeneous Poisson process. Aggregated patterns, where individuals are clustered together, are the norm in plant and animal communities and can be produced by several mechanisms. Dispersal limitation is a primary cause of small-scale aggregation, as seeds often fall near the parent plant, creating local clusters of conspecific individuals (Shang et al., 1991; Zhang, 1992; Seidler and Plotkin, 2006; Levine and Murrell, 2003). Niche differentiation combined with the occupation of marginal or patchy habitats can also generate aggregation when species are restricted to specific microsites within a heterogeneous landscape (Harms et al., 2001; Zhang, 2007a-b). The environmental filtering of habitat heterogeneity thus imposes an aggregated spatial structure on species that track specific abiotic conditions. Furthermore, asymmetric interspecific interactions, such as competition with dominant species, can confine rare species to spatially restricted refuge zones, enhancing their apparent aggregation (Tilman, 1982).

2.3 Principal Methods for Analyzing Rare Species' Spatial Patterns

A wide array of methods has been developed to quantify spatial pattern, each with specific assumptions and sensitivities.

Spatial point pattern analysis provides a rigorous framework for analyzing mapped point locations. The Ripley's K function and the pair correlation function $g(r)$ are central tools that summarize second-order structure across a continuous range of spatial scales (Ripley, 1977; Baddeley et al., 2015). The K function evaluates the expected number of additional points within distance r of a typical point, standardized by intensity; significant deviation from the theoretical value under complete spatial randomness, indicates aggregation or uniformity. Wiegand and Moloney (2004) emphasized the use of ring-based statistics and null models for ecological applications, addressing common edge-effect corrections and hypothesis testing via Monte Carlo simulation envelopes.

Aggregation indices provide summary scalars that quantify the degree of nonrandomness. The variance to mean ratio is one of the simplest indices, equal to unity under a random Poisson distribution, greater than one under aggregation, and less than one for uniform patterns (Shang et al., 1991; Zhang and Shang, 1992; Southwood and Henderson, 2000; Zhang, 1992, 1993, 2007a). Taylor's power law relates the log variance to the log mean abundance across samples, with the slope parameter serving as a measure of aggregation that is often consistent within taxonomic groups (Taylor, 1961; Zhang, 2007a). Morisita's index of dispersion is designed to be less sensitive to sample size and mean density than the variance to mean ratio and is widely used in ecological studies (Morisita, 1959). Lloyd's index of mean crowding and patchiness offers a biologically intuitive measure of the degree to which individuals experience conspecific crowding (Lloyd, 1967). Perry (1995) developed the spatial analysis by distance indices methodology, which measures

aggregation by clustering of counts into patches and gaps and provides a randomization test for nonrandomness.

Species distribution models have become essential for predicting the geographic ranges of rare species, but their application is fraught with difficulties. Guisan et al. (2006) showed that standard SDMs calibrated with presence absence data perform poorly for rare species because of extreme class imbalance and sparse data. To address this, Breiner et al. (2015) proposed ensembles of small models that build bivariate models and combine them, improving predictive performance for species with few occurrence records. Maximum entropy modeling (Maxent) remains one of the most robust approaches for modeling rare species distributions when only presence data are available (Phillips et al., 2006; Zhang, 2016, 2018). A recent conceptual guide by Jeliazkov et al. (2022) reviewed the entire sampling and modeling pipeline for rare species, advocating for hierarchical modeling and integrated data approaches to explicitly handle detection imperfections and data sparsity.

Multiscale analysis methods are crucial because spatial patterns are inherently scale dependent (Levin, 1992). Additive partitioning of diversity can reveal the relative contributions of different spatial scales to overall species richness (Lande, 1996). For spatial patterns, the behavior of the pair correlation function across a continuous range of radii provides a scale dependent signature; a species may be aggregated at distances up to 20 meters but randomly distributed at larger scales. This scale continuity requires analytical approaches that do not impose arbitrary discrete scales.

2.4 Main Findings and Unresolved Issues

Empirical comparisons between rare and common species consistently report higher aggregation for rare species. In the Barro Colorado Island 50 ha forest plot, Condit et al. (2000) found that most tree species were more aggregated than expected under random distribution, with the degree of aggregation inversely related to abundance. Plotkin et al. (2000) formalized the aggregation-abundance relationship, showing that aggregation as measured by the negative binomial parameter k (Zhang, 2026; Zhang and Qi, 2025) or the pair correlation function decreases as local abundance increases. Magurran and Henderson (2003) explored the excess of rare species in species abundance distributions and found that many rare species are transient or form temporary aggregations linked to specific recruitment events.

A central unresolved issue remains the statistical reliability of aggregation estimates for species with very few individuals. The sampling variance of aggregation indices such as k or Morisita's index inflates dramatically as sample size decreases, leading to a situation where the null hypothesis of randomness cannot be rejected even for genuinely clustered populations, simply due to insufficient power (Hurlbert, 1990; Dale, 1999). Many empirical studies report only a point estimate of the aggregation index without any measure of its precision, such as a standard error or confidence interval. This practice systematically obscures the degree of uncertainty and can lead to spurious claims of pattern discrimination. The existing literature lacks a systematic framework that ties together sample size limitations, estimation uncertainty, and multiscale pattern detection specifically for the effective discrimination of rare species' spatial patterns.

3 Theoretical Dilemmas and Methodological Reflections on Effective Discrimination of Rare Species' Spatial Patterns

3.1 Defining Effective Discrimination

Effective discrimination is defined here as the unified achievement of statistical significance and ecological realism. It entails three distinct but interrelated dimensions of differentiation. The first is pattern type discrimination: whether the observed point configuration is best described as aggregated, random, or uniform. The second is mechanism discrimination: distinguishing deterministic processes such as habitat filtering from

stochastic processes such as ecological drift and limited dispersal. The third is scale effect discrimination: identifying at which spatial scales aggregation is manifest and at which scales it dissolves into randomness or uniformity. An adequate framework must address all three dimensions simultaneously.

3.2 Fundamental Defects in Existing Methodologies

The current methodological landscape suffers from several interconnected deficiencies. The sample size dilemma is the most fundamental (Zhang, 2024). A rare species may be represented by as few as five to twenty individuals within a sampled quadrat system or mapped plot. At such low densities, the empirical distribution of distances or quadrat counts is necessarily sparse, and the second-order summary statistics become highly variable (Dale and Fortin, 2005). Statistical tests for aggregation, such as Monte Carlo envelope tests based on complete spatial randomness, lack power because the simulation envelopes become so wide that only extreme configurations lead to a significant departure.

