Simple and robust models of ecological abundance

John Alroy

School of Natural Sciences, Macquarie University, NSW, Australia

Email: john.alroy@mq.edu.au

Abstract

1. Counts of species in ecological samples are of interest when they tell us about community assembly processes. Older process-based models of count distributions are either complex, widely rejected, or not able to predict high unevenness.

2. I leverage a general strategy for deriving simple one-parameter models. A distribution of abundances *x* on a continuous scale is predicted from a transform of a uniform distribution *U*; *U* is solved for to yield one minus a cumulative distribution function (CDF) for *x*; and the result is differenced and rounded to down to yield a probability mass function. The same workflow has long been used to derive the geometric series from the exponential distribution. Three variants are proposed, respectively based on the transforms $\mu/U - \mu = (\mu - U)/U$ where μ is a fitted constant (a scaled odds); $[-\ln(U)/\lambda]^2$ where $-\ln U$ is just an exponential random variate and λ is the constant; and $[-\ln(2/U - 1)/\gamma]^4$ where γ is the constant. They collectively cover the range of functions that lead from some U to a non-negative real number.

3. The distributions are all consistent with simple population dynamical models in which recruitment rates, and sometimes death rates, vary randomly amongst species and are fixed for each species. The number of recruited offspring produced during each interval by each species is Poisson-distributed, and death rates are per-capita. Population counts are equilibrial, allowing co-existence in the absence of competition.

4. Large-scale surveys of corals, fishes, butterflies, and trees are consistent with the distributions, as are local-scale inventories of trees and assorted vertebrate and insect groups. Each inventory is used to predict the counts of another one that is matched based on group representation, biogeography, and richness. Based on examining decisive differences between the resulting likelihoods, the new models routinely outperform eight different rivals.

5. Thanks to their simplicity, grounding in non-competitive equilibrial population dynamics, and predictive power, the new approaches have considerable relevance throughout ecology.

KEYWORDS

half-power distribution, log series, negative binomial distribution, Poisson log normal distribution, quarter-power distribution, scaled odds distribution, Weibull distribution

1 | INTRODUCTION

The rules of community assembly are of fundamental interest to ecologists, and debate over them goes back to the conflict between the Gleasonian and Clementsian schools in the early 20th century (Eliot 2007; Presley et al., 2010). Community assembly is grounded in rates of birth, death, and immigration (Kendall, 1948). Rate variation is responsible for complex patterns at local scales such as vegetational succession and predator-prey cycles. However, the rates also scale up to govern speciation and extinction processes. Thus, they indirectly control or correlate with everything that it is interesting in community ecology and macroecology, including biogeographic patterns, species-area relationships, diversity gradients, and trait distributions.

There may be no agreement about which assembly processes are the most important, but the business of ecology is the same as the business of science in general: establishing process by studying pattern. The problem is that there are highly distinct strategies for drawing inferences. For example, presence-absence matrices that compare assemblages may signal several processes (Leibold & Mikkelson, 2002; Henriques-Silva et al., 2013), and species diversity patterns can likewise suggest different population processes such as colonisation and local extirpation (MacArthur & Wilson, 1963; Loreau & Mouquet, 1999).

While that literature is important and interesting, the common currency of community ecology is more basic: simple inventories of species found in particular locations at particular times. The problem is that isolated inventories are generally thought not to contain enough information to indicate assembly processes with any real specificity (Lawton, 1999; McGill et al., 2007; Matthews & Whittaker, 2014). This explains why authors have discussed alternative approaches such as seeing how abundance distributions, which are counts of individuals grouped into species, vary across temporal scales (Magurran, 2007) or spatial scales (Borda-de-Água et al., 2011; Antão et al., 2021).

In this paper, I suggest that individual real-world distributions do have the power to differentiate quite different assembly processes. In particular, I present three new and extremely simple models of population dynamics that all generate simple species abundance distributions. I show that their predicted patterns are common in tree and animal data. Importantly, the new distributions are not only plausible but distinct, so it is possible to reject their underlying models and thereby exclude their assumptions.

Population models have been used in this way before. For example, Kendall (1948) predicted the log series distribution of Fisher et al. (1943) with a completely random, per-capita birth-death process; MacArthur (1960) pointed out that the log normal should result if all populations grow exponentially; and Saether et al. (2013) showed how weak density dependence could also generate the log normal distribution. Meanwhile, the influential zero-sum multinomial (ZSM) distribution of Hubbell (1997, 2001) encompasses the log series and other shapes. It can be derived from a population model that makes clear assumptions about dispersal, speciation, competition, and so on.

These are all long-established ideas. But except for the ZSM, newer species abundance models such as those of Tokeshi (1990) have often not gained much traction. A potential exception is the gambin model of Ugland et al. (2007), which has attracted other attention (Matthews et al., 2014, 2019). This model is difficult to assess for reasons outlined later. Comparative analyses (e.g., Baldridge et al., 2016) have therefore focused on classic alternatives such as the log series (Fisher et al., 1943) and Poisson log normal (Bulmer, 1974).

With all of this previous work, it would be natural to think that nothing more needs to be said. Don't we already have far too many models? I will argue this is not true. But even if the general theory proposed here proves superfluous, stimulating a wider discussion may better our understanding of ecological processes. In addition, the particular new models all have built-in species richness estimators that provide maximum likelihood values when the model assumptions are met. So if the theory is any good, then these estimators might see widespread application.

2 | MATERIALS AND METHODS

2.1 | Workflow for deriving distributions

Throughout this paper, I draw a distinction between two mathematical means of summarising count data: (1) rank-abundance distributions (RADs), which are simply lists of counts ordered from greatest to least; and (2) species-abundance distributions (SADs) sensu stricto, which are lists of counts of species sharing counts (Fisher et al., 1943). Although some researchers prefer to fit data to models by examining RADs (e.g., Hughes, 1986; Ulrich et al., 2018), I emphasise fitting data to SADs by likelihood, as done by Prado et al. (2018), for reasons explained further in the discussion of the preferred fitting method. I use RADs for illustrative purposes because it is easier to grasp them quickly.

A fitted SAD is just a probability mass function (PMF) in the standard statistical sense, which is another good reason to work with SADs. As statisticians well understand, integer-value PMFs can be derived from continuous-value cumulative distribution functions (CDF). A general strategy is to start with a transform of a uniform random variate *U* into a non-uniform random variate *X*:

$$X = f(U) \tag{1}$$

Next, *U* is solved for in terms of *x* to yield $U = f(X)$. The resulting expression is just one minus a CDF if (1) it declines monotonically to zero as *X* approaches infinity, and (2) it either starts with a value of 1 when $X = 0$ or can be scaled easily to do so. In other words, many expressions like $1 - f(X)$ can be CDFs:

$$F_X(x) = P(X \leq x) = 1 - U = 1 - f(X) \tag{2}$$

Finally, a PMF is produced by rounding down the first differences of the CDF:

$$p_X(x) = P(X = x) = [1 - f(x + 1)] - [1 - f(x)] = f(x) - f(x + 1) \tag{3}$$

where *x* is an integer value. The derivation of the geometric series from the exponential distribution is then as follows:

$$X = -\ln U \tag{4}$$

$$F_X(x) = 1 - U = 1 - \exp(-X) \tag{5}$$

$$p_X(x) = \exp(-x) - \exp[-(x + 1)] \tag{6}$$

To confirm that this yields the geometric distribution, let its governing parameter $p = 1 - \exp(-\lambda)$ where λ governs the exponential. Suppose $\lambda = 3$. In R, symbolise λ as l and then compute:

```
l = 3
p = 1 - exp(-l)
x = 0:9
exp(-l * x) - exp(-l * (x + 1))
dgeom(x,p)
```

2.2 | New equations

The exact equations for the three new distributions examined in this paper follow easily from the workflow. First, we consider a distribution related to the discrete Weibull (Nakagawa & Osaki, 1975), whose general form can be derived from the exponential distribution in this way:

$$X = [-\ln(U)/\lambda]^p \qquad (7)$$

$$F_X(x) = 1 - \exp(-\lambda\, x^{1/p}) \qquad (8)$$

$$p_X(x) = \exp(-\lambda\, x^{1/p}) - \exp\{-[\lambda\, (x + 1)^{1/p}]\} \qquad (9)$$

where λ and p are constants, the former just being the familiar rate parameter of the exponential distribution.

