Received: 27 February 2018 | Revised: 19 November 2018 | Accepted: 25 November 2018

DOI: 10.1111/geb.12879

RESEARCH PAPER

WILEY Global Ecology and Biogeography A Journal of Macroecology

Systematic variation in North American tree species abundance distributions along macroecological climatic gradients

Thomas J. Matthews[1,2,3] | Jon P. Sadler[1,2] | Yasuhiro Kubota[4] | Christopher W. Woodall[5] | Thomas A. M. Pugh[1,2]

[1]GEES (School of Geography, Earth and Environmental Sciences), The University of Birmingham, Birmingham, UK

[2]The Birmingham Institute of Forest Research, The University of Birmingham, Birmingham, UK

[3]CE3C – Centre for Ecology, Evolution and Environmental Changes/Azorean Biodiversity Group and Universidade dos Açores – Depto de Ciências e Engenharia do Ambiente, Angra do Heroísmo, Açores, Portugal

[4]Faculty of Science, University of the Ryukyus, Nishihara, Japan

[5]U.S. Department of Agriculture, Northern Research Station, Durham, New Hampshire

Correspondence
Thomas J. Matthews, School of Geography, Earth and Environmental Sciences, University of Birmingham, Birmingham B15 2TT, UK.
Email: t.j.matthews@bham.ac.uk

Editor: Nick Isaac

Abstract

Aim: The species abundance distribution (SAD) is a fundamental pattern in macroecology. Understanding how SADs vary spatially, and identifying the variables that drive any change, is important from a theoretical perspective because it enables greater understanding of the factors that underpin the relative abundance of species. However, precise knowledge on how the form of SADs varies across large (continental) scales is limited. Here, we use the shape parameter of the gambin distribution to assess how meta-community-scale SAD shape varies spatially as a function of various climatic variables and dataset characteristics.

Location: Eastern North America (ENA).

Time period: Present day.

Major taxa studied: Trees.

Methods: Using an extensive continental-scale dataset of 863,930 individual trees in plots across ENA sampled using a standardized method, we use a spatial regression framework to examine the effect of temperature and precipitation on the form of the SAD. We also assess whether the prevalence of multimodality in the SAD varies spatially across ENA as a function of temperature and precipitation, in addition to other sample characteristics.

Results: We found that temperature, precipitation and species richness can explain two-thirds of the variation in tree SAD form across ENA. Temperature had the largest effect on SAD shape, and it was found that increasing temperature resulted in more logseries-like SAD shapes (i.e. SADs with a relatively higher proportion of rarer species). We also found spatial variation in SAD multimodality as a function of temperature and species richness.

Main conclusions: Our results indicate that temperature is a key environmental driver governing the form of ENA tree meta-community-scale SADs. This finding has implications for our understanding of local-scale variation in tree abundance and suggests that niche factors and environmental filtering are important in the structuring of ENA tree communities at larger scales.

KEYWORDS

climate, compound distribution, gambin, macroecology, sampling effects, species abundance distributions

1 | INTRODUCTION

The species abundance distribution (SAD) describes how the number of individuals is distributed across all species in a sample or community and is one of the fundamental patterns in macroecology (Gaston & Blackburn, 2000; May, 1975; McGill, 2011; McGill et al., 2007). Although a multitude of different SAD models have been proposed (see McGill et al., 2007), SADs can be grouped into two main classes: logseries- and lognormal-like-shaped distributions (Ulrich, Kusumoto, Shiono, & Kubota, 2016; Ulrich, Ollik, & Ugland, 2010). The logseries distribution itself results from the Poisson sampling of a γ-distribution after a certain relevant limit is taken, and it is characterized by a right-hand-skewed curve with a modal value of one (Fisher, Corbet, & Williams, 1943). The lognormal distribution represents a situation in which the logarithms of abundances follow a Gaussian distribution, and it is characterized by a community in which species of intermediate abundance are most prevalent (May, 1975; Preston, 1948). Both the logseries and lognormal distributions are unimodal models, which have been the focus of many studies until recently (but see Ugland & Gray, 1982). However, recent work has indicated that a small proportion of empirical SADs are, in fact, multimodal (e.g. Antão, Connolly, Magurran, Soares, & Dornelas, 2017; Dornelas & Connolly, 2008; Matthews & Whittaker, 2015; Vergnon, van Nes, & Scheffer, 2012). For example, a recent synthesis of 117 datasets found significant evidence of multimodality in *c*. 20% of cases (Antão et al., 2017). A number of potential causes of multimodality in SADs have been put forward, such as the amalgamation of different types of species within a single sample (e.g. core and satellite species; Magurran & Henderson, 2003; Matthews & Whittaker, 2015) and the increasing taxonomic breadth, sampling variation and spatial extent (i.e. increasing ecological heterogeneity; Antão et al., 2017) of a study. However, variation in the prevalence of SAD multimodality at large scales and across ecological gradients is largely unknown.