The systematic neglect of estimation uncertainty exacerbates the problem. The common practice of reporting a single aggregation index value, such as a negative binomial k of 0.5, without its associated variance gives a false sense of precision. Perry (1995) stressed that aggregation indices from count data exhibit substantial sampling variability, yet most ecological textbooks still advocate simple index thresholds for pattern classification. Two populations with identical true aggregation might yield estimated indices of 0.3 and 0.7 simply due to sampling error when sample sizes are small, leading to contradictory interpretations.

The arbitrary choice of indices further complicates synthesis and comparison. The variance to mean ratio, Morisita's index, the negative binomial k, and Taylor's b each embed different mathematical assumptions and scale references (Taylor, 1961; Morisita, 1959). These indices do not map onto each other in a simple monotonic fashion across all parameter ranges. A species might appear strongly aggregated by one index but only weakly so by another, especially when the underlying distribution deviates from the assumed theoretical form (Zhang, 2007a).

The oversimplification of scale represents another critical gap. Many studies either pick a single quadrat size or a single distance r for analysis, ignoring the fact that spatial pattern is a continuously varying function of scale (Levin, 1992). A dichotomous decision at one chosen scale ignores the rich multiscale structure that could reveal mechanistic details, such as a dispersal kernel causing aggregation up to 20 meters superimposed on a broad-scale habitat gradient causing aggregation at hundreds of meters. Finally, standard statistical models often assume asymptotic properties and well-behaved likelihood surfaces that break down for small samples. The negative binomial distribution, for instance, requires joint estimation of the mean and the aggregation parameter, and the maximum likelihood estimator of k is notoriously unstable when the mean is low and the true aggregation is moderate to high (Anscombe, 1949; Bliss and Fisher, 1953). These model violations mean that even if an index is computed, the underlying inferential framework may be invalid.

3.3 In-depth Analysis of the Theoretical Dilemma

At the heart of the difficulty lies a signal-to-noise problem. The true ecological signal of aggregation in a rare species is a product of the intensity and scale of the underlying point process. The noise arises from finite sampling variability and demographic stochasticity. When abundance is low, the signal is weak and easily swamped by noise, creating a zone of ambiguity where the observed pattern is consistent with both a weakly aggregated true process and a purely random process (Plotkin et al., 2000).

I propose the concept of an impossibility triangle linking rarity, sample size, and pattern identifiability. For a fixed low abundance, achieving high identifiability requires a very large sampling extent or extremely fine grain, which may be impractical. Conversely, for a given sampling design, there exists a lower bound of abundance below which no degree of aggregation can be statistically distinguished from randomness with a required confidence. This non-negotiable constraint implies that many rare species are destined to remain in an

indeterminate state under traditional binary classification schemes. The logical leap from pattern description to mechanism inference, therefore, carries a grave risk: a failure to reject the null hypothesis of randomness is frequently interpreted as evidence for neutral or stochastic processes, when in fact the test simply lacked the power to detect true niche-based aggregation (Hubbell, 2001). This misinterpretation has profound consequences for our understanding of community assembly.

4 Innovative Theoretical Framework for Effective Discrimination of Rare Species' Spatial Patterns

4.1 Core Philosophy: From Binary Choice to a Multidimensional Continuum

Initially inspired by our earlier studies on continuous spatial distribution models (Zhang, 1993; Zhang and Shang, 1992), I advocate a fundamental shift away from the binary classification of aggregated versus random. The new paradigm conceives the degree of aggregation as a continuous variable, forming a spatial pattern continuum (Zhang, 1993; Zhang and Shang, 1992). A species is not simply aggregated; it possesses a specific level of aggregation with an associated uncertainty. I introduce the concept of pattern identifiability as a quantitative measure of how reliably the true aggregation can be determined given the available data. Identifiability is high when the confidence interval of an aggregation index is narrow and entirely on one side of the randomness threshold; it is low when the interval straddles the threshold, indicating that the true state could be on either side (Zhang, 2007a). This reframing transforms the objective from classification to quantification with explicit uncertainty.

4.2 Three Pillars of the Framework

Pillar 1: Pattern Heterogeneity under Multidimensional Rarity

The seven forms of rarity proposed by Rabinowitz (1981) imply distinct spatial pattern signatures. A species that is rare because of restricted geographic range but locally abundant may form a single large dense cluster. A species that is a sparse habitat specialist will be strongly associated with rare habitat patches, exhibiting strong aggregation at the patch scale. A species that is globally sparse everywhere may show weak, diffuse aggregation. The framework explicitly hypothesizes that the spatial pattern continuum is stratified by rarity type, and that the internal heterogeneity among rare species is far greater than the average difference between rare and common species. Hence, any analysis must first classify rarity type and then analyze pattern conditional on that type.

Pillar 2: Scale Continuity and Pattern Polymorphism

Spatial pattern is a function of spatial scale. The framework adopts a continuous-scale perspective, using functions like the pair correlation function $g(r)$ or the O-ring statistic $O(r)$ to characterize pattern across all measurable distances (Wiegand and Moloney, 2004). I introduce the concept of a pattern shift threshold, defined as the spatial scale at which the expected pattern shifts from statistically significant aggregation to randomness or uniformity. A two-dimensional scale-pattern phase diagram is proposed as a visualization tool, where the x-axis is the spatial scale r, and the y-axis is the aggregation index estimate with its confidence band. The phase diagram displays zones of credible aggregation, credible randomness, and uncertainty, delineated by the pattern shift thresholds.

Pillar 3: Quantification and Integration of Estimation Uncertainty

The framework mandates that no pattern inference be made without an accompanying uncertainty measure. I define a statistically credible pattern as a state where the estimated 95% confidence interval (or more confidential, >99%; Zhang, 2022) for the chosen aggregation index theta is entirely located on one side of a predefined discrimination threshold. For the most common scenario, where complete spatial randomness serves as the null model, the threshold for an index normalized to unity under randomness is 1.0. If the entire confidence interval lies above 1.0, the pattern is credibly aggregated. If it lies below 1.0, it is credibly uniform.

If the interval includes 1.0, the pattern is in an ambiguous or indeterminate state; the data do not permit reliable discrimination. An uncertainty-guided decision tree then classifies each species at each scale into one of these three credibility categories.

4.3 Mathematical Formalization and Operationalization

Let θ denote a generic, scale-dependent aggregation index whose expected value under complete spatial randomness is θ_0. Without loss of generality, we focus on the variance to mean ratio $VMR = \sigma^2/\mu$, for which $\theta_0 = 1$. Let $\hat{\theta}$ be the sample estimate and $SE(\hat{\theta})$ its standard error. The approximate 95% confidence interval is $CI = \hat{\theta} \pm 1.96\ SE(\hat{\theta})$. The decision rule is as follows. If the lower bound of the *CI* is greater than 1, the species is classified as credibly aggregated at that scale. If the upper bound is less than 1, it is credibly uniform. Otherwise, the pattern is indeterminate.