The specific distribution used here, called the half-power (HP), follows from setting $p = 2$:

$$X = [-\ln(U)/\lambda]^2 \qquad (10)$$

$$F_X(x) = 1 - \exp(-\lambda\, x^{0.5}) \qquad (11)$$

$$p_X(x) = \exp(-\lambda\, x^{0.5}) - \exp\{-[\lambda\, (x + 1)^{0.5}]\} \qquad (12)$$

The $p = 2$ assumption is made because a very simple population dynamics model discussed below implies this value. Assuming any other value would require burdening the model with extra assumptions.

Because $\exp(-\lambda\ 0^{0.5}) = 1$ and $\exp[-(\lambda\ 1^{0.5})] = \exp(-\lambda)$, this equation yields a remarkably simple species richness estimator:

$$R = S/\exp(-\lambda) \qquad (13)$$

where R = estimated richness and S = the observed number of species.

The second distribution, called the scaled odds, uses a scaling constant μ and has a simplified PMF:

$$X = \mu\ (1/U - 1) \qquad (14)$$

$$F_X(x) = 1 - \mu/(x + \mu) \qquad (15)$$

$$p_X(x) = [\mu/(x + \mu)] - [\mu/(x + 1 + \mu)]$$

$$p_X(x) = 1/[(x + \mu)\ (x + 1 + \mu)] \qquad (16)$$

$$R = (\mu + 1)/\mu\ S \qquad (17)$$

Crucially, $1/U - 1$ can be rearranged as $(1 - U)/U$. This ratio is nothing other than the gambler's odds of a random outcome – where the probability of that outcome is itself a random uniform variate. Odds distributions range from zero to infinity, meeting the requirement that abundances on a continuous or discrete scale must fall into that range.

Finally, the quarter-power distribution incorporates features of both equations. Specifically, a modified odds component is logged, scaled, and raised to a power. The power term could be freed to create a two-parameter model comparable to, say, the Weibull. Very close fits to real and simulated data are seen with a power of 4, implying that the expression's form is realistic and the constant is canonical. The constant may reflect an equilibrium state: a different one would presumably result in unstable and transient communities. It is denoted with the symbol γ:

$$X = [-\ln(2/U - 1)/\gamma]^4 \qquad (18)$$

Note that $–\ln(2/U – 1)$ has bounds of zero and infinity, with a self-evident median of ln 3 and a computable mean of ln 2. It is very important that the expressions $–\ln(U)$, $1/U – 1$, and $–\ln(2/U – 1)$ collectively encompass the set of simple expressions that can convert U into this range.

The other equations are:

$$\gamma X^{1/4} = –\ln(2/U – 1)$$

$$U = 2/[\exp(–\gamma X^{1/4}) + 1] \quad (19)$$

$$F_X(x) = 1 – 2/[\exp(–\gamma x^{1/4}) + 1] \quad (20)$$

$$p_X(x) = 2/\{\exp[–\gamma (x + 1)^{1/4}] + 1\} – 2/[\exp(–\gamma x^{1/4}) + 1] \quad (21)$$

The richness estimate requires a little work:

$$p_X(0) = 2/[\exp(–\gamma 1^{1/4}) + 1] – 2/[\exp(–\gamma 0^{1/4}) + 1]$$

$$p_X(0) = 2/[\exp(–\gamma) + 1] – 1$$

$$1 – p_X(0) = 2 – 2/[\exp(–\gamma) + 1]$$

$$1 – p_X(0) = 2 \exp(–\gamma)/[\exp(–\gamma) + 1]$$

$$R = [\exp(–\gamma) + 1]/[2 \exp(–\gamma)] S \quad (22)$$

It is important to stress two other things. First, unlike the log series (Fisher et al. 1943), all of these distributions directly imply the total species richness of a community (eqns. 13, 17, and 22). Likewise, a richness estimate can be gotten out of a Poisson log normal fit because it too indicates the proportion of species with non-zero counts (Grøtan & Engen, 2008). There are issues with that distribution such as its failure to remove sample size biases, its imprecise estimates, and its poor prediction of patterns. The first two topics merit a fuller discussion elsewhere. The third problem is demonstrated in the results. On a conceptual level, I take up what it means to estimate richness from an ecological sample in the discussion.

Second, all of the new models have a single scaling parameter and no shape parameter. In other words, they posit that all differences between species inventories stem from just two properties – the richness of the overall species pool and the number of drawn individuals. Suppose a real-world distribution is ably described by any such distribution. Then all measures that concern distributional evenness here are irrelevant, because if a shape doesn't vary, then there is nothing for an "evenness" metric to describe. I discuss later how this deduction bears on the widespread use of Hill numbers (Hill, 1973; Chao et al., 2014).

2.3 | Additional distributions

There is a large literature on species-abundance distributions in the general sense (McGill et al., 2007). I restrict my discussion to eight published models that have received substantial attention from ecologists at different points in history. (1) The geometric series distribution (Motomura, 1932) was originally applied to RADs. This application has been thought to yield unrealistic fits to data, and the model is no longer considered viable in such a form (Alroy, 2015; Baldridge et al., 2016). However, its fate is different in the current analysis, which applies the distribution to SADs instead. (2) The log series (Fisher et al., 1943) is fundamental to ecology and already considered by some to be a good descriptor of many communities (Baldridge et al., 2016), especially local ones (Antão et al., 2021). This explains why it is still routinely used in biodiversity studies, including very large-scale ones (e.g., Buzas et al., 2002; Cazzolla Gatti et al., 2022). (3) The broken stick distribution (MacArthur, 1957) has a distinct theoretical basis and makes distinct predictions about the shapes of SADs, so it is investigated here even though modern studies reject it (Alroy, 2015). The remaining distributions must be considered because of their recent advocacy. (4) The Poisson log normal (PLN: Bulmer, 1974) was applied to large-scale marine data sets by Connolly et al. (2005, 2009). (5) The zero-sum multinomial (ZSM: Hubbell, 1997, 2001) is widely advocated and has long been the subject of much debate (e.g., McGill, 2003). (6) The negative binomial was explored by Connolly et al. (2009) and Connolly and Thibaut (2012) and also applied by Tovo et al. (2017) and ter Steege et al. (2020), as part of a broader study. (7) The Weibull, a standard statistical distribution, was put forth as a good description of ecological count data by Ulrich et al. (2018). I consider the discrete version of the Weibull

(Nakagawa & Osaki, 1975). (8) The Zipf is another classic distribution and was thought to be a good general descriptor of ecology communities by Su (2018).

I put aside the gambin distribution (Ugland et al., 2007; Matthews et al., 2019) for the same reasons as Ulrich et al. (2018): it is a heuristic pattern descriptor not based in a process model and one that is fit by binning the data, so a direct comparison based on fitting alternatives to proper SADs is not possible. In particular, the *gambin* R library (Matthews et al., 2014) was not designed to fit SADs. I also do not consider niche preoccupation models such as the ones proposed by Sugihara (1980) and Tokeshi (1990) because these RAD-based theories are no longer endorsed, depend on strong assumptions about competition, and do not make clear predictions about SADs.

2.4 | Likelihood-based fitting method

Fitting models to abundance distributions is a challenging problem (Connolly & Thibaut, 2012; Matthews & Whittaker 2014; Ulrich et al., 2018). Earlier researchers sought to do so by sorting counts into $\log_2$ bins (Preston, 1962). However, even when maximum likelihood methods are used (McGill, 2003) this loses much information. Thus, it is impractical when dealing with routine ecological surveys including only 10, 20 or even 30 species (Ulrich et al., 2018). Meanwhile, directly fitting RADs (e.g., Ulrich et al., 2018) is problematic because (1) it depends on frequentist methods such as least-squares or major axis regression; (2) there is no way to specify an error distribution that should apply fairly to all theoretical models; and (3) the data violate the standard statistical requirement of independence between x- and y-values. Specifically, it is not possible to model error in ranks sensibly because stochastic variation in counts would generate swaps in ranks. I therefore follow others (Bulmer, 1974; Connolly et al., 2005, 2017; Connolly & Thibaut, 2012; Prado et al., 2018; Antão et al., 2021) in evaluating model fit by computing the likelihoods of empirical SADs. Again, the term SAD is used here for a list of counts of species sharing particular counts of individuals.

