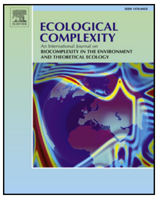
Ecological Complexity 36 (2018) 110–116

Contents lists available at ScienceDirect

Ecological Complexity

journal homepage: www.elsevier.com/locate/ecocom



Original Research Article

The two-parameter Weibull distribution as a universal tool to model the variation in species relative abundances

Werner Ulrich[a,*], Ryosuke Nakadai[b], Thomas J. Matthews[c,d,e], Yasuhiro Kubota[f]

[a] Department of Ecology and Biogeography, Nicolaus Copernicus University, Toruń, Poland
[b] Faculty of Science, University of the Ryukyus, Nishihara, Japan
[c] School of Geography, Earth and Environmental Sciences, University of Birmingham, UK
[d] CE3C – Centre for Ecology, Evolution and Environmental Changes/Azorean Biodiversity Group and Univ. dos Açores – Depto de Ciências e Engenharia do Ambiente, PT-9700-042 Angra do Heroísmo, Açores, Portugal
[e] Birmingham Institute of Forest Research, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK
[f] Marine and Terrestrial Field Ecology, Tropical Biosphere Research Center, University of the Ryukyus, Nishihara, Japan

ARTICLE INFO

Keywords:
Preston plot
Species abundance distribution
Species assembly
Statistical fitting
Weibull distribution
Whittaker plot

ABSTRACT

The study of species abundance distributions (SADs) needs a precise modeling of their drivers and ecological implications. We introduce the two-parameter Weibull distribution as a versatile tool to fit various kinds of observed SADs and to compare observed and theoretically expected values at the species level. We show that the shape and the scale parameter of this distribution have precise ecological interpretations, the first being a measure of the excess of either rare or common species, and the second as a quantification of the proportion of persistent species in the focal community. Applying the Weibull model to 534 global tree communities we demonstrate that plots of the parameters of the Weibull distribution demark ecologically impossible species abundance distributions. This promises new insight into the ecological constraints on community assembly.

1. Introduction

1.1. Theoretical background

Within ecological assemblages, species generally differ widely in abundances (Magurran, 2004; Matthews and Whittaker, 2015). Often, the dominant species exceed the least abundant species by more than five orders of magnitude (Ulrich et al., 2010). Since Motomura (1932) formally introduced the concept of the species relative abundance distribution (SAD), the question of which ecological processes are responsible for this large variance in abundance has been controversially discussed (Tokeshi, 1999; Hubbell, 2001; McGill et al., 2007; Locey and White, 2013). Whilst early SAD models focused on the role of niche (Sugihara, 1980) vs. stochastic processes (May, 1975), recent discussion has largely centered on the impact on the SAD of dispersal (Hubbell, 2001; Dexter et al., 2017) vs. persistence (Magurran and Henderson, 2003). These discussions have sparked the development of a large number of different SAD models, each based on a specific set of assumptions about community assembly (Magurran, 2005; McGill et al., 2007). Irrespective of the ecological and theoretical background, SADs are commonly used for biodiversity assessment and monitoring (Matthews and Whittaker, 2015).

Models of relative abundances need to be fitted to observed distributions, a task that is far from being straightforward (Ulrich et al., 2010; Mathews and Whittaker, 2014; Baldridge et al., 2016). To be of ecological value these fits need to be compared among different assemblages, for instance to infer environmental or geographical gradients (Ulrich et al., 2016a, 2016b) and variation in abiotic conditions and biotic interactions (Erlén and Morris, 2015). Traditionally, SAD models have mostly been fitted to the statistical distribution (Matthews and Whittaker, 2014), where abundances are classified into $\log_2$ bins prior to fitting (Preston, 1948; Fig. (1) left panels). As the selection of the logarithmic base and the way in which abundances are binned are arbitrary decisions, different binning procedures might have major impacts on model fitting (Nekola et al., 2008; Connolly and Dornelas, 2011). Further, any sufficient estimate of class frequencies needs an appropriate number of species, making frequency distribution fits reliable only for larger communities (Wilson, 1993). In contrast, plots that use ranked abundances for all species (Whittaker plots: Whittaker, 1975; Bazzaz, 1975) do not loose information due to data binning (Fig. 1 right panels) or the pooling of species to observed numbers of individuals. As a consequence, fits become reliable at much