For the critical case of rare species, the width of the confidence interval is a decreasing function of sample size n and an increasing function of the true degree of aggregation. I propose the concept of minimum distinguishable sample size, denoted $n_{\min}$, which is the smallest number of independent sampling units for which a true aggregation level θ_{true} greater than 1 can be statistically distinguished from 1 with a prescribed power $(1 - \beta)$ and significance level α. Assuming an approximate normal distribution for $\hat{\theta}$, this condition is met when

$$n_{\min} = (z_{\alpha/2} + z_{\beta})^2\, CV^2 / (\theta_{\text{true}} - 1)^2,$$

where *CV* is the coefficient of variation of $\hat{\theta}$, which itself may depend on θ_{true} and μ. For the variance to mean ratio under a negative binomial model with mean mu and aggregation parameter k, I derive a detailed approximation for *CV* in Appendix. The resulting formula links sampling effort directly to the expected pattern identifiability. For example, a rare species with a mean of 2 individuals per quadrat and true $VMR = 2.0$ (moderate aggregation) may require over 40 quadrats just to achieve an 80% chance of credible identification, while a more common species with a mean of 20 may require fewer than 10 quadrats.

4.4 Comparison with and Transcendence of Existing Theories

The proposed framework reconciles and extends classical pattern ecology. Traditional point pattern analysis, as summarized by Diggle (2003) and Dale and Fortin (2005), focuses on significance testing against a null model of complete spatial randomness. By adding a rigorous uncertainty layer, our framework transforms the binary outcome of significance tests into a continuous measure of evidential strength. It integrates with the aggregation-abundance relationship theory by explicitly modeling how the variance of aggregation estimates inflates at low abundances, thus explaining the high observed scatter in the lower end of the relationship (Taylor, 1961). Furthermore, the framework is compatible with both neutral and niche theory. A neutral model, such as that of Hubbell (2001), serves as a specific stochastic null model that generates a known continuum of aggregation purely from dispersal limitation and ecological drift. Our framework can evaluate whether an observed pattern falls within the credible region of a neutral prediction or whether deterministic habitat filtering must be invoked to explain a deviation.

5 An Analytical Methodology System for Effective Discrimination

5.1 General Design Principles

The methodology is governed by four principles. First, data-driven sample size assessment takes priority: before any analysis, the raw count data must be evaluated to compute the expected precision of aggregation estimates, and the feasibility of discrimination at various scales must be determined. Second, multi-index cross-validation is mandatory: at least three aggregation indices from different mathematical families, such as

the *VMR*, Morisita's I_Δ, and the integral of the pair correlation function, should be computed to check for agreement. Third, analysis must be scale explicit: all computations are performed across a continuous range of spatial lags or grain sizes. Fourth, uncertainty must be quantified and transparently reported: every estimate is accompanied by its bootstrap confidence interval or Bayesian credible interval (Zhang, 2022).

5.2 Core Analytical Methods

5.2.1 Improved Aggregation Estimation

To overcome the instability of plug-in point estimators, I adopt a bootstrap resampling approach. For a set of n sampling units (quadrats, plots, or distance bins), we draw B resamples with replacement and compute the target aggregation index for each resample. The percentile method yields a 95% confidence interval for θ (Efron, 1979). This nonparametric approach avoids strong distributional assumptions and adapts naturally to the sample size.

A Bayesian alternative is recommended when prior ecological knowledge is available. I model the counts using a negative binomial likelihood with an uninformative prior on the aggregation parameter k or a transformed parameter $\varphi = 1/k$. Markov Chain Monte Carlo sampling yields the full posterior distribution of the aggregation index, and the 95% highest posterior density interval is used as the credible interval (Clark, 2005). Bayesian methods are particularly powerful for small samples because they naturally regularize extreme estimates by shrinking them toward prior means.

A spatially explicit neutral model, such as a simulation of the unified neutral theory on a lattice landscape (Hubbell, 2001; Rosindell et al., 2011), provides a customized null model that accounts for the observed species abundance distribution and dispersal limitation. The aggregation index of the observed rare species is compared against the distribution of indices obtained from many neutral model realizations, thereby testing whether the observed pattern departs from the neutral expectation.

5.2.2 Multiscale Pattern Recognition

I construct a multiscale pattern curve by computing the chosen aggregation index or summary function across a dense sequence of spatial scales. For point pattern data, the pair correlation function $g(r)$ is evaluated at distances from the minimum resolvable distance to one quarter of the plot side, with a step of 1 meter. For quadrat count data, grain size is varied by aggregating adjacent quadrats. The pattern shift threshold is identified as the scale at which the confidence band of the aggregation index first crosses the randomness threshold. A multiscale pattern fingerprint is then defined as the vector of credible pattern states (aggregated, indeterminate, uniform) across the scale sequence. This fingerprint facilitates the comparison of different species or different populations.

5.2.3 Uncertainty Quantification in Pattern Discrimination

I assign a probability to each pattern classification. For a given aggregation index and scale, the probability of credible aggregation is defined as the proportion of bootstrap samples for which the index exceeds the randomness threshold. If this probability is above 0.975, the pattern is credibly aggregated; if below 0.025, it is credibly uniform; otherwise, indeterminate. This probabilistic approach gives a continuous measure of confidence. Sensitivity analysis is performed by varying the block size in the bootstrap or by using alternative indices, and the stability of the classification is recorded.

5.3 Decision Workflow of the Methodology

The five-step workflow is as follows.

Step 1: Sample size assessment and feasibility judgment. Compute the expected coefficient of variation of the aggregation index based on the observed mean and a preliminary estimate of aggregation. If the required sample size $n_{\min}$ exceeds the available n, flag the analysis as of limited discriminatory power.

Step 2: Multi-index computation and confidence interval estimation. Compute *VMR*, Morisita's index, and the

pair correlation function integral with their 95% bootstrap CIs.

Step 3: Multiscale pattern curve plotting. Generate the aggregation profile across scales with confidence bands.

Step 4: Uncertainty-threshold-based pattern classification. Apply the credible pattern decision rule at each scale and assign an overall fingerprint.

Step 5: Contextualized mechanism inference. Formulate competing hypotheses for the observed fingerprint based on the rarity type and the ecological setting, and assess which processes are most consistent with the scale-specific patterns.

6 Empirical Research Design: A Validation Pathway for the Framework and Methodology

6.1 Simulation Study Design

To validate the framework, I propose a comprehensive simulation study using spatially explicit neutral models. A metacommunity is simulated on a 500 by 500 grid with periodic boundaries. The fundamental biodiversity number θ_H is set to 50, and the immigration rate m is varied to produce communities of different sizes. After a burn-in period, the local community of 200 by 200 cells is extracted. Individuals are then thinned to create a realistic species abundance distribution, yielding a set of rare species with abundances ranging from 5 to 50 individuals. True aggregation is known exactly because each individual's location is known. For each rare species, apply the proposed methodology to test whether the credible pattern classification correctly identifies the true state. Systematically vary sample size (the number of quadrats used for analysis) to evaluate the statistical power and Type I error rates. The performance of the framework is compared against traditional methods that use a simple point estimate of Morisita's index with a chi-squared test.