Although a large proportion of previous (unimodal) SAD studies has focused either on finding the best-fitting model given a set of local-scale ecological data (e.g. Ulrich et al., 2010) or on using the SAD to test the performance of a particular theory or model (e.g. Volkov, Banavar, Hubbell, & Maritan, 2003), there has been increasing recognition of the importance of assessing how different SAD properties change across ecological gradients, such as climate, succession and disturbance gradients (e.g. Dornelas, Soykan, & Ugland, 2011; Matthews et al., 2014; Matthews, Borges, de Azevedo, & Whittaker, 2017; Ulrich et al., 2016). Traditionally, these SAD gradient studies have mostly been undertaken at relatively local scales (e.g. Bazzaz, 1975; Matthews et al., 2014). However, probably owing to the increased availability of open-source SAD datasets, in combination with an increase in computer processing power, there has been an increase in the number of SAD studies focusing on larger, macroecological, scales (e.g. Kubota, Kusumoto, Shiono, Ulrich, & Jabot, 2015; Ulrich et al., 2010, 2016; White, Thibault, & Xiao, 2012). Generally speaking, macroecological-scale analyses are characterized by a trade-off between global coverage and local/regional resolution; that is, studies that analyse datasets from across multiple continents (e.g. Ulrich et al., 2016) tend not to have very high coverage in any particular region/continent, and vice versa. Thus, most global-scale SAD analyses have large gaps within any given region. Although this is not a criticism of global SAD analyses, which are able to identify broad-scale patterns, it often involves analysing datasets from multiple studies that use different sampling methods and have varying aims, which may result in some patterns of interest being obscured. A different and more effective approach involves extensively sampling one large region using a standardized sampling protocol. This approach is arguably better at identifying spatial variation in SAD properties because it allows for more variables to be controlled, but, owing to the resources required to undertake the standardized sampling, it has been used less frequently in SAD studies (but see Locey & White, 2013; White et al., 2012).

Understanding how SADs vary spatially is important from a theoretical perspective because it enables greater understanding of: (a) what underpins the relative abundance of species (MacArthur, 1960, 1972; Matthews et al., 2017; May, 1975); and (b) large-scale species richness gradients (Currie et al., 2004; Rosenzweig, 1995). However, precise knowledge on how the form of SADs varies across large scales, and the role of different processes driving this change, is limited (Ulrich et al., 2016). It can be theorized that variation in climate across space will be important in driving variation in SAD form. Climatic variables are known to be important drivers of species richness gradients at macroecological scales (Brown, Gillooly, Allen, Savage, & West, 2004; Currie & Paquin, 1987; Field et al., 2009). In particular, temperature and precipitation are known to be the primary limiting drivers of richness variation in North American trees (Allen, Brown, & Gillooly, 2002; Currie & Paquin, 1987; Whittaker, Willis, & Field, 2003). The effect of variation in climate on the shape of the SAD is largely unknown, but based on the findings of previous studies on species richness gradients (e.g. Currie & Paquin, 1987; Currie et al., 2004), we predict logseries SAD shapes to be more prevalent with increasing temperature and precipitation. Primary productivity correlates strongly with climatic variables, and higher energy and productivity is known to: (a) result in finer-scale divisions of niche space (Whittaker et al., 2003); and (b) enable areas to support more individuals (Currie et al., 2004) at smaller minimum viable population sizes (Hawkins et al., 2003). In addition, it has been argued that, in contrast to the theoretical predictions of the species–energy hypothesis, populations of species are smaller in more productive environments (Currie et al., 2004). These factors have been postulated to result in greater richness in productive environments, but together would also mean a higher proportion of rarer species and thus logseries SAD shapes. The role of climatic variables in driving multimodality in SADs has not, to our knowledge, been assessed previously at this scale.