---

* Corresponding author.
E-mail addresses: ulrichw@umk.pl (W. Ulrich), T.J.Matthews@bham.ac.uk (T.J. Matthews).

https://doi.org/10.1016/j.ecocom.2018.07.002
Received 3 April 2018; Received in revised form 14 June 2018; Accepted 7 July 2018
1476-945X/

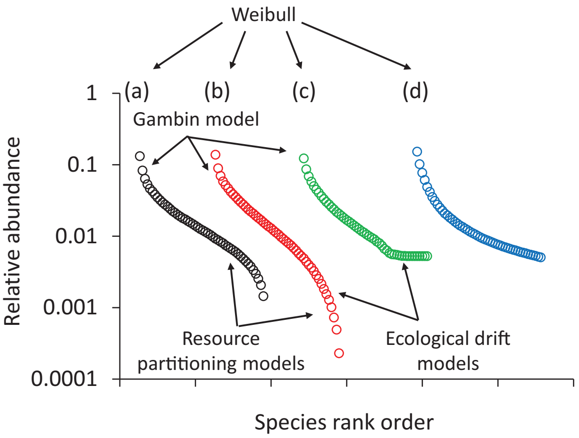


Fig. 1. Four typical shapes of species abundance distributions (Whittaker species – rank order plots). (a) a lognormal distribution, (b) left skewed lognormal with an excess of rare species, (c) a log-series distribution, (d) a power function. Each community contains 50 species. Arrows show which types of distributions the selected flexible SAD models can mimic.

lower species richness. Ulrich et al. (2010) recommended ten species as the lower boundary for model fit, whereas Wilson (1993) reported that even 40 species may not be enough to reliably identify particular types of statistical distributions. Importantly, deviation of the fits from the original rank – abundance plot can be directly traced down to single species and possibly to the underlying processes that influence the abundance of these species. However, rank – abundance fits might fail if an excess of few very abundant or very rare species biases the metric used for fitting (often ordinary least squares in combination with Akaike information maximization). Further, the octaves of the SAD models that are based on a statistical distribution (e.g. the lognormal distribution) and not on an algorithm that directly generates for each species the expected abundance (e.g. most niche division based models, Tokeshi, 1996) need to be interpolated to species abundances prior to fitting, introducing a degree of subjectivism.

As observed SADs differ widely in shape and scale (the range in abundances), current models, that are based on one (shape, for instance the lognormal and most niche division models) or two (shape and scale, particularly the log-series) parameters, have limited variation in shape. Most are not able to accurately mimic all of the main different SAD types and do not cover the whole possible SAD space (i.e. all of the different empirically observed SAD forms). For instance, in recent global comparisons of dryland plant (Ulrich et al., 2016a) and forest tree (Ulrich et al., 2016b) communities, the latitudinal variation in SAD shape required separate fits of two different models. In this situation it became difficult to identify any gradual latitudinal trend. In this respect, Baldridge et al. (2016) reported common distribution based SAD models had weak discrimination power, making any ecological inference challenging.