Before continuing, I note that the same likelihood calculation is used in this paper for two purposes: (1) quantifying the fit of each and every rival model to any given SAD, and (2) finding the best value of the parameters of the new models. The function is also used to fit the broken stick, geometric series, negative binomial, and discrete Weibull, which lack trivially computed parameters (the log series has one) and lack existing R functions that fit the parameters by maximum likelihood (the Poisson log normal has one).

The math depends on first computing the independent probability p_i that a given species will fall in its observed count class i, i.e., the likelihood. The overall likelihood is just the product of all the p_i values for the counts (Prado et al., 2018). Of course, only the observed counts can be predicted and the sum of p_i over all observable classes has to be 1. However, zero counts can't be observed and do feature in the PMF equations given above. Therefore, the p values have to be divided by $1 - p_0$ (meaning standardised). Connolly et al. (2017, their eqn. 8) used the same correction.

Connolly and Thibaut (2012) proposed a multinomial equation for fitting SADs instead of a binomial equation. Nothing is wrong with that. However, when it comes to actual computation the distinction is not important: the only difference between an indepent-draws equation and a multinomial equation is the inclusion of combinatorial terms made up of S and s_i. Those values are fixed, so the combinatorial terms are fixed across all possible parameter values, leading to identical maximum likelihood solutions. Thus, users of these methods can choose the interpret the fitting procedure as "really" based on a multinomial model if they so choose.

2.5 | Simulations of population dynamics

Simple simulations are used to demonstrate sufficient if not necessary conditions for the geometric series and the three new distributions to arise. The simulations each assume a species pool of 100,000 with initial population sizes of 100, and they continue for 1000 time steps. Death is always a binomial process, meaning that it is per-capita (based on the initial number of adults) with a probability that any one individual will die. Counts of recruits ("births") are randomly drawn from the Poisson distribution. Similar results can be obtained using models that drawn birth counts from the geometric series. A non-capita birth process is assumed because the system is assumed to be either (1) open to a steady influx of propagules, or (2) saturated with subadults that have been generated over a series of intervals instead of arising over just one time step. Therefore, the models could apply either to open or closed systems.

The geometric series model assumes that the death rate is fixed at some fraction (0.1 in the illustrated trials), and that the Poisson parameter of the recruitment rate is a simple random exponential variate with a rate λ. All the other models are variants. The half-power model assumes that the death probability p is a function of the birth rate λ, specifically $p = 1/(\lambda + 1)$.

So the rates are negatively correlated: when $1/(\lambda + 1) = 1$ or 9, $p = 0.5$ or 0.1. The odds model assumes a fixed death probability, here 0.5, and a birth probability of $\exp(-\lambda)/\lambda$. Finally, the quarter-power model also assumes a uniform death rate, again illustrated as 0.5, and a birth rate of λ^3. In the illustrated trial, the birth rate is scaled up by 3 to allow comparison with the other curves.

So the models assume different relationships between birth and death – but populations must somehow stay in a viable range. How is co-existence maintained?

The counter-intuitive reason is that the simulations reach an equilibrium total population size K for each species. For example, let p = the death probability and d = the expected death count, equal to the current population size $p\ n$. Also let b = the expected birth count, equal to $-\ln p$ in this hypothetical model. At equilibrium, then, $d = b$ and $p\ K = \ln p$, so $K = \ln(p)/p$. Below equilibrium, $n < \ln(p)/p$ because $n < K$ and $K = \ln(p)/p$. Therefore, $d < b$: $n < b/p$ because $b = \ln p$, $p\ n < b$ by rearrangement, and $d < b$ because $d = p\ n$. As a result, n will climb towards K. Above K, $n > 1/p^2$ and $d > b$, so n will fall to K. Similar proofs apply to the preceding models. They relate closely to the equilibrial theory of island biogeography (MacArthur & Wilson, 1963), which also assumed per-capita "death" (extinction) and steady, non-per-capita "birth" (immigration).

The fact that all of this is true is easily confirmed by simulation. It is highly important because it specifically predicts that species producing more recruits in total per time step are more common at equilibrium. There are truly "winner" and "loser" species in this paradigm, but all of them have equilibrial population dynamics, so all of them can co-exist.

All of the models assumes high but predictable variance among species in recruitment rates because of fixed differences in traits, but little variance among individuals. Models assuming a geometric sampling process for recruitment would build in greater variance. They are not explored in this paper because low variance may be more intuitive to many ecologists.

2.6 | Empirical data

Four large-scale data sets and one database of local-scale species inventories were used to benchmark the distributions. Data for communities of fishes and corals spread across the western and central Pacific were drawn from Connolly et al. (2017). A regional data set of 18 butterfly communities from Colombia was taken from Cómbita et al. (2021). Combined abundances of trees inventoried in 1946 plots across the Amazon basin were drawn from ter

Steege et al. (2020). Finally, all 3257 available inventories of local tree, insect, and vertebrate communities from around the world were drawn directly from the Ecological Register database (Alroy, 2015, 2024). A large majority apply to a single trophic level and a small local area. There was no combination of inventories and multiple inventories from the same publications were allowed to be included. After discarding inventories with less then four species, a maximum count of less than four, or entirely identical counts, 3095 remained.

2.7 | Assessment of model fit

The fit of the 11 models to each of the local data sets was assessed by computing the corrected Akaike information criterion (AICc) for each combination (Hurvich & Tsai, 1993). Antão et al. (2021) did the same thing. The above-mentioned likelihood calculation was used as the basis of the computations, which were implemented in the *richness* R package (https://github.com/johnalroy/richness/releases/tag/v2.4). The Zipf and ZSM distributions were fit first using the *sads* library (Prado et al., 2018), which uses the same likelihood equation as *richness* for all of its SAD fitting. The *poilog* library (Grøtan & Engen, 2008) in combination with the *richness* function *pln* was used to fit the PLN. The other models were fit using this paper's maximum likelihood equation, as implemented in the *richness* package.

The AICc statistic penalises weakly for the number of parameters in a model (either one or two in all cases), so it tends to favour more complex ones. Many data sets are small in terms of both the number of species and the number of individuals, so raw AICcs can be misconstrued to indicate meaningful differences. To avoid being misled by stochastic variation in the fits, I tallied cases where differences (Δs) in AICcs yielded a weight of > 20, i.e., where $\exp(\Delta\text{AICc}/2) > 20$.

Complex models are able to fit a wide range of distribution shapes by definition, but this does not necessarily mean they are good predictors of community structure. The reason is that they overfit, so they commit strongly to a pattern that may result from random variation in counts. To show whether models could generalise, I carried out more head-to-head comparisons by (1) fitting each model to each species inventory; (2) for each inventory, selecting another one that represented the same ecological group and the same biogeographic realm (ecozone) and had the most similar numbers of non-singleton and singleton species based on the sum of log ratios of those counts (with the first-encountered inventory being chosen when there was a tie); and (3) computing the log likelihoods (LLs) of the second distribution based on the first one's models. The above methodology was used to obtain the likelihoods. A likelihood weight cutoff of > 20, meaning exp(ΔLL) > 20, was used to flag the decisive comparisons.

2.7 | Multivariate ordination based on fit statistics

Differential sampling of the range of possible SADs might skew tallies of the best distributions for the inventories. Therefore, it is more illuminating to see which shapes across the range are able to to fit which distributions, and whether the new models can account for most or all of this variation. If so, then it is possible that most communities are indeed generated by processes conforming with the key assumptions: per-capita death rates that may or may not be species-specific combined with species-specific, highly variable, and not per-capita recruitment rates.

Principal components analysis of the LLs is used to explore the range of shapes. A level playing field has to be created to make this possible. Specifically, the average magnitude of LLs regardless of the model tracks richness and sample size, rising with both. To account for this, the LLs for each inventory are first standardised to fall in the range between the minimum and maximum. So if the LLs for three models are 10, 13, and 20, then the standardised values are 0, 0.3, and 1. Alternative approaches would depend on making strong assumptions, such as strong and linear tracking between average LLs and either richness, sample size, or both somehow combined.

3 | Results

3.1 | Simulated SADs

Patterns closely consistent with the distributions are yielded by the appropriate simulations. Fits of models to counts are almost precise (Fig. 1). The same patterns can be seen in almost every single trial – these were selected arbitrarily.