Macroecological SAD studies have tended to compare the fit of different models, and then assessed spatial variation in the best-fitting model (e.g. Ulrich et al., 2016). However, this approach does not necessarily provide accurate information about the shape of the

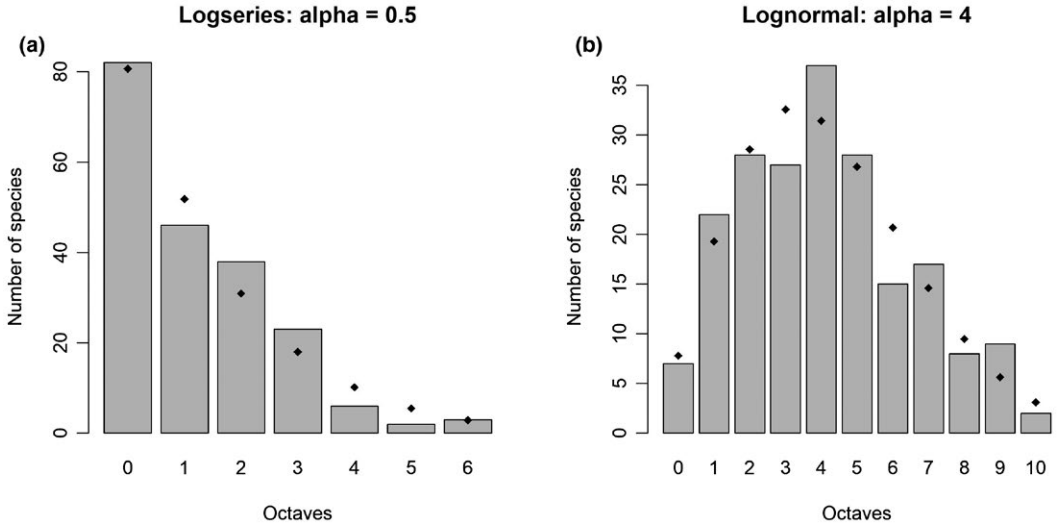

FIGURE 1 The two most commonly observed species abundance distribution (SAD) shapes: (a) logseries-like distributions, and (b) lognormal-like distributions. In both (a) and (b), the gambin model (black circles) has been fitted to the data binned into logarithmic octaves (grey bars). The data were simulated by sampling random values from gambin distributions with α parameters of 0.5 (a) and 4 (b); in both plots, the number of species was set to 200

SAD, because certain SAD models are relatively flexible and can fit a range of SAD forms, and choosing a best-fitting model does not necessarily mean that it fits the data well (i.e. none of the models in the comparison might provide an accurate representation of the SAD shape). An alternative approach focuses on a single value that characterizes the shape of the SAD (Ulrich, Nakadai, Matthews, & Kubota, 2018), such as the shape parameter of the gambin model (Ugland et al., 2007). Gambin is a stochastic model, which combines the γ-distribution with a binomial sampling method. The unimodal gambin model has a single free parameter (α), which characterizes the distribution shape: low values indicate logseries-shaped curves, whereas higher values indicate more lognormal-shaped curves (for an example, see Figure 1). Gambin has been shown to provide good fits to a wide variety of empirical datasets, and α has been found to represent a useful metric that can be used to assess the effect of different variables on SAD shapes (Arellano et al., 2017; Dornelas et al., 2011; Matthews et al., 2014). Recent methodological developments (Matthews et al., 2018) have derived the likelihood functions for multimodal gambin models (multimodality also being a measure of the shape of the SAD), thus providing a means of easily assessing multimodality in SAD datasets.

Here, we analyse an extensive dataset of 863,930 individual trees in 33,282 plots across Eastern North America (ENA), sampled using a standardized method, to examine the effect of climate on the form of ENA tree SADs across broad spatial scales. We combined adjacent plots (within grid squares of *c*. 44 km × 44 km) to create coarse-scale SADs; thus, we are analysing SADs at the metacommunity scale. We use gambin's α to assess how SAD shape varies spatially as a function of various climatic variables and dataset characteristics. We hypothesized that, owing to the arguments outlined above, we would observe a shift from lognormal- to logseries-shaped SADs with increasing temperature and precipitation. We also assessed the prevalence of SAD multimodality and whether SAD multimodality varies spatially as a function of temperature and precipitation.