6.2 Empirical Case Study Design

The framework will be applied to data from the Barro Colorado Island 50 ha forest dynamics plot, for which all freestanding trees with a diameter at breast height of 1 cm or greater have been mapped (Condit et al., 2000). Rare species are defined as those with fewer than 50 individuals in the entire plot. For each species, the pair correlation function $g(r)$ and the O-ring statistic $O(r)$ are computed up to r = 100 meters. Bootstrap confidence intervals are constructed using a block-of-quadrats resampling scheme to preserve spatial autocorrelation. The multiscale fingerprint and the probability of credible aggregation are calculated. The results are then compared with existing knowledge on habitat associations derived from independent environmental data for the same species (Harms et al., 2001).

6.3 Expected Results and Theoretical Verification

We can expect that the proposed framework will yield a higher proportion of species classified as indeterminate at small sample sizes compared to traditional methods that force a binary decision, thus reducing false positives. As sample size increases, the proportion of species with a credible determination should rise. The multiscale fingerprints are predicted to reveal previously hidden complexities, such as species that are credibly aggregated only at very short distances but indeterminate at larger distances, a pattern consistent with tight dispersal limitation but broad-scale environmental randomness. These outcomes will provide strong empirical support for the theoretical claims made in Section 4.

7 Discussion

7.1 Ecological Significance of the Theoretical Framework

The framework carries substantial implications for community assembly research. The reliability of inferences about deterministic versus stochastic processes hinges on the accurate identification of spatial patterns in rare species. A study that erroneously classifies a truly aggregated rare species as random due to low power might incorrectly conclude that stochastic processes dominate community assembly, when in fact niche-based habitat

filtering is at work (Chase and Myers, 2011). By making the uncertainty explicit, our framework prevents such overconfident inferences and directs attention to the data-deficient species for which conclusions must be suspended. For species coexistence theory, the protective effect of intraspecific aggregation on rare species persistence via the Janzen-Connell mechanism critically depends on the strength and scale of aggregation. The multiscale fingerprints here can test whether the observed aggregation is of sufficient strength and spatial extent to stabilize coexistence relative to heterospecific competitors.

For macroecology, the accurate estimation of the aggregation-abundance relationship is fundamental to models of species richness and abundance distributions. The common observation of a negative relationship between aggregation and abundance is confounded by increased measurement error at low abundances, potentially inflating the relationship artificially. By applying the minimum distinguishable sample size concept, researchers can filter out species whose aggregation cannot be reliably estimated, leading to less biased scaling relationships.

7.2 Practical Implications for Conservation Biology

The framework offers direct practical benefits. For rare species survey and monitoring design, the minimum distinguishable sample size formula provides an evidence-based target for the number of quadrats or transect segments required to achieve a desired level of confidence in pattern identification. Conservation practitioners can use this information to allocate limited resources efficiently, ensuring that monitoring efforts are sufficient to answer the central question of whether a rare species is clustered in a protectable habitat patch or is diffusely distributed. Communicating distribution predictions with quantified uncertainty, rather than as binary presence-absence maps, will improve the reliability of conservation prioritization. For instance, a map of probability of credible aggregation could guide the placement of protected area boundaries around high-confidence clusters while acknowledging areas of ambiguous status. Lan et al. (2025) point out that 70% of China's protected areas exhibit a "conservation mismatch", which manifests in two ways: a spatial mismatch: 70% of these protected areas are located in the sparsely populated western regions, where extinction risks are relatively low; and a conservation gap: over 90% of the country's plant species are distributed in the eastern regions, where extinction risks are high but protection is severely insufficient, with small and highly fragmented reserves and nearly 20% of core areas already lost. This study reveals the "hidden extinction crisis" confronting China's plant diversity.

7.3 Limitations of the Framework and Future Research Directions

The framework rests on the assumption that the sampling scheme is adequate for revealing the underlying point pattern, which may be violated if the species is highly cryptic or detection probability is low and heterogeneous. The integration of detection probability models, such as N-mixture models or occupancy models with spatial autocorrelation, represents a vital future extension (MacKenzie and Royle, 2005). The current framework is primarily developed for static spatial patterns; extending it to spatiotemporal dynamics of rare species will allow the discrimination of transient aggregations from persistent ones. The integration with modern machine learning tools and joint species distribution models (Pollock et al., 2014) that borrow strength across species could further improve pattern identifiability for the rarest species. Finally, the framework should be extended to functional and phylogenetic dimensions of rarity, testing whether functionally rare species exhibit characteristic spatial fingerprints.

8 Conclusion

This paper has presented a comprehensive reexamination of how ecologists discriminate the spatial distribution patterns of rare species. The central argument is that the traditional binary paradigm of aggregated

versus random is fundamentally inadequate in the face of the profound estimation uncertainty caused by low abundance. I proposed a new theoretical framework that redefines aggregation as a spatial pattern continuum, integrates multidimensional rarity and scale continuity, and places estimation uncertainty at the core of inference. The concepts of statistically credible pattern, pattern identifiability, and minimum distinguishable sample size provide an operational vocabulary for rigorous and honest inference. The analytical methodology, employing bootstrap and Bayesian uncertainty estimation with multiscale fingerprinting, offers a practical pathway for implementation. Through this work, I call on the ecological community to adopt a new standard that mandates the reporting of uncertainty measures in all studies of spatial pattern for rare species, thereby enabling more robust assessments of biodiversity and more effective conservation action.

Appendix

Derivation of Minimum Distinguishable Sample Size for the Variance to Mean Ratio

This appendix provides the detailed derivation of the minimum sample size required to distinguish an aggregated pattern ($VMR > 1$) from randomness, using the variance to mean ratio as the aggregation index.

A.1 Theoretical Setup

Consider a negative binomial distribution with mean mu and aggregation parameter *k*. The probability mass function is:

$$P(X = x) = \Gamma(x + k) / (\Gamma(k)\, x!)\, (\mu / (\mu + k))^x\, (k / (\mu + k))^k$$

The variance is $\sigma^2 = \mu + \mu^2/k$, so the true *VMR* is $\theta = \sigma^2/\mu = 1 + \mu/k$. Under complete spatial randomness (*k* to infinity), $\theta = 1$.

Given a sample of *n* independent quadrat counts, compute the sample mean $\bar{m} = \Sigma X_i / n$ and the sample variance $s^2 = \Sigma (X_i - \bar{m})^2 / (n - 1)$. The plug-in estimator is $\hat{\theta} = s^2/\bar{m}$.