The geometric series (Fig. 1A) is the most general, with fixed per-capita death rates and a simple exponential distribution of birth rates. The half-power (HP) model (Fig. 1B) assumes coupling between rates. Finally, the scaled odds and quarter-power (QP) distributions assume fixed death rates and high-variance birth rates (Figs. 1C, D).

3.2 | Descriptions of empirical SADs

The QP distribution fits all for regional data sets with great accuracy (Fig. 2). The scaled odds distribution fits the Pacific coral and fish data and the Colombian tree data next-best (Figs. 2A, B, C). The log series is second-best with the composite Amazonian tree inventories (ter Steege et al., 2020), which span a huge spatial scale (Fig. 2D). Thus, it is not clear that a multi-parameter model like the negative binomial (ter Steege et al., 2020) is really needed for this data set.

In terms of the local-scale data, an initial vetting of the models can be based on head-to-head comparisons that yield large differences in AICcs (AICc weights > 20: Table 1). Here, the three new distributions are decisively better than the broken stick, geometric series, negative binomial, and Zipf. They also beat the zero-sum multinomial (ZSM). The QP overwhelmingly beats the log series while the others fall to it. The Poisson log normal (PLN) and Weibull fare worse again against the QP. This is a mixed result for the HP and odds, and it suggests that the QP is the strongest of all considered distributions.

The fair performance of the two-parameter PLN, Weibull, and ZSM models may be an artefact of (1) the AICc's weak penalisation for model complexity, (2) overfitting, and (3) the ability of complex models to mimic distributions generated by simpler processes, including those that underlie the four models emphasised here.

3.3 | Predictions of empirical SADs

The differences are much more dramatic when fitted SADs are used to predict matched SADs (Table 2). The QP distribution now trumps all of the old models at least 80% of the time when the likelihood weight is > 20. The scaled odds distribution is also strong, with a minimum win percentage of 71. The HP more or less ties the three distributions that predict gently declining, J-shaped RADs: the log series, PLN, and Weibull. The HP and Zipf also tie.

In sum, because accurate prediction is more important than simple description in science, the large differences in favour of all three new models, and especially QP, yield them considerable credence. This conclusion is strengthened by limiting the comparisons to complex distributions (those having highest log likelihoods across all models > 100). This time, the QP and scaled odds respectively beat the two-parameter distributions at least 93 and

86% of the time in all cases. The HP also performs better, still basically tying the Zipf (45%) but now overcoming the log series (80%), PLN (67%), and Weibull (62%).

There is some important variation among 13 ecological groups with respect to relative model performance. QP is far and away the strongest, not falling below 50% in any of the 13 x 8 = 104 comparisons with older models. The scaled odds distribution is also favoured strongly, but not so much over the log series, which it usually beats about 70 – 80% of the time. However, this ranges from 49% (mosquitoes) to 81% (birds). Support for the HP distribution is less impressive (sometimes < 50% in various comparisons) when it comes to four major groups: ants, dung beetles, mosquitoes, and trees. The three insect groups often feature steep distributions that are well-explained by the odds and QP models. There is no obvious latitudinal pattern in the data.

3.4 | Multivariate ordination patterns

The ordination is even more interesting because it shows which shapes go with which distributions, and thus which shapes are broadly applicable (Fig. 3). The classic J-shaped RAD pattern is only seen at left. The other side encompasses flattened and symmetrical RADs only well described by two classic but underlooked one-parameter distributions: the broken stick and much more often the geometric series (red points). The log series (yellow points) is common only at upper left, and specifically matches RADs that start with a hook and trail off into a straight line (as illustrated).

Importantly, the two-parameter distributions (turquoise points) that are of so much interest to ecologists are only common in the central zone of the space, plus part of the branch to the right (Fig. 3). In particular, they explain some J-shaped RADs that are curved in the middle instead of running straight. In other words, the Poisson log normal, and Weibull mostly serve to wrap around unremarkable distributions.

Finally, numerous data sets fit at least one of the three new models well, with relevant inventories (light blue points) falling almost everywhere to the left of the small "flat RAD" zone (Fig. 3). Thus, the new distributions are jointly able to account for most shapes. They are also distinct (Fig. 4). The HP distribution spans a wide region (dark blue points). The odds distribution (violet points) and QP distribution (green points) split the densely-populated left side, which includes many distributions that are J-shaped but steep. Like the Zipf, they can fit broad distributions with hyper-abundant dominant species. But they can also account for the straightness of the log series-type RADs.

4 | Discussion

4.1 | Inference of process

For many years, ecologists were optimistic about inferring processes from species abundance distributions (Fisher et al., 1943; MacArthur, 1957; Preston, 1962; May, 1975; Sugihara, 1980; Hughes, 1986; Tokeshi, 1990; Hubbell, 2001). However, influential papers such as McGill et al. (2007) have more recently argued that because there are so many models making such similar predictions, the entire enterprise is doomed.

This perspective overlooks the basic logic of the current analysis: whenever a population model M exactly predicts a distribution D, rejecting D based on empirical data also rejects M. Thus, fitting SADs can be considerably informative – but only when distributions are simple and grounded in models. In fact, the three new one-parameter distributions actually do predict patterns well (Figs. 1 – 4, Tables 1 and 2). Therefore, they actually do inform us about fundamental ecological processes. By contrast, two-parameter distributions may serve no real purpose because (1) they are not needed to predict the full range of possible SADs (Fig. 3); (2) they are mostly not grounded in simple population dynamical models (as opposed to Fig. 1); and (3) science operates on the principle that simple theories are better.

The proposed population models are ecologically interesting and important for several other major reasons. (1) All of them are not only simple, but simple variants of each other. (2) They assume high variance in recruitment rates among species but low variance among individuals within species. By contrast, the fully neutral log series model assumes no consistent, trait-based variation in demographic rates among species (Kendall, 1948; Hubbell, 2001). In the new models, species do have systematically different demographic rates and equilibrium population sizes because of their traits, so there are "winners" and "losers" in perpetuity. (3) The models imply that populations reach equilibrium strictly because of demographic tradeoffs (Fig. 1). There is no role for competition, niche preoccupation, assembly rules, speciation, extinction, or any other non-local, non-random process. Thus, they are bona fide null models that are even simpler and less assumption-laden than that of Hubbell (1997, 2001).

4.2 | Implications for quantifying biodiversity

In recent years, ecologists have also moved to the idea that communities should be assessed by computing Hill numbers (Hill, 1973) such as Shannon's *H* and Simpson's *D* (Roswell et al., 2021). Chao et al. (2014) seems to have provided much momentum in this direction. Hill numbers blend information about richness and evenness, and ecologists use them in the hope that the latter can be quantified independent of sample size. But this hope may be in vain for three reasons.

First, blended statistics are dubious from a philosophical point of view. Statisticians prefer to develop one descriptive statistic per property. Second, evenness is a transient property of ecosystems driven by the random success of particular species in particular places at particular times. By contrast, richness is non-transient because it is governed by processes operating on geological time scales: speciation, extinction, and dispersal. Third, one-parameter distributions vary based on sampling intensity (scale) and richness but not based on shape, and Hill numbers vary meaningfully only when "evenness" varies. Because these distributions often hold, Hill numbers only indicate that some distributions are intrinsically steep and some are shallow, with this steepness being an inflexible property of no interest on its own.

A further motivation for the evenness-not-richness philosophy is the notion that the richness of any community is not only unknown from raw data, but unknowable in general. There are actually two arguments of this kind. The first is just that existing methods don't work because their estimates are usually either too low or highly imprecise (Roswell et al., 2021). When the assumptions of the new methods are met, their estimates cannot be greatly biased because they depend on maximum likelihood estimates of single parameters. Likewise, the arithmetic mean of a legitimately normal distribution can't be consistently biased because the mean is the maximum likelihood value of the central tendency. Although there is no room here to say much more about the matter, the fact that such estimates are accurate and precise would merit a fuller discussion elsewhere.

The second proposition is that the effective sampling universe is a function of the size of an inventory: the more individuals counted, the spatiotemporally larger and therefore richer the sampled community. This argument conflates two things: (1) the number of species that would be found in an infinitely large inventory, and (2) the number of species that existed in the spatiotemporal realm that encompassed the sampling point (i.e., the community). This paper's richness equations are about the latter, not the former.