2 | MATERIALS AND METHODS

2.1 | Data and sampling methodology

Our analyses were based on publicly available plot-level data produced by the U.S. Department of Agriculture, Forest Service's Forest Inventory and Analysis Program (FIA; http://fia.fs.fed.us/). The FIA Program conducts a systematic and consistent inventory of all forest land in the USA, with a comprehensive summary of the associated data and sampling methodology provided by O'Connell et al. (2017). Briefly, inventory plots are systematically distributed across the entire USA, with remotely sensed information used to identify plots that are located in a forest land use. Each FIA plot comprises four circular subplots of area 0.017 ha, each located within a circular 0.10 ha macroplot. All free-standing woody stems (live and dead) with a diameter ≥ 12.7 cm are sampled within each subplot. Within each subplot, there is a circular 0.001 ha microplot in which all live stems with a diameter ≥ 2.54 cm are sampled (O'Connell et al., 2017). The full dataset contained sampling data from two time periods; that is, each plot was re-sampled a second time on average 5 years later. For the present study, we used only the data from the first sampling period. The location of the plots is illustrated in Figure 2.

Annual mean temperature and annual mean precipitation data for each plot were sourced from the WorldClim database (version 2.0; 2.5 min resolution; Fick & Hijmans, 2017) using averages based on annual means (1970–2000) and extracted using the "raster" R package (Hijmans, 2017; for further details, see Appendix S1). Climatic