A.2 Asymptotic Variance of the Estimator

Employ the delta method to approximate the variance of $\hat{\theta}$. Let $\mu_2 = \sigma^2$, μ_3 the third central moment, and μ_4 the fourth central moment of the negative binomial. The necessary moments are:

$$\mu_2 = \mu + \mu^2/k,$$
$$\mu_3 = \mu + 3\mu^2/k + 2\mu^3/k^2,$$
$$\mu_4 = \mu + 7\mu^2/k + 6\mu^3/k^2 + 3\mu^4/k^3 + 3\mu^2 + 6\mu^3/k + 3\mu^4/k^2.$$

The covariance matrix of $(\bar{m}, s^2)$ for i.i.d. data can be derived from standard asymptotic theory. The asymptotic variance of $\hat{\theta}$ is:

$$\mathrm{Var}(\hat{\theta}) \approx (1/n) * [\, (\mu_4 - \sigma^4)/\mu^2 - 4\sigma^2\mu_3/\mu^3 + 4\sigma^4/\mu^2].$$

This expression is exact in large samples. For small *n*, a correction factor of $(n/(n-1))$ may be applied, and the jackknife or bootstrap is recommended for practical use (Zhang, 2007a).

A.3 Coefficient of Variation and Sample Size Formula

The coefficient of variation of hat_theta is $CV = \sqrt{\mathrm{Var}(\hat{\theta})}/\theta$. Substituting $\theta = 1 + \mu/k$ and the variance expression yields a function $CV(\mu, k, n)$. To achieve a significant distinction of θ from 1 with a two-sided test at level α and power 1 - θ, we require:

$$n = (z_{\alpha/2} + z_{\beta})^2 \, CV_0^2 / (\theta - 1)^2,$$

where $CV_0 = CV$ evaluated at a pilot estimate of μ and k. The minimum distinguishable sample size $n_{\min}$ is the smallest integer n satisfying this equation. For a typical scenario with $\mu = 2$, $k = 2$ ($\theta = 2$), and $\alpha = 0.05$, $\beta = 0.2$, the formula yields $n_{\min} = 43$ quadrats.

A.4 Computational Algorithm for the Bootstrap Method

To avoid reliance on asymptotic approximations for small samples, I implement the following bootstrap algorithm:

1. From the original n quadrat counts, draw B bootstrap samples of size n with replacement.
2. For each bootstrap sample b, compute $\hat{\theta}_b$.
3. The 95% percentile confidence interval is the 2.5th and 97.5th percentiles of the bootstrap distribution of $\hat{\theta}$.
4. If the lower percentile exceeds 1.0, the pattern is credibly aggregated. If the upper percentile is below 1.0, it is credibly uniform.
5. The probability of credible aggregation is the proportion of $\hat{\theta}_b > 1$.
6. The minimum sample size is determined by simulation: for a given true distribution, repeatedly generate data of increasing n, apply the bootstrap decision rule, and record the proportion of trials achieving credible aggregation. The smallest n for which this proportion exceeds 0.8 is the empirical $n_{\min}$.

Below is a complete JavaScript implementation of the bootstrap algorithm for credible aggregation discrimination, along with the minimum sample size search via simulation. The code includes robust generators for negative binomial random variates (using a Gamma–Poisson mixture) and a gamma random variate generator based on well-established numerical methods. All functions are self-contained and can be run in Node.js or any modern browser console.