4.3 | Adequacy of the new analyses and models

It has long been agreed that a comparative study of species abundance distributions must compare multiple models by investigating multiple data sets (McGill et al., 2007). However, previous analyses have tended to consider quite different and often limited sets of distributions (Hughes, 1986; Ulrich & Ollik, 2005; Ugland et al., 2007; Ulrich et al., 2010; Connolly et al., 2014; Matthews et al., 2014, 2019; Alroy, 2015; Baldridge et al., 2016; Su, 2018; Antão et al., 2021). Many have included one version or another of both the log normal and log series (e.g., Antão et al., 2021), if not always (e.g., Su, 2018). For example, the log series is a special case of the negative binomial (Fisher et al., 1943) and the latter has been tested against the Poisson log normal (Connolly et al., 2014). Past that, coverage is eclectic.

Thus, few studies are comparable to this one. In the face of this comprehensiveness, support for the new distributions is jointly clear when one considers their ability to predict new sets of counts from old ones (Table 2, Figs. 3, 4). It is reasonable to ask whether additional one-parameter distributions might also be sound from both a descriptive view (Table 1) and a predictive view (Table 2). But only the geometric series and log series come even close to passing both of these tests. The latter is profoundly skeptical because it assumes that communities are drawn from pools with infinite richness (Fisher et al., 1943). It also assumes that species are identical in terms of population dynamics, in which respect it may take null modelling a bit too far. After all, this assumption discards the entire premise of trait-based ecology. Thus, the three newly proposed distributions are not only jointly adequate but arguably more sensible. One way or another, it is fair to suggest that the structure of many or even most communities does actually result from extremely simple dynamical processes.

Acknowledgments

I thank Michael Foote, Matthew Kosnik, and colleagues at Macquarie University, the University of Nebraska, Lincoln, and the University of Tasmania for helpful discussions. Hans ter Steege and Werner Ulrich commented on the text. The author is the recipient of a Discovery Project Award (project number DP210101324) funded by the Australian Government.

Conflict of interest statement

The author has no conflict of interest to declare.

Data availability statement

The data are available from the Dryad digital repository (https://datadryad.org/stash/dataset/doi:10.5061/dryad.brv15dvdc).

References

Alroy, J. (2010). The shifting balance of diversity among major marine animal groups. *Science,* 329, 1191-1194.

Alroy, J. (2015). The shape of terrestrial abundance distributions. *Science Advances,* 1, e1500082.

Alroy, J. (2024). Data from: three models of ecological community assembly: terrestrial species inventories. *Dryad.* https://doi.org/10.5061/dryad.brv15dvdc

Antão, L. H., Magurran, A. E., & Dornelas, M. (2021). The shape of species abundance distributions across spatial scales. *Frontiers in Ecology and Evolution,* 9, 626730. https://doi.org/10.3389/fevo.2021.626730

Baldridge, E., Harris, D. J., Xiao, X., & White, E. P. (2016). An extensive comparison of species-abundance distribution models. *PeerJ,* 4, e2823.

Borda-de-Água, L., Borges, P. A. V., Hubbell, S. P., & Pereira, H. M. (2011). Spatial scaling of species abundance distributions. *Ecography,* 35, 549-556.

Bulmer, M. G. (1974). On fitting the Poisson lognormal distribution to species-abundance data. *Biometrics,* 30, 101-110.

Buzas, M. A., Collins, L. S., & Culver, S. J. (2002). Latitudinal difference in biodiversity caused by hgigher tropical rate of increase. *Proceedings of the National Academy of Sciences USA,* 99, 7841-7843.

Cazzolla Gatti, R. et al. (2022). The number of tree species on Earth. *Proceedings of the National Academy Sciences USA,* 119, e2115329119.

Chao, A. (1984). Nonparametric estimation of the number of classes in a population. *Scandinavian Journal of Statistics,* 11, 265- 270.

Chao, A., Gotelli, N. J., Hsieh, T. C., Sander, E. L., Ma, K. H., Colwell, R. K., & Ellison, A. M. (2014). Rarefaction and extrapolation with Hill numbers, a framework for sampling and estimation in species diversity studies. *Ecological Monographs,* 84, 45-67.

Cómbita, J. L., Giraldo, C. E., & Escobar, F. (2021). Data from: Environmental variation associated with topography explains butterfly diversity along a tropical elevation gradient, Dryad, Dataset, https://doi.org/10.5061/dryad.vx0k6djsn.

Connolly, S. R., et al. (2014). Commonness and rarity in the marine biosphere. *Proceedings of the National Academy of Sciences USA,* 111, 8524-8529.

Connolly, S. R., Dornelas, M., Bellwood, D. R., & Hughes, T. P. (2009). Testing species abundance models, a new bootstrap approach applied to Indo-Pacific coral reefs. *Ecology,* 90, 3138-3149.

Connolly, S. R., Hughes, T. P., & Bellwood, D. R. (2017). A unified model explains commonness and rarity on coral reefs. *Ecology Letters,* 20, 477-486.

Connolly, S. R., Hughes, T. P., Bellwood, D. R., & Karlson, R. H. (2005). Community structure of corals and reef fishes at multiple scales. *Science,* 309, 1363-1365.

Connolly, S. R., & Thibaut, L. M. (2012). A comparative analysis of alternative approaches to fitting species-abundance models. *Journal of Plant Ecology,* 5, 32-45.

Eliot, C. (2007). Method and metaphysics in Clements's and Gleason's ecological explanations. *Studies in History and Philosophy of Science C,* 38, 85-109.

Fisher, R.A., Corbet, A.S., & Williams, C.B. (1943). The relation between the number of species and the number of individuals in a random sample of an animal population. *Journal of Animal Ecology,* 12, 42– 58.

Grøtan, V., & Engen, S. (2008). poilog: Poisson lognormal and bivariate Poisson lognormal distribution. R package version 0.4.

Henriques-Silva, R., Lindo, Z., & Peres-Neto, P. R. (2013). A community of metacommunities: exploring patterns of species distributions across large geographical areas. *Ecology,* 94, 627-639.

Hill, M. O. (1973). Diversity and evenness: a unifying notation and its consequences. *Ecology,* 54, 627-639.

Hubbell, S. P. (1997). A unified theory of biogeography and relative species abundance and its application to tropical rain forests and coral reefs. Coral Reefs, 16, S9-S21.

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

Hughes, R. G. (1986). Theories and models of species abundance. *American Naturalist,* 128, 879-899.

Hurvich, C. M., & Tsai, C. L. (1993). A corrected Akaike information criterion for vector autoregressive model selection. *Journal of Time Series Analysis,* 14, 271-279.

Kendall, D. G. (1948). On some modes of population growth leading to R. A. Fisher's logarithmic series distribution. *Biometrika,* 35, 6-15.

Lawton, J. H. (1999). Are there general laws in ecology? *Oikos,* 84, 177-192.

Leibold, M. A., & Mikkelson, G. M. (2002). Coherence, species turnover, and boundary clumping: elements of metacommunity structure. *Oikos,* 97, 237-250.

Loreau, M., & Mouquet, N. (1999). Immigration and the maintenance of local species diversity. *American Naturalist,* 154, 427-440.

MacArthur, R. H. (1957). On the relative abundance of bird species. *Proceedings of the National Academy of Sciences USA,* 43, 293-295.

MacArthur, R. H. (1960). On the relative abundance of species. *American Naturalist,* 94, 25-36.

MacArthur, R. H., & Wilson, E. O. (1963). An equilibrium model of insular zoogeography. *Evolution,* 17, 373-387.

Magurran, A. E. (2007). Species abundance distributions over time. *Ecology Letters,* 10, 347-354.

Matthews, T. J., Borregaard, M. K., Gillespie, C. S., Rigal, F., Ugland, K. I., Ferreira Krüger, R., Marques, R., Sadler, J. P., Borges, P. A. V., Kubota, Y., & Whittaker, R. J. (2019). Extension of the gambin model to multimodal species abundance distributions. *Methods in Ecology and Evolution,* 10, 432-437.

Matthews, T. J., Borregaard, M. K., Ugland, K. I., Borges, P. A. V., Rigal, F., Cardoso, P., & Whittaker, R. J. (2014). The gambin model provides a superior fit to species abundance distributions with a single free parameter: evidence, implementation and interpretation. *Ecography,* 37, 1002-1011

Matthews, T. J., & Whittaker, R. J. (2014). Fitting and comparing competing models of the species abundance distribution: assessment and prospect. *Frontiers of Biogeography,* 6, 67-82.