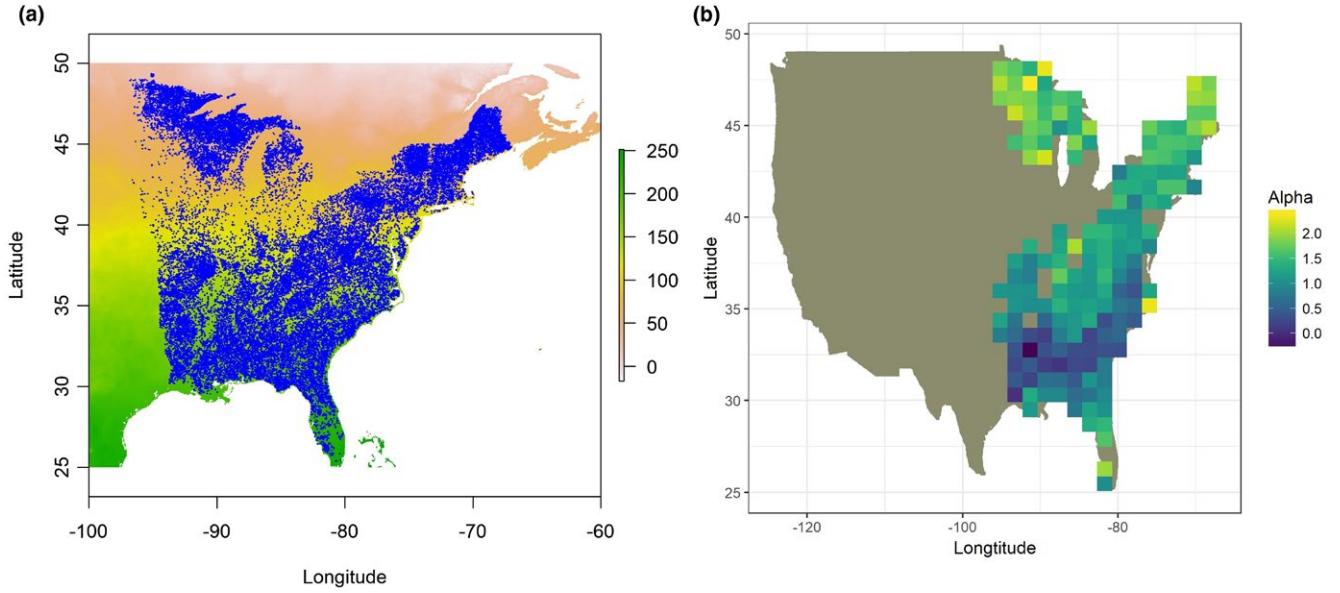

FIGURE 2 (a) The distribution of 33,282 Forest Inventory and Analysis (FIA) Program tree plots (blue dots) across Eastern North America, overlaid on a heat map of temperature values. Temperature data were sourced from the WorldClim database and represent the annual mean temperature at 2.5 min resolution (temperature data are in the form °C × 10). (b) A heat map showing spatial variation in the α parameter ($\log_e$-transformed) of the gambin model; α values were grouped into 30 bins and the median value displayed. Lower values of α correspond to logseries-like SAD shapes, whereas higher values correspond to more lognormal-like SAD shapes (see Figure 1). The α values were generated from fitting the gambin SAD model to tree data from 737 coarse-scale samples, where a sample is a collection of FIA tree monitoring plots within a *c*. 44 km × 44 km grid [Colour figure can be viewed at wileyonlinelibrary.com]

seasonality variables were also extracted but not used further owing to multicollinearity issues (based on variance inflation factors).

2.2 | Dataset format and fitting the gambin model

Although the full dataset had a very high spatial resolution (i.e. coverage of plots within ENA), the individual plots did not contain sufficient individuals to fit SAD models confidently (see McGill, 2011). Thus, we pooled all plots within a given distance to create meta-community-scale SADs. To achieve this, we divided the ENA into a grid of squares of $x_1 \times x_2$. For each grid square, we then pooled all plots with centre points within the boundaries, thus creating individual meta-community samples of individuals (hereafter, "samples") for each grid square. For the main analyses, we used 0.4° of latitude (i.e. $x_1 = 0.4°$). Given that the length of a degree of longitude varies with latitude, we varied the selected degrees of longitude (x_2) at different latitudinal bands to ensure that the grid squares were all approximately the same size (*c*. 44 km × 44 km).

Given that SAD model parameters are known to be biased when sample size is small (Matthews et al., 2014; McGill, 2011), we removed all samples with < 500 individuals. For the remaining samples, we fitted the one-component (unimodal) gambin model to each sample using the gambin R package (version 2.4; Matthews et al., 2014). Given that SAD model parameters are sensitive to variations in sample size (McGill, 2011), we used a procedure where, for each sample, we subsampled 500 individuals, fitted the unimodal gambin model to this subsample and stored the α parameter value. Given that this subsampling procedure is stochastic, we repeated the process 100 times for each sample and took the mean α value. The 100 subsamples were also used to create estimates of the *SE* of the mean value. Given that comparing SAD model parameters makes sense only if the model provides a reasonable fit to the data, for each model we fitted, we also stored the χ^2 goodness-of-fit statistic and its associated *p*-value; the mean values of the 100 subsamples were then calculated. We then discarded all samples where the mean *p*-value was < .05. Occasionally, the model did not converge fully, and the fit generated unrealistically high values of α (e.g. 100). Following previous work and earlier versions of the gambin package (Matthews et al., 2014), we discarded all samples that had a mean α value > 15. Species identities were taken from O'Connell et al. (2017; see their appendix F).

2.3 | Spatial regression analyses

To determine whether temperature and precipitation could explain any of the variation in SAD form across our samples, we used a spatial linear regression modelling approach. For the response variable, we used the α values from the fits of the unimodal gambin model to the samples. The distribution of α values was skewed; therefore, it was $\log_e$-transformed to enable use of standard Gaussian linear models. We included three predictor variables (temperature, precipitation and species richness of the sample), and all predictor variables were standardized to have a mean of zero and a *SD* of one to enable comparison of effect sizes. Species richness was simply the number of species in a sample, and precipitation and temperature were taken as the mean values of the plots within a sample. The