These problems regarding model fitting and comparison, the biases introduced by the binning of distributions, and the need for multiple model fits call for the development of flexible statistical SAD descriptors that are able to mimic various SAD shapes. In this respect, Ulrich et al. (2010) demonstrated that SADs can generally be classified into three basic shapes. The first shape is a lognormal type statistical distribution characterized by a larger number of species with intermediate abundance and fewer species with high and low abundance (Fig. 1a). In empirical SADs, there is often an excess of rare species than predicted by a lognormal distribution (Fig. 1b). The second shape is equivalent to a log-series sample distribution (Fisher et al., 1943), which is characterized by a few abundant and a larger number of relatively rare species (Fig. 1c). A small number of assemblages, particularly arthropod samples, follow a third shape, a power function SAD characterized by a heavy tail of rare species (Pueyo, 2006; Ulrich et al., 2010) (Fig. 1d). Here, we argue that a versatile SAD model must be able to fit these three basic shapes.

Hughes (1986) was the first to develop a flexible SAD model. However, fitting this model is not straightforward and the model has received little attention. Tokeshi (1996) described a one parameter resource division model that is able to fit the symmetric and skewed lognormal, but not other SAD shapes (Fig. 1). The dynamic model of Dewdney (2000) provides good fits to log-series shaped distributions and possibly also to power functions. Ugland et al. (2007) developed the Gambin model, based on a discrete version of the statistical gamma distribution. These authors argued that Gambin is able to mimic several observed distribution shapes by variation of a single parameter (α); small values of α characterize log-series SAD shapes, while higher values indicate lognormal curve shapes. As Gambin is intended to be a descriptor only, the parameter has no clear ecological interpretation, rather, it is simply a measure of the shape of the SAD. In addition, Gambin is based on a statistical distribution (as used by Preston, 1948) and involves binning the abundance data into octaves before fitting. There is no straightforward way to rescale the Gambin distribution to species abundances.

Of course, neutral, ecological drift models (Hubbell, 2001) provide a mechanistic interpretation of observed abundance distributions with ecologically well-defined parameters. Depending on the probability of dispersal, speciation rates, local abundances, and meta-community size they provide predictions (Fig. 1) that are close to those of either the skewed lognormal or log-series (Hubbell, 2001). These models do not predict power function SADs (Fig. 1), commonly observed in arthropods (Siemann et al., 1999; Borda-de-Água et al., 2017). In addition, neutral models are notoriously difficult to fit and certain models require information on the structure of the underlying meta-community (Rosindell et al., 2010). Maximum likelihood fitting of neutral models requires complex assumptions about macroevolutionary processes (Etienne et al., 2007) that can result in unfeasible (or unmeasurable) parameter values, such as speciation rates (Ricklefs, 2003) and long-distance dispersal (Rosindell and Cornell, 2009). These issues have limited the application of neutral models for the prediction of species abundances.

Here, we strongly argue that a flexible descriptive SAD model needs to predict the abundances of each species directly and that abundance – rank orders are superior to distribution approaches. Only such species-focused models make it possible to trace deviations from observation and possible ecological drivers directly to the species level. There is also clearly a need to link a SAD model to species functional traits or phylogenetic relationships, for instance to infer how traits influence the dominance order of species abundances (Jones et al., 2017). Consequently, a flexible SAD model should ideally be based on Whittaker plot data.

Stauffer (1979) was apparently the first to propose the Weibull distribution (Weibull, 1951) as a model to explain observed species abundance distributions in forest trees. However, his derivation of the model and parameter interpretation were based on the broken stick model (Mac Arthur, 1957), a model that was subsequently found to be unrealistic. As his approach was distribution based the work did not receive the attention it deserves. Recently, Storch et al. (2018) used the one-parameter Weibull distribution to mimic the log-series. The R package SADs (Prado et al., 2017) provides functionality to fit the Weibull model to the statistical distribution, but not to rank abundances.

Here, we reintroduce the Weibull distribution as a flexible descriptive model that meets the above defined requirements. We show that this distribution is able to fit the most commonly observed SAD shapes. This ability enables us to compare the respective parameter values across SADs from different studies within the same modeling framework. Although Weibull distributions result from several stochastic processes (Rinne, 2008), we do not claim that the model is directly linked to ecological processes. We also demonstrate that the