May, R. M. (1975). Patterns of species abundance and diversity. In M. L. Cody & J. M. Diamond (Eds.), *Ecology and evolution of communities.* Belknap.

McGill, B. J. (2003). A test of the unified neutral theory of biodiversity. *Nature,* 422, 881-885.

McGill, B.J., Etienne, R.S., Gray, J.S., Alonso, D., Anderson, M.J., Benecha, H.K., et al. (2007). Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. *Ecology Letters,* 10, 995– 1015.

Motomura, I. (1932). A statistical treatment of associations. *Japanese Journal of Zoology,* 44, 379-383.

Nakagawa, R., & Osaki, S. (1975). The discrete Weibull distribution. *IEEE Transactions on Reliability,* 24, 300-301.

Prado, P. I., Dantas Miranda, M., & Chalom, A. (2018). sads: maximumum likelihood models for species abundance distributions. R package version 0.4.2.

Presley, S. J., Higgins, C. L., & Willig, M. R. (2010). A comprehensive framework for the evaluation of metacommunity structure. *Oikos,* 119, 908-917.

Preston, F. W. (1962). The canonical distribution of commonness and rarity of species. *Ecology,* 43, 410-432.

Roswell, M., Dushoff, J., & Winfree, R. (2021). A conceptual guide to measuring species diversity. *Oikos,* 130, 321-338.

Saether, B. E., Engen, S., & Grøtan, V. (2013). Species diversity and community similarity in fluctuating environments: parametric approaches using species abundance distributions. *Journal of Animal Ecology,* 82, 721-738.

Su, Q. (2018). A general pattern of the species abundance distribution. *PeerJ,* 6, e5928.

Sugihara, G. (1980). Minimal community structure: an explanation of species abundance patterns. *American Naturalist,* 116, 770-787.

ter Steege, H. et al. (2020). Biased-corrected richness estimates for the Amazonian tree flora. *Scientific Reports,* 10, 10130.

Tokeshi, M. (1990). Niche apportionment or random assortment: species abundance patterns revisited. *Journal of Animal Ecology,* 59, 1129-1146.

Tovo, A., Suweis, S., Formentin, M., Favretti, M., Volkov, I., Banavar, J. R., Azaele, S., & Maritan, A. (2017). Upscaling species richness and abundances in tropical forests. *Science Advances, 3,* e1701438.

Ugland, K. I., Lambshead, P. J. D., McGill, B., Gray, J. S., O'Dea, N., Ladle, R. J., & Whittaker, R. J. (2007). Modelling dimensionality in species abundance distributions: description and evaluation of the Gambin model. *Evolutionary Ecology Research,* 9, 313-324.

Ulrich, W., Nakadai, R., Matthews, T. J., & Kubota, Y. (2018). The two-parameter Weibull distribution as a universal tool to model the variation in species relative abundances. *Ecological Complexity,* 36, 110-116.

Ulrich, W., & Ollik, M. (2005). Limits to the estimation of species richness: the use of relative abundance distributions. *Diversity and Distributions,* 11, 265-273.

Ulrich, W., Ollik, M., & Ugland, K. I. (2010). A meta-analysis of species-abundance distributions. *Oikos,* 119, 1149-1155.

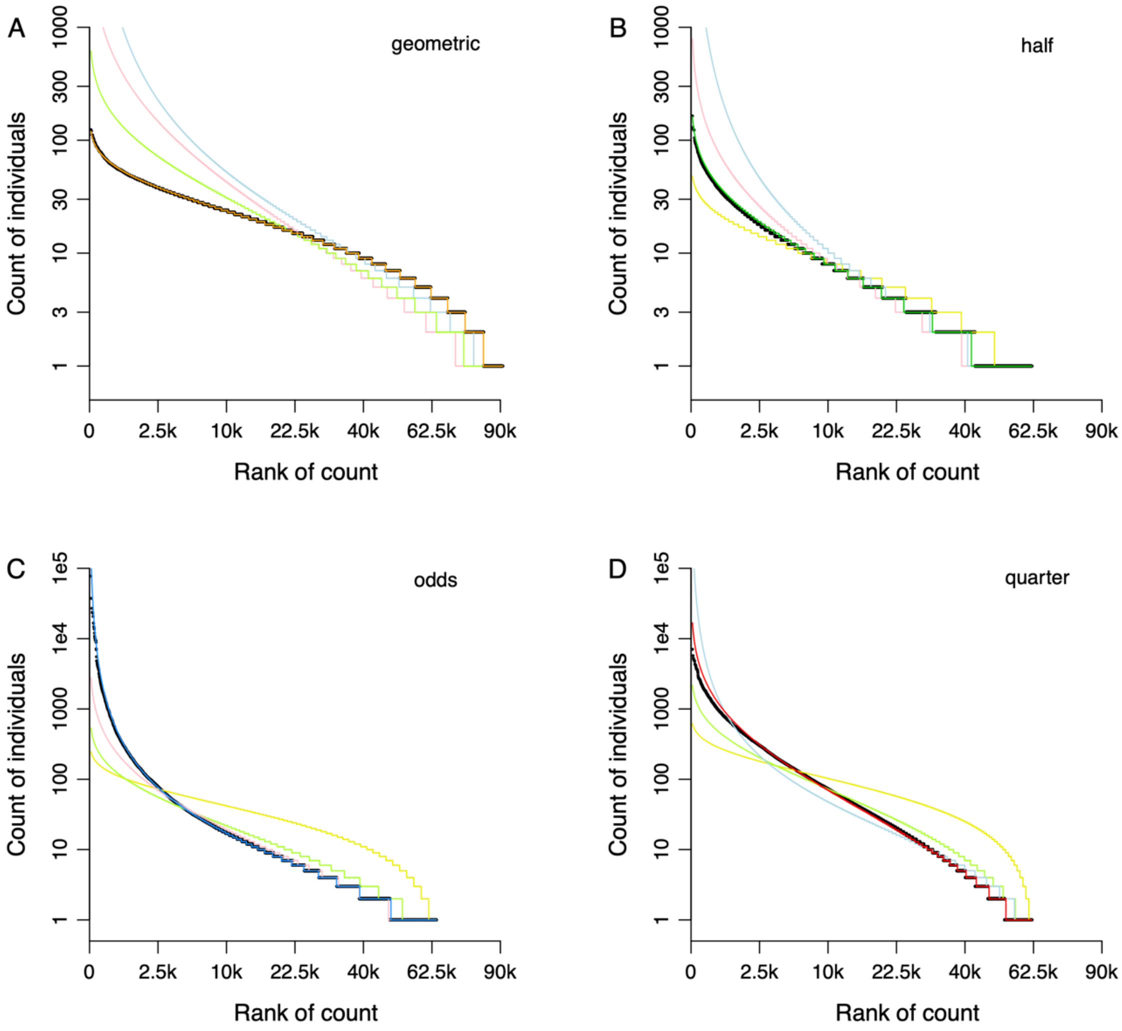

Figure 1. Simulated rank-abundance distributions for pools of 100,000 species. Curves show the raw counts (black lines), geometric series (orange lines), half-power (half) distribution (green lines), scaled odds distribution (blue lines), and quarter-power distribution (red lines). Distributions best-fitting a given model are illustrated in bolder colours. x-axes are square-root transformed; y-axes are log transformed. Recruitment ("birth") counts in each time step follow a Poisson distribution; death counts follow a binomial distribution. Birth rates vary exponentially. (A) Geometric series: the death probability is fixed at 0.1. (B) Half-power model: the death probability is the birth rate λ rescaled as 1/(λ + 1). (C) Scaled odds model: the death probability is 0.5 and the birth rate is exp(–λ)/λ. (D) Quarter power model: the death probability is 0.5 and the birth rate is λ^3.