variance inflation factors of all predictors were below three. The number of individuals was not included as a predictor because this was standardized before fitting the gambin model. We then fitted a standard linear model using all predictors. We tested for spatial autocorrelation in the residuals of this model fit using a permutation test (999 permutations) for Moran's *I* statistic and the "spdep" R package (Bivand, 2017); spatial weights for neighbour lists were calculated using the "knearneigh" ($k = 4$) and "nb2listw" functions and row standardized weights. This test revealed strong spatial autocorrelation in the residuals (Moran's $I = .38$, $p = .001$). To account for this, we used a spatial regression framework (Bivand, 2017; Ward & Gleditsch, 2008). We fitted both a spatial lag model (i.e. a spatial simultaneous autoregressive lag model) and a spatial error model (i.e. a spatial simultaneous autoregressive error model) using the "spdep" R package and compared the models using Akaike's information criterion (AIC). Owing to the large number of data points, it was not necessary to use a corrected AIC. Using the best-fitting spatial regression model, we fitted the global model (i.e. with all predictors) and models with all possible predictor combinations that included species richness; species richness was included in all models because it was a predictor we wanted to control for. We also fitted a null model (only an intercept term; Mac Nally, Duncan, Thomson, & Yen, 2018). Model comparisons used an information theoretic approach (Burnham & Anderson, 2002). All global and best-fitting models were rechecked for residual spatial autocorrelation, and we also calculated Nagelkerke's pseudo-R^2 to assess model fit. To validate models, we extracted the fitted values and the residuals from the spatial regression model fit object, and then constructed Q–Q plots to check for residual normality and plotted the fitted values against the residuals to check for homoscedasticity. We constructed partial regression plots to assess the effect of each variable after taking into account the effect of the other predictors.

2.4 | Assessing multimodality using multiple-component gambin models

To assess whether the prevalence of multimodality in the SAD varied as a function of the predictor variables, for each sample (from the 0.4° of latitude grid) we fitted both one-component and two-component gambin models using the gambin R package (version 2.4; Matthews et al., 2018) and derived the Bayesian information criterion (BIC) values. We used BIC here rather than AIC because the former is known to penalize more complex models more strictly than AIC (Burnham & Anderson, 2004), and this is a desirable property in this context, because arguably, a test of multimodality should be conservative. The two-component model was considered the best-fit model if it had a ΔBIC value lower than the one-component model (Burnham & Anderson, 2002, 2004). Given that we were not interested in comparing parameter values across samples in this part of the analysis, we fitted the models without standardizing for the number of individuals in the samples. We excluded all samples where neither the one-component or the two-component model had a χ^2 p-value > .05. We converted the number of times the two-component model was the best-fitting model into a binomial variable to be used as a response variable in a binomial generalized linear model (GLM), using temperature and precipitation as predictor variables. Given that we did not standardize by sample size, we also included the number of individuals and the number of species in a sample as predictors. All predictors were standardized to have a mean of zero and a *SD* of one. The variance inflation factors of all parameters were below five. To deal with spatial autocorrelation, we created an autocovariate to be used in autologistic regression, using the "autocov_dist" function (type = "inverse"; style = "W") in the spdep R package. We set the neighbourhood radius to 50 km to ensure that there were few regions that included points with zero links to other points (.04% of regions; average number of links = 2.61). We used the MuMIn R package (Bartoń, 2012) to fit a complete set of models, considering all predictors; the autocovariate was set as fixed. As with the unimodal model selection analysis, we used AIC to compare regression models. Weight of evidence (WoE) values for each predictor variable were calculated by summing the AIC weights for all models in which a variable was present (Burnham & Anderson, 2002; Giam & Olden, 2016). McFadden's pseudo-R^2 was calculated for each model using the formula: 1 − (residual deviance/null deviance).

2.5 | Sensitivity analyses

In order to determine whether our results were influenced by the location and size of the grid squares, by the pooling of data within grid squares in general, and to test the effect of potentially including managed plots on our results, we ran a number of sensitivity analyses to account for these factors (the full methods are provided in Supporting Information Appendix S1).

Inspection of plots of the *SEM* α values (the mean of the α values from the 100 subsamples) indicated that the *SE* increased with increasing mean α (Supporting Information Appendix S2, Figure S1). Thus, to ensure that this did not bias our results: (a) we re-ran the main analyses using unstandardized α values (i.e. the gambin model was fitted without any subsampling); and (b) we used the *SE*s as weights in a linear regression model selection. The *SE*s were normalized between zero and one, and we used the inverse of these values as the weights. It was necessary to use a standard linear model because it was not possible to add weights to our spatial regression models. Re-running the analyses using unstandardized α values also enabled us to check that our random subsampling did not affect the results. For example, this could happen because our subsampling procedure randomly sampled individuals from a plot, which disregards the possibility that individuals of a species are spatially aggregated rather than randomly distributed within a plot.