javascript

```
/**
 * Effective Discrimination of Spatial Patterns for Rare Species
 * Bootstrap uncertainty quantification for the Variance-to-Mean Ratio (VMR)
 * and determination of minimum distinguishable sample size.
 *
 * All algorithms are fully reproducible and documented.
 */

// ------------------------------------------------------------
// 1. Random number generators for Gamma and Negative Binomial
// ------------------------------------------------------------
```

```
/**
 * Generate a Gamma(shape=alpha, scale=theta) random variate.
 * Uses:
 *    - For alpha > 1: Marsaglia & Tsang (2000) method
 *    - For alpha < 1: transformation of a uniform variate via the property
 *        Gamma(alpha, theta) ~ Gamma(alpha+1, theta) * U^(1/alpha)
 * Edge case alpha = 1 treated as exponential.
 */
function randomGamma(alpha, theta) {
   if (alpha <= 0) throw new Error('alpha must be > 0');
   if (alpha === 1) {
      // Exponential distribution
      return -Math.log(Math.random()) * theta;
   }
   if (alpha < 1) {
      // Boost alpha by 1 and rescale
      const g = randomGamma(alpha + 1, theta);
      const u = Math.random();
      return g * Math.pow(u, 1 / alpha);
   }
   // alpha > 1: Marsaglia and Tsang algorithm
   const d = alpha - 1 / 3;
   const c = 1 / Math.sqrt(9 * d);
   while (true) {
      let x, v;
      do {
         x = standardNormal();
         v = 1 + c * x;
      } while (v <= 0);
      v = v * v * v;
      const u = Math.random();
      if (u < 1 - 0.331 * x * x * x * x) {
         return (d * v) * theta;
      }
      if (Math.log(u) < 0.5 * x * x + d * (1 - v + Math.log(v))) {
         return (d * v) * theta;
      }
   }
}

/**
 * Standard normal variate (Box–Muller)
 */
```

```
function standardNormal() {
   let u1, u2;
   do {
      u1 = Math.random();
   } while (u1 === 0);
   u2 = Math.random();
   return Math.sqrt(-2 * Math.log(u1)) * Math.cos(2 * Math.PI * u2);
}

/**
 * Generate a negative binomial count with mean mu and aggregation parameter k.
 * NB(r=k, p = k/(mu+k)). Done via Poisson–Gamma mixture:
 * lambda ~ Gamma(k, mu/k), then X ~ Poisson(lambda).
 */
function randomNegativeBinomial(mu, k) {
   if (mu <= 0 || k <= 0) throw new Error('mu and k must be positive');
   const lambda = randomGamma(k, mu / k);
   return randomPoisson(lambda);
}

/**
 * Generate a Poisson random variate with mean lambda.
 */
function randomPoisson(lambda) {
   if (lambda <= 0) return 0;
   // For small lambda, use exponential interarrivals
   if (lambda < 30) {
      let L = Math.exp(-lambda);
      let k = 0;
      let p = 1;
      do {
         k++;
         p *= Math.random();
      } while (p > L);
      return k - 1;
   }
   // Normal approximation for larger lambda
   return Math.max(0, Math.round(
      lambda + Math.sqrt(lambda) * standardNormal()
   ));
}

// ------------------------------------------------------------
// 2. Statistical estimators and bootstrap
```

```
// ------------------------------------------------------------

/**
 * Compute sample variance-to-mean ratio (VMR) from an array of counts.
 */
function computeVMR(sample) {
  const n = sample.length;
  if (n < 2) return NaN;
  const mean = sample.reduce((a, b) => a + b, 0) / n;
  if (mean === 0) return 0; // all zeros: VMR defined as 0
  const variance = sample.reduce((s, x) => s + (x - mean) ** 2, 0) / (n - 1);
  return variance / mean;
}

/**
 * Bootstrap confidence interval and classification for a given sample.
 * Returns an object with:
 *   vmr: point estimate
 *   lowerCI, upperCI: 95% percentile bootstrap confidence limits
 *   credibleAggregation: probability that VMR > 1 across bootstrap replicates
 *   classification: 'aggregated', 'uniform', or 'indeterminate'
 */
function bootstrapVMR(sample, B = 5000) {
  const n = sample.length;
  const obsVMR = computeVMR(sample);
  const bootVMRs = new Array(B);
  for (let i = 0; i < B; i++) {
    // draw resample of size n with replacement
    const resample = new Array(n);
    for (let j = 0; j < n; j++) {
      resample[j] = sample[Math.floor(Math.random() * n)];
    }
    bootVMRs[i] = computeVMR(resample);
  }
  bootVMRs.sort((a, b) => a - b);
  const lower = bootVMRs[Math.floor(0.025 * B)];
  const upper = bootVMRs[Math.floor(0.975 * B)];
  const probAgg = bootVMRs.filter(v => v > 1).length / B;
  let classification = 'indeterminate';
  if (lower > 1) {
    classification = 'credibly aggregated';
  } else if (upper < 1) {
    classification = 'credibly uniform';
  }
```

```
    return {
        pointEstimate: obsVMR,
        lowerCI: lower,
        upperCI: upper,
        probCredibleAggregation: probAgg,
        classification: classification,
    };
}

// ------------------------------------------------------------
// 3. Minimum distinguishable sample size (n_min) search
// ------------------------------------------------------------

/**
 * Determine the minimum sample size required to reliably detect a given true VMR > 1
 * using bootstrap classification. A trial is successful if the classification is
 * 'credibly aggregated' at significance level alpha=0.05.
 *
 * @param {number} mu - true mean abundance per sampling unit
 * @param {number} k - true negative binomial aggregation parameter (smaller = more aggregated)
 * @param {number} power - desired statistical power (default 0.8)
 * @param {number} maxN - maximum sample size to consider
 * @param {number} nTrials - number of simulation replicates per sample size
 * @param {number} B - bootstrap replicates per trial
 * @returns {number|null} - n_min, or null if not found within maxN
 */
function findMinSampleSize(mu, k, power = 0.8, maxN = 200, nTrials = 200, B = 5000) {
    const trueVMR = 1 + mu / k;
    if (trueVMR <= 1) throw new Error('True VMR must be > 1 for aggregation detection');
    for (let n = 5; n <= maxN; n++) {
        let successes = 0;
        for (let trial = 0; trial < nTrials; trial++) {
            // Generate a sample of size n from the true distribution
            const sample = Array.from({ length: n }, () => randomNegativeBinomial(mu, k));
            const result = bootstrapVMR(sample, B);
            if (result.classification === 'credibly aggregated') {
                successes++;
            }
        }
        const observedPower = successes / nTrials;
        if (observedPower >= power) {
            return n;
        }
```

```
    }
    return null; // not found within maxN
}

// ------------------------------------------------------------
// 4. Example usage
// ------------------------------------------------------------

// Example with observed quadrat counts from a hypothetical rare species
const observedCounts = [0, 2, 0, 1, 5, 0, 0, 3, 1, 0, 4, 2];
console.log('Observed counts:', observedCounts);
const result = bootstrapVMR(observedCounts);
console.log(`Point estimate VMR: ${result.pointEstimate.toFixed(3)}`);
console.log(`95% bootstrap CI: [${result.lowerCI.toFixed(3)}, ${result.upperCI.toFixed(3)}]`);
console.log(`Probability of credible aggregation: ${result.probCredibleAggregation.toFixed(3)}`);
console.log(`Classification: ${result.classification}`);

// Determine minimum sample size for a species with mu=2, k=2 (true VMR=2)
const nMin = findMinSampleSize(2, 2, 0.8, 100, 100, 5000);
console.log(`\nFor mu=2, k=2, minimum sample size for 80% power: ${nMin}`);
```

Explanation of the algorithm

1. Data preparation: The input is an array of quadrat counts (e.g., number of individuals per plot).
2. *VMR* computation: The function `computeVMR()` calculates the sample variance-to-mean ratio *CV*, which equals 1 under complete spatial randomness and >1 for aggregation.
3. Bootstrap resampling: The function `bootstrapVMR()` draws *B* resamples of size *n* with replacement from the original counts, recomputes the *VMR* for each resample, and builds a 95% percentile confidence interval.
4. Classification rule: If the entire confidence interval lies above 1, the pattern is *credibly aggregated*. If it lies entirely below 1, it is *credibly uniform*. Otherwise, the discrimination is *indeterminate*. The proportion of bootstrap replicates with *VMR* > 1 gives the probability of *credible aggregation*, a continuous measure of confidence.
5. Minimum sample size search: The function `findMinSampleSize()` simulates data from a known negative binomial distribution (with mean μ and aggregation parameter *k*) for increasing sample sizes. At each candidate *n*, it repeatedly applies the bootstrap classification and records the fraction of trials where the outcome is *credibly aggregated*. The smallest *n* for which this power exceeds the desired threshold (e.g., 0.8) is the *minimum distinguishable sample size*. This quantifies the sampling effort required to reliably separate aggregation from randomness.

The code is self-contained and uses only standard JavaScript `Math` functions, making it readily reproducible in any environment that supports ES6. For clarity, the gamma variate generator employs established algorithms (Marsaglia and Tsang (2000) for shape > 1, and a simple transformation for shape < 1).