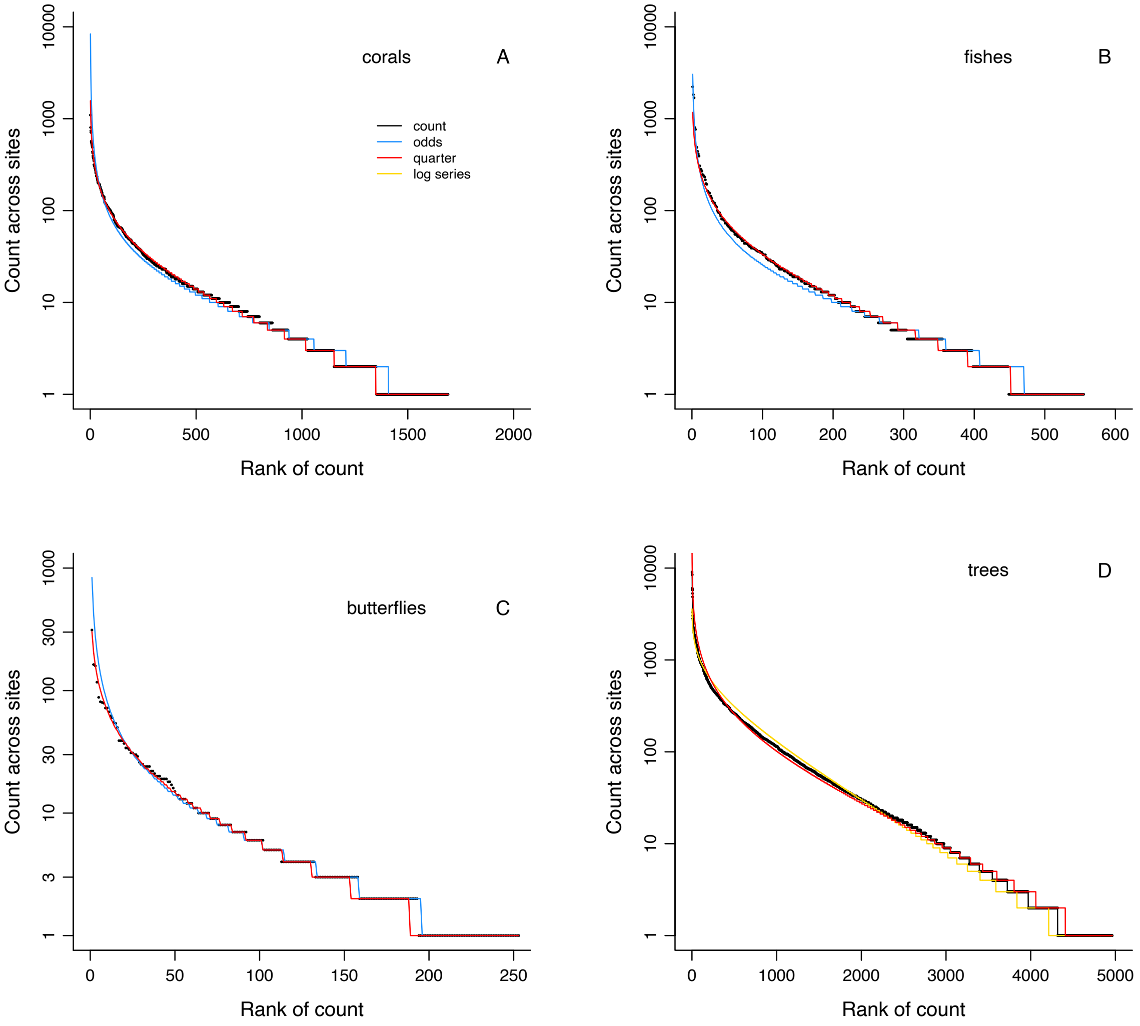

Figure 2. Examples of regional rank-abundance distributions. Black lines; raw counts; light blue lines: scaled odds distribution; red lines: quarter-power distribution; yellow line in (D): log series. The best two distributions in each case are illustrated: the quarter-power model is always best. (A) Corals from the Pacific Ocean (Connolly et al., 2017). Scaled odds is second. (B) Fishes from the Pacific Ocean (Connolly et al., 2017). Odds is second. (C) Butterflies from Colombia (Cómbita et al., 2021). Odds is second. (D) Trees from Amazonia (ter Steege et al., 2020). Log series is second.

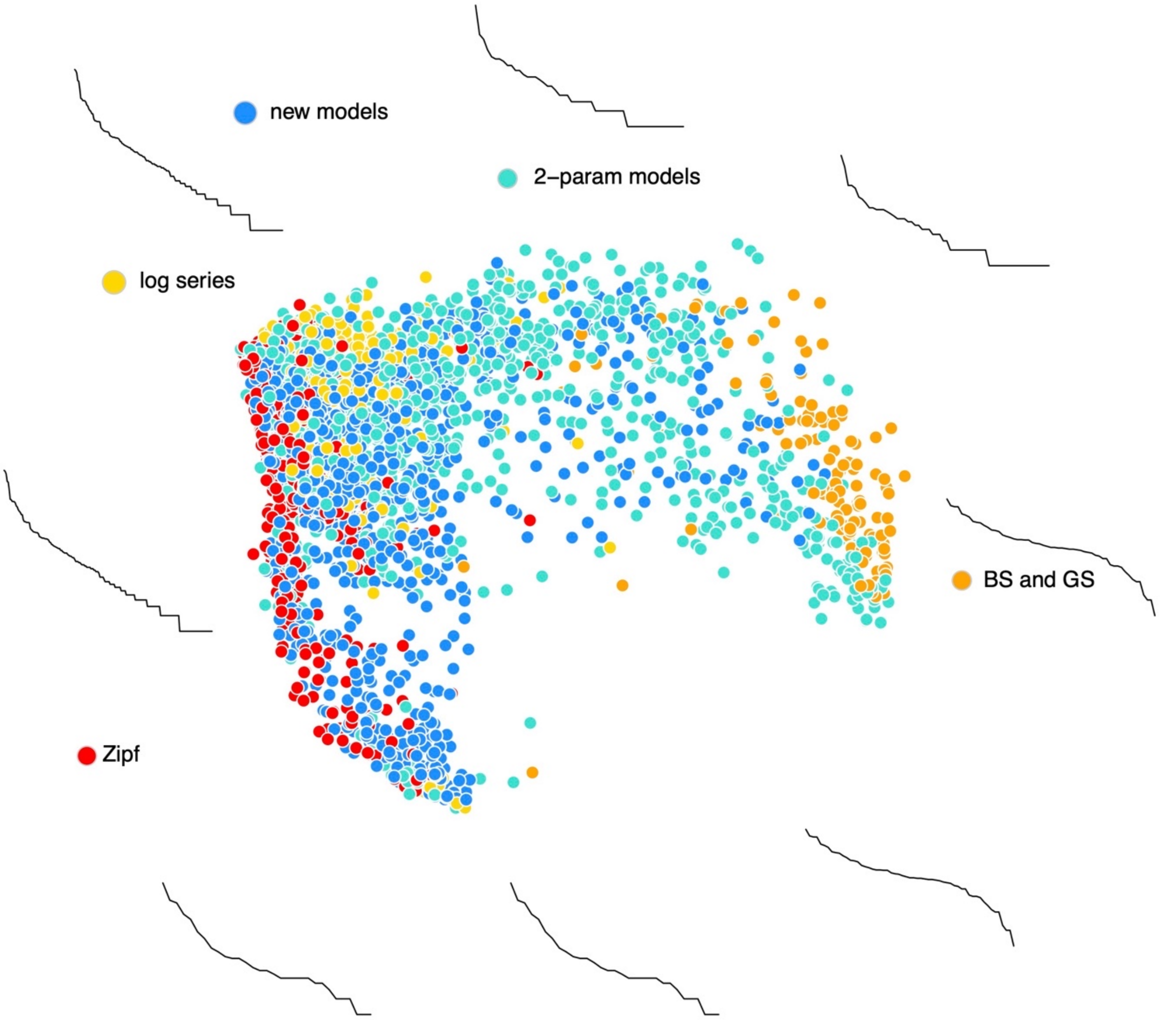

Figure 3. Ordination of species inventories based on the fits of 11 models. Points closer together yield similar log likelihoods. Likelihoods are produced by fitting models to inventories and using the fits to predict distributions for other inventories matched by considering ecological groups, biogeographic regions, and species counts (see text). Data come from the Ecological Register (Alroy, 2015, 2024). Eight lines at the edges illustrate representative rank-abundance distributions each including at least 30 species. Point colours indicate the models that best fit each inventory's data. Blue = the three new methods (half-power exponential, scaled odds, and quarter-power); turquoise = two-parameter models (negative binomial, Poisson log normal, Weibull, and zero-sum multinomial); orange = flat one-parameter models (BS = broken stick and GS = geometric series); red = the Zipf model; yellow = the log series. See the text for references.