In addition, when the number of species in a sample is low, the shape of the SAD is constrained (Locey & White, 2013). Thus, to test whether our results are simply an artefact of low richness in certain samples we ran an additional simulation analysis, in which we assessed the effect of the number of species in a sample on the

α value of the unimodal gambin model (full details can be found in Supporting Information Appendix S1).

All analyses were undertaken in R (version 3.5.1; R Core Team, 2017). The R code is provided in an online repository on GitHub (txm676/NEA_SADs).

3 | RESULTS

Across the 33,282 plots there were 863,930 individual trees, representing 214 species. Using a *c.* 44 km × 44 km grid square and a minimum number of individuals threshold of 500, there were 763 coarse-scale SAD samples. After the removal of samples to which the one-component gambin model did not provide a good fit (according to the χ^2 statistic or an unreasonably high α value), we were left with 737 samples distributed across ENA. The mean species richness of the samples was 30 (*SD* = 8), and the mean number of individuals was 904 (*SD* = 343, although the number of individuals in each sample was standardized before model fitting).

3.1 | Variation in gambin's α along macroecological gradients

When all predictors were considered, the spatial error model (AIC = 644.8) had a lower AIC value than the spatial lag model (AIC = 647.8) and the standard non-spatial linear model (AIC = 822.8), and the residuals were no longer autocorrelated (Moran's *I* of the global model = −.03, *p* = .91). Thus, the spatial error model was used in subsequent analyses.

There was substantial spatial variation in α (see Figure 2). The best spatial error model (i.e. the model with the lowest AIC) contained temperature and species richness (Table 1); temperature had the largest effect on α (i.e. this variable had the largest coefficient), followed by species richness. A second model containing all three predictors was also within two ΔAIC values of the best model (Table 1). The model with the lowest AIC value explained a large amount of the variance in SAD form (pseudo-R^2 = .65). The ΔAIC value of the null model was 132.5 (Table 1), meaning that the best model provided a substantially better fit than an

TABLE 1 Spatial error model selection results

Rank	Temperature	Precipitation	Species richness	AIC	ΔAIC	*Z*-value
1	−0.39 (0.03)	–	−0.12 (0.02)	643.4	0.0	14.64*
2	−0.36 (0.04)	−0.04 (0.04)	−0.12 (0.02)	644.8	1.4	14.85*
3	–	−0.27 (0.04)	−0.14 (0.02)	694.6	51.2	22.80*
4	–	–	−0.15 (0.02)	730.0	86.6	29.57*
5	–	–	–	775.8	132.5	32.96*

Note. The response variable in all models was the α parameter of the gambin species abundance distribution model ($\log_e$-transformed). The three predictors (temperature, precipitation and species richness) were standardized to have a mean of zero and *SD* of one. The data were pooled samples (*n* = 737) of North American tree monitoring plots. The abundance data of each sample were standardized to ensure that all samples contained 500 individuals. For each model, the variable coefficient estimate is given with the *SE* in parentheses. The Akaike information criterion (AIC), ΔAIC and the *Z*-value of each model are also provided.

*Significant *Z*-value at the $p \leq .001$ level.