References

Anscombe FJ. 1949. The statistical analysis of insect counts based on the negative binomial distribution. Biometrics, 5(2): 165-173. https://www.jstor.org/stable/3001918

Baddeley A, Rubak E, Turner R. 2015. Spatial Point Patterns: Methodology and Applications with R. Chapman and Hall/CRC, New York, USA. https://www.taylorfrancis.com/books/mono/10.1201/b19708/spatial-point-patterns-adrian-baddeley-rolf-turner-ege-rubak

Bliss CI, Fisher RA. 1953. Fitting the negative binomial distribution to biological data. Biometrics, 9(2): 176-200. https://www.jstor.org/stable/3001850?origin=crossref&seq=1

Breiner FT, Guisan A, Bergamini A, Nobis MP. 2015. Overcoming limitations of modelling rare species by using ensembles of small models. Methods in Ecology and Evolution, 6(10): 1210-1218. https://doi.org/10.1111/2041-210X.12403

Chase JM, Myers JA. 2011. Disentangling the importance of ecological niches from stochastic processes across scales. Philosophical Transactions of the Royal Society B, 366(1576): 2351-2363. https://royalsocietypublishing.org/rstb/article-abstract/366/1576/2351/21612/Disentangling-the-importance-of-ecological-niches?redirectedFrom=fulltext

Clark JS. 2005. Why environmental scientists are becoming Bayesians. Ecology Letters, 8(1): 2-14. https://onlinelibrary.wiley.com/doi/10.1111/j.1461-0248.2004.00702.x

Condit R, Ashton PS, Baker P, et al. 2000. Spatial patterns in the distribution of tropical tree species. Science, 288(5470): 1414-1418. https://www.science.org/doi/10.1126/science.288.5470.1414

Dale MRT. 1999. Spatial Pattern Analysis in Plant Ecology. Cambridge University Press, Cambridge, USA. https://www.cambridge.org/core/books/spatial-pattern-analysis-in-plant-ecology/A5B47D9CCFA2E1CE1267B9E66E7F9208

Dale MRT, Fortin MJ. 2005. Spatial Analysis: A Guide for Ecologists, 2nd edn. Cambridge University Press, Cambridge, USA. https://assets.cambridge.org/97805211/43509/frontmatter/9780521143509_frontmatter.pdf

Diggle PJ. 2003. Statistical Analysis of Spatial Point Patterns, 2nd edn. Edward Arnold, London, UK. https://research.lancaster-university.uk/en/publications/statistical-analysis-of-spatial-point-patterns/

Efron B. 1979. Bootstrap methods: Another look at the jackknife. The Annals of Statistics, 7(1): 1-26. https://projecteuclid.org/journals/annals-of-statistics/volume-7/issue-1/Bootstrap-Methods-Another-Look-at-the-Jackknife/10.1214/aos/1176344552.full

Gaston KJ. 1994. Rarity. Springer, Dordrecht, Netherlands. https://link.springer.com/book/10.1007/978-94-011-0701-3

Guisan A, Broennimann O, Engler R, Vust M, et al. 2006. Using niche-based models to improve the sampling of rare species. Conservation Biology, 20(2): 501-511. https://conbio.onlinelibrary.wiley.com/doi/epdf/10.1111/j.1523-1739.2006.00354.x

Harms KE, Condit R, Hubbell SP, Foster RB. 2001. Habitat associations of trees and shrubs in a 50-ha neotropical forest plot. Journal of Ecology, 89(6): 947-959. https://besjournals.onlinelibrary.wiley.com/doi/10.1111/j.1365-2745.2001.00615.x

Hartley S, Kunin WE. 2003. Scale dependency of rarity, extinction risk, and conservation priority. Conservation Biology, 17(6): 1559-1570.

https://conbio.onlinelibrary.wiley.com/doi/epdf/10.1111/j.1523-1739.2003.00015.x

He FL, Legendre P. 2002. Species diversity patterns derived from species–area models. Ecology, 83(5): 1185-1198. https://esajournals.onlinelibrary.wiley.com/doi/10.1890/0012-9658(2002)083%5B1185:SDPDFS%5D2.0.CO;2

Hubbell SP. 2001. The Unified Neutral Theory of Biodiversity and Biogeography. Princeton University Press, Princeton, USA. https://press.princeton.edu/books/paperback/9780691021287/the-unified-neutral-theory-of-biodiversity-and-biogeography?srsltid=AfmBOoraOYKH3XkmQ-0yGLju28dW9r6rTVDw9fYYVCQbK-ZQvpfdK4Nu

Hurlbert SH. 1990. Spatial distribution of the montane unicorn. Oikos, 58(3): 257-271. https://www.jstor.org/stable/3545216

Jeliazkov A, Gavish Y, Marsh CJ, et al. 2022. Sampling and modelling rare species: Conceptual guidelines for the neglected majority. Global Change Biology, 28: 3754–3777. https://onlinelibrary.wiley.com/doi/10.1111/gcb.16114

Kunin WE. 1998. Extrapolating species abundance across spatial scales. Science, 281(5382): 1513-1515. https://www.science.org/doi/10.1126/science.281.5382.1513

Kunin WE, Gaston KJ. 1993. The biology of rarity: Patterns, causes and consequences. Trends in Ecology & Evolution, 8(8): 298-301. https://www.sciencedirect.com/science/article/abs/pii/016953479390259R?via%3Dihub

Lan TY, Pimm SL, Chase JM, et al. 2025. Revealing hidden extinction risks in China's flora through dynamic habitat assessment. One Earth, 8(12): 101429. https://www.sciencedirect.com/science/article/pii/S2590332225002556?via%3Dihub.

Lande R. 1996. Statistics and partitioning of species diversity, and similarity among multiple communities. Oikos, 76(1): 5-13. https://www.jstor.org/stable/3545743?origin=crossref&seq=1

Levin SA. 1992. The problem of pattern and scale in ecology. Ecology, 73(6): 1943-1967. https://esajournals.onlinelibrary.wiley.com/doi/10.2307/1941447

Levine JM, Murrell DJ. 2003. The community-level consequences of seed dispersal patterns. Annual Review of Ecology, Evolution, and Systematics, 34: 549-574. https://www.annualreviews.org/content/journals/10.1146/annurev.ecolsys.34.011802.132400

Lloyd M. 1967. Mean crowding. Journal of Animal Ecology, 36(1): 1-30. https://www.jstor.org/stable/3012?origin=crossref&seq=1

Liu GH, Zhang WJ. 2011. Computer generation of initial spatial distribution for cell automata. Computational Ecology and Software, 1(4): 244-248. http://www.iaees.org/publications/journals/ces/articles/2011-1(4)/computer-generation-of-initial-spatial-distribution.pdf

Lyons KG, Brigham CA, Traut BH, Schwartz MW. 2005. Rare species and ecosystem functioning. Conservation Biology, 19(4): 1019-1024. https://conbio.onlinelibrary.wiley.com/doi/epdf/10.1111/j.1523-1739.2005.00106.x

MacKenzie DI, Royle JA. 2005. Designing occupancy studies: General advice and allocating survey effort. Journal of Applied Ecology, 42(6): 1105-1114. https://besjournals.onlinelibrary.wiley.com/doi/10.1111/j.1365-2664.2005.01098.x

Magurran AE, Henderson PA. 2003. Explaining the excess of rare species in natural species abundance distributions. Nature, 422: 714-716. https://www.nature.com/articles/nature01547

Marsaglia G, Tsang WW. 2000. A simple method for generating gamma variables. ACM Transactions on