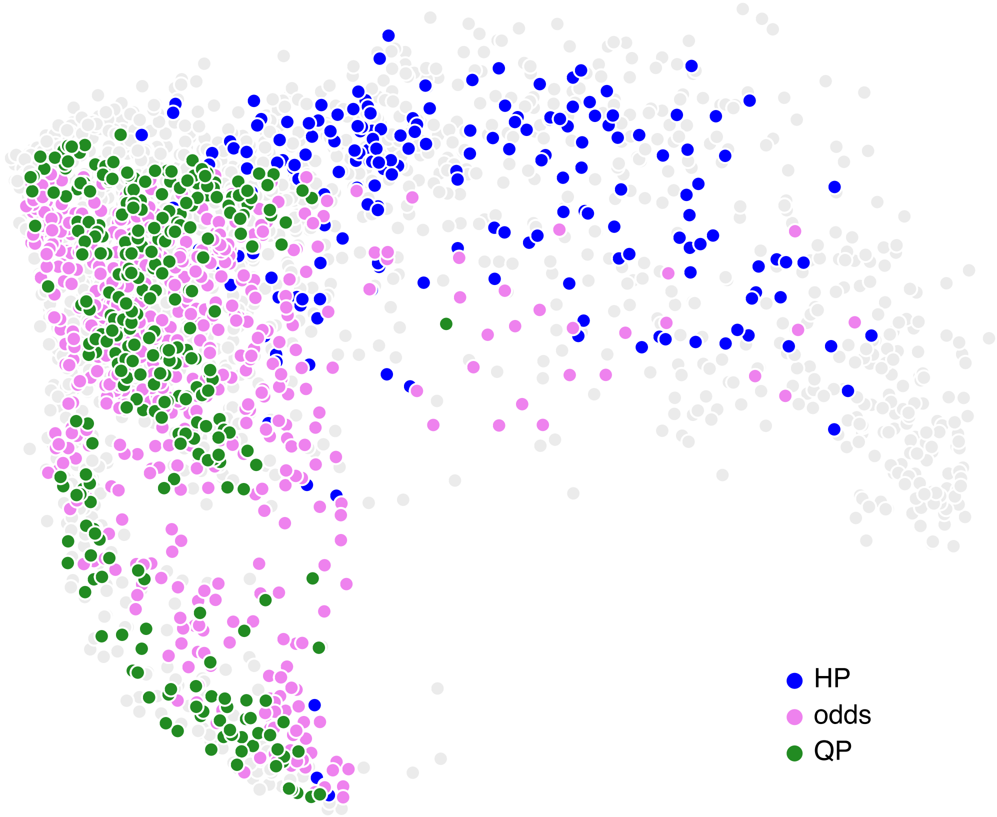

Fig. 4. Ordination of species inventories highlighting the newly proposed distribution models. Data and methods are the same as in Fig. 3. Colours indicate the best models. HP = half-power (blue points); odds = scaled odds (violet); QP = quarter-power (green). Points best fitting the other distributions are in grey.

Table 1. Head-to-head comparisons of 11 species abundance distribution models. Each pair of numbers shows how many published terrestrial ecological inventories are better fit to the column's distribution than the row's distribution according to the corrected Akaike information criterion (Hurvich & Tsai, 1993) with a weight > 20. Proportions > 0.5 are in bold. Data are local-scale inventories drawn from the Ecological Register and reposited on Dryad (Alroy, 2024). Models are explained and referenced in the text. HP = half-power; odds = scaled odds; QP= quarter power; geom. series = geometric series; n. binomial = negative binomial; PLN = Poisson log normal; ZSM = zero-sum multinomial.

	HP	odds	QP	broken stick	geom. series	log series
HP		**222/391**	**297/390**	84/1875	74/1711	**228/346**
odds	169/391		**195/277**	134/1818	119/1653	**316/510**
QP	93/390	82/277		202/1914	188/1777	20/130
broken stick	**1797/1875**	**1684/1818**	**1712/1914**		**389/396**	**1755/1947**
geom. series	**1637/1711**	**1534/1653**	**1589/1777**	7/396		**1633/1814**
log series	118/346	194/510	**110/130**	192/1947	181/1814	
Zipf	**1711/1878**	**1538/1606**	**1748/1773**	897/2181	953/2104	**1811/1864**
n. binomial	**1911/1952**	**1678/1723**	**1803/1889**	821/1400	931/1373	**1899/1996**
PLN	236/478	**246/465**	**247/394**	195/1746	195/1610	**301/512**
Weibull	178/468	179/436	**158/316**	162/1751	167/1624	170/389
ZSM	**474/621**	**575/755**	**390/400**	455/2017	468/1907	**144/145**

	Zipf	n. binomial	PLN	Weibull	ZSM
HP	167/1878	41/1952	**242/478**	**290/468**	147/621
odds	68/1606	45/1723	219/465	**257/436**	180/755
QP	25/1773	86/1889	147/394	**158/316**	10/400
broken stick	**1284/2181**	579/1400	**1551/1746**	**1589/1751**	**1562/2017**

geom. series	**1151/2104**	442/1373	**1415/1610**	**1457/1624**	**1439/1907**
log series	53/1864	97/1996	211/512	**219/389**	1/145
Zipf		482/1460	**1296/1437**	**1352/1419**	**1159/1326**
n. binomial	**978/1460**		**1316/1328**	**1371/1375**	**1351/1500**
Poisson LN	141/1437	12/1328		**88/115**	80/446
Weibull	67/1419	4/1375	27/115		3/375
ZSM	167/1326	149/1500	**366/446**	**372/375**	

Table 2. Head-to-head comparisons of 11 species abundance distribution models based on predictions of counts in matched inventories. Each model is fitted to each inventory in the overall Ecological Register data set (Alroy, 2024) and then projected onto another inventory with similar singleton and non-singleton species counts that represents the same ecological group and ecozone. Each pair of numbers shows how many inventories better fit to the column's distribution than the row's distribution according to the log likelihood of the second count vector, with a relative weight > 20. Proportions > 0.5 are in bold. Data and models are explained and referenced in the text; abbreviations are as in Table 1.

	HP	odds	QP	broken stick	geom. series	log series
HP		**643/781**	**680/710**	22/2321	26/2100	232/576
odds	138/781		**209/380**	20/2247	45/2019	265/902
QP	30/710	171/380		44/2307	57/2097	43/611
broken stick	**2299/2321**	**2227/2247**	**2263/2307**		**1291/1292**	**2273/2318**
geom. series	**2074/2100**	**1984/2029**	**2040/2097**	1/1292		**2055/2114**
log series	**344/576**	**637/902**	**568/611**	45/2318	59/2114	
Zipf	**661/1256**	**730/923**	**756/949**	215/2265	313/2120	**650/1152**
n. binomial	**1677/1697**	**1652/1678**	**1732/1769**	188/1494	625/1433	**1738/1777**
Poisson LN	**459/841**	**508/692**	**544/634**	56/2290	139/2036	**474/770**
Weibull	414/834	**525/739**	**525/609**	32/2272	155/2023	**378/725**
ZSM	**734/930**	**911/1172**	**859/908**	37/2314	330/2086	**642/644**

	Zipf	n. binomial	PLN	Weibull	ZSM
HP	595/1256	20/1697	382/841	420/834	196/930
odds	193/923	26/1678	184/692	218/739	261/1172
QP	193/949	37/1769	90/634	84/609	49/908
broken stick	**2050/2265**	**1306/1494**	**2234/2290**	**2240/2272**	**2277/2314**

geom. series	**1807/2120**	**808/1433**	**1897/2036**	**1868/2023**	**1756/2086**
log series	502/1152	39/1777	296/770	347/725	2/644
Zipf		201/1654	**715/1243**	**719/1245**	662/1468
n. binomial	**1453/1654**		**1652/1706**	**1607/1682**	**1437/1775**
PLN	528/1243	54/1706		**283/532**	252/983
Weibull	526/1245	75/1682	249/532		245/986
ZSM	**806/1468**	338/1755	**731/983**	**741/986**	