Mathematical Software, 26(3): 363-372. https://dl.acm.org/doi/10.1145/358407.358414

Morisita M. 1959. Measuring of the dispersion of individuals and analysis of the distributional patterns. Memoirs of the Faculty of Science, Kyushu University, Series E (Biology), 2(4): 215-235. https://reference.morisita.or.jp/paper_pdf/55.pdf

Perry JN. 1995. Spatial analysis by distance indices. Journal of Animal Ecology, 64(3): 303-314. https://doi.org/10.2307/5892

Phillips SJ, Anderson RP, Schapire RE. 2006. Maximum entropy modeling of species geographic distributions. Ecological Modelling, 190(3-4): 231-259. https://www.sciencedirect.com/science/article/pii/S030438000500267X?via%3Dihub

Pielou EC. 1977. Mathematical Ecology. Wiley, New York, USA. https://www.amazon.com/Mathematical-Ecology-C-Pielou/dp/0471019933

Plotkin JB, Potts MD, Leslie N, Manokaran N, et al. 2000. Species-area curves, spatial aggregation, and habitat specialization in tropical forests. Journal of Theoretical Biology, 207(1): 81-99. https://www.sciencedirect.com/science/article/pii/S0022519300921581

Pollock LJ, Tingley R, Morris WK, Golding N, O'Hara RB, et al. 2014. Understanding co-occurrence by modelling species simultaneously with a joint species distribution model (JSDM). Methods in Ecology and Evolution, 5(5): 397-406. https://besjournals.onlinelibrary.wiley.com/doi/10.1111/2041-210X.12180

Rabinowitz D. 1981. Seven forms of rarity. In: The Biological Aspects of Rare Plant Conservation (Synge H, ed). 205-217. Wiley, Chichester, UK. https://scispace.com/papers/seven-forms-of-rarity-2atqg7aixm

Ripley BD. 1977. Modelling spatial patterns. Journal of the Royal Statistical Society: Series B, 39(2): 172-192. https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.2517-6161.1977.tb01615.x

Rosindell J, Hubbell SP, Etienne RS. 2011. The unified neutral theory of biodiversity and biogeography at age ten. Trends in Ecology & Evolution, 26(7): 340-348. https://www.sciencedirect.com/science/article/pii/S0169534711000942?via%3Dihub

Seidler TG, Plotkin JB. 2006. Seed dispersal and spatial pattern in tropical trees. PLoS Biology, 4(11): e344. https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.0040344

Shang HS, Zhang WJ, Jing JX. 1991. Spatial distribution analysis of primary inoculum of *Gibberella zeae* (Schw.) in wheat field and its sampling method. Acta Univ. Agric. Boreali-Occidentalis, 19(Suppl): 66-70. https://www.zhangqiaokeyan.com/academic-journal-cn_journal-northwest-university-natural-science-edition_thesis/02012110547682.html

Southwood TRE, Henderson PA. 2000. Ecological Methods, 3rd edn. Blackwell Science, Oxford, UK. https://www.researchgate.net/publication/260051655_Ecological_Methods_3rd_edition

Taylor LR. 1961. Aggregation, variance and the mean. Nature, 189: 732-735. https://www.nature.com/articles/189732a0

Tilman D. 1982. Resource Competition and Community Structure. Princeton University Press, Princeton, USA. https://press.princeton.edu/books/paperback/9780691083025/resource-competition-and-community-structure

Wiegand T, Moloney KA. 2004. Rings, circles, and null-models for point pattern analysis in ecology. Oikos, 104(2): 209-229. https://ui.adsabs.harvard.edu/abs/2004Oikos.104..209W/abstract

Wiegand T, Moloney KA. 2014. Handbook of Spatial Point-Pattern Analysis in Ecology. Chapman and Hall/CRC, New York, USA. https://www.taylorfrancis.com/books/mono/10.1201/b16195/handbook-spatial-point-pattern-analysis-ecology-thorsten-wiegand-kirk-moloney

Zhang WJ. 1992. Spatial distribution pattern for source of inoculum of wheat scab in Guanzhong district. Acta

Phytopathologica Sinica, 22(2): 162. https://www.cnki.com.cn/Article/CJFDTotal-ZWBL199202018.htm

Zhang WJ. 1993. Epidemic Simulation and Fungicide Control Decision System of Wheat Scab. PhD Thesis. Northwestern Agricultural University, Yangling, Shannxi, China

Zhang WJ. 2007a. Methodology on Ecology Research. Sun Yat-sen University Pess, Guangzhou, China. https://baike.baidu.com/item/%E7%94%9F%E6%80%81%E5%AD%A6%E7%A0%94%E7%A9%B6%E6%96%B9%E6%B3%95/7057682

Zhang WJ. 2007b. Pattern classification and recognition of invertebrate functional groups using self-organizing neural networks. Environmental Monitoring and Assessment, 130: 415-422. https://link.springer.com/article/10.1007/s10661-006-9432-1

Zhang WJ. 2016. Selforganizology: The Science of Self-Organization. World Scientific, Singapore. https://www.worldscientific.com/worldscibooks/10.1142/9685#t=aboutBook

Zhang WJ. 2018. Fundamentals of Network Biology. World Scientific Europe, London, UK. https://www.worldscientific.com/worldscibooks/10.1142/q0149#t=aboutBook

Zhang WJ. 2022. p-value based statistical significance tests: Concepts, misuses, critiques, solutions and beyond. Computational Ecology and Software, 12(3): 80-122. http://www.iaees.org/publications/journals/ces/articles/2022-12(3)/p-value-based-statistical-significance-tests.pdf

Zhang WJ. 2024. SampSizeCal: The platform-independent computational tool for sample sizes in the paradigm of new statistics. Network Biology, 14(2): 100-155. http://www.iaees.org/publications/journals/nb/articles/2024-14(2)/5-Zhang-Abstract.asp

Zhang WJ. 2026. probFunCal2: A web-based calculator for calculating probability distribution functions as binomial, Poisson, exponential, gamma, beta, uniform, Weibull, lognormal, Cauchy, geometric, and negative binomial distributions. Selforganizology, 13(3-4): 26-65. http://www.iaees.org/publications/journals/selforganizology/articles/2026-13(3-4)/1-Zhang-Abstract.asp

Zhang WJ, Qi YH. 2025. probDistriCal: The online calculator for probability distributions. Computational Ecology and Software, 15(2): 57-77. http://www.iaees.org/publications/journals/ces/articles/2025-15(2)/3-Zhang-Abstract.asp

Zhang WJ, Shang HS. 1992. Continuous spatial distribution models of biological population and their applicatiuon. Chinese Jounal of Applied Ecology, 3(2): 169-172. http://dianda.cqvip.com/Qikan/Article/Detail?id=925323&from=Qikan_Article_Detail