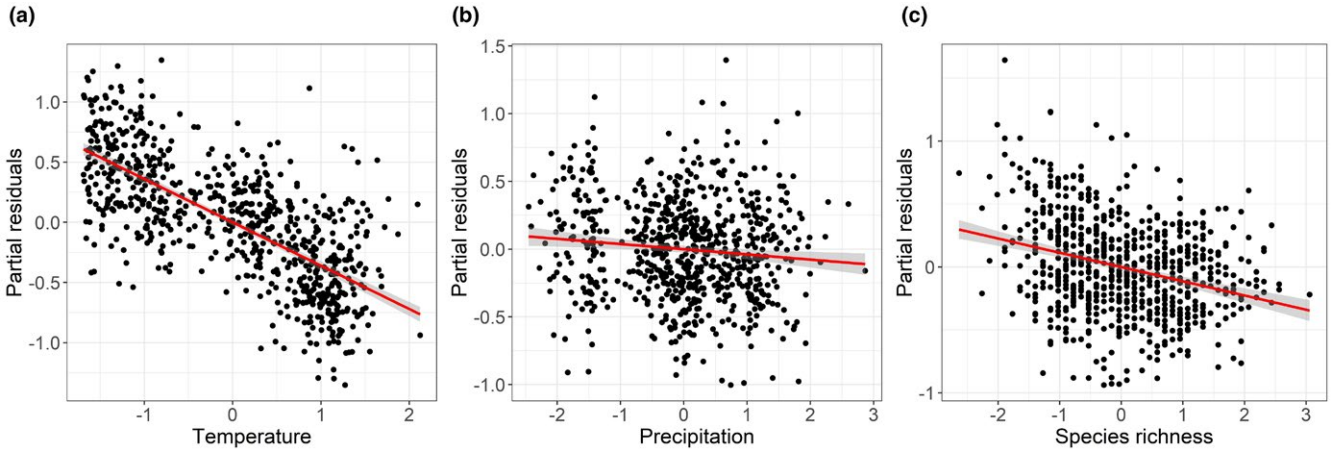

FIGURE 3 Partial residual plots showing the effect of temperature (a), precipitation (b) and species richness (c) on the α parameter of the gambin SAD model ($\log_e$-transformed), after taking into account the effect of the other independent variables in the global model (i.e. the model containing all three predictors). The red continuous line in each plot represents the best-fitting linear model to the partial residual data. The α values were generated from fitting the gambin SAD model to tree data from 737 coarse-scale samples, where a sample is a collection of FIA tree monitoring plots within a *c.* 44 km × 44 km grid [Colour figure can be viewed at wileyonlinelibrary.com]

intercept-only model. The partial regression plots (Figure 3) of the global model (i.e. the model with all three predictors) showed a stronger negative effect of temperature on α (Figure 3a) relative to the effects of species richness (Figure 3c) and precipitation (Figure 3b). The residuals of the best spatial error model were not significantly autocorrelated (Moran's I = −.03, p = .91) and were observed to be normally distributed according to both the Q–Q plot and the histogram (Shapiro–Wilk normality test on the residuals: w = 1.00, p = .15).

3.2 | Multimodal model results

After filtering out samples according to our acceptance criteria (i.e. number of individuals and χ^2 goodness-of-fit test), we were left with 653 samples. Across these, the two-component (bimodal) gambin model had the lowest BIC value in 65 cases (10%). Examples of unimodal and bimodal SADs are provided in Figure 4a,b. A heat map of the bimodal BIC weights is provided as Supporting Information (Appendix S2, Figure S2). The best binomial GLM contained temperature, number of species and number of individuals, in addition to the spatial autocovariate (Table 2). The pseudo-R^2 of the best model was low (.09), and there was no residual spatial autocorrelation (Moran's I = .03, p = .10). There were two additional models with ΔAIC values less than two; temperature and species richness were included in both (Table 2). Temperature had a WoE value of one, whereas precipitation and the number of individuals and species had values of .34, .60 and .98, respectively (Table 2). The fit of a simple non-spatial logistic regression model using only temperature as a predictor is shown in Figure 4c; it shows an increasing probability of a one-component model providing the best fit with increasing temperature.

3.3 | Sensitivity analyses

Removing the potentially managed plots or re-running the analyses from different starting points to create the grid cells or with smaller grid squares did not affect the overall results; the spatial model selection results, partial regression plots and binomial GLM selection analyses produced largely similar outcomes (see Supporting Information Appendix S2). The main difference was the performance of the precipitation variable, which had a positive effect in some of the models with smaller grid squares (in both the unimodal and multimodal model selection analyses) and in two of the analyses with shifted grid squares (Supporting Information Appendix S2).

Re-running the spatial error model selection using α values generated from fitting gambin to the raw plot data also did not affect the overall results. Regardless of the threshold used for the number of individuals (i.e. 25 or 50; which resulted in 9,669 and 633 plots, respectively), the model selection results were very similar (Supporting Information Appendix S2, Tables S6 and S7). Again, the main difference was the performance of the precipitation variable, which was included in the best model and had a positive effect, in both cases. Re-running the spatial error model selection using unstandardized α values did not change the overall results (Supporting Information Appendix S2, Table S8 and Figure S5). In addition, using the *SE* (of mean α) values as weights in a standard linear model resulted in a very similar global model (Supporting Information Appendix S2, Table S9).

The species richness simulations indicated that the α parameter of a sample is an accurate estimate of the population α value when the number of species is > 10 (see Supporting Information Appendix S2, Figure S6). When the number of species is < 10, the α value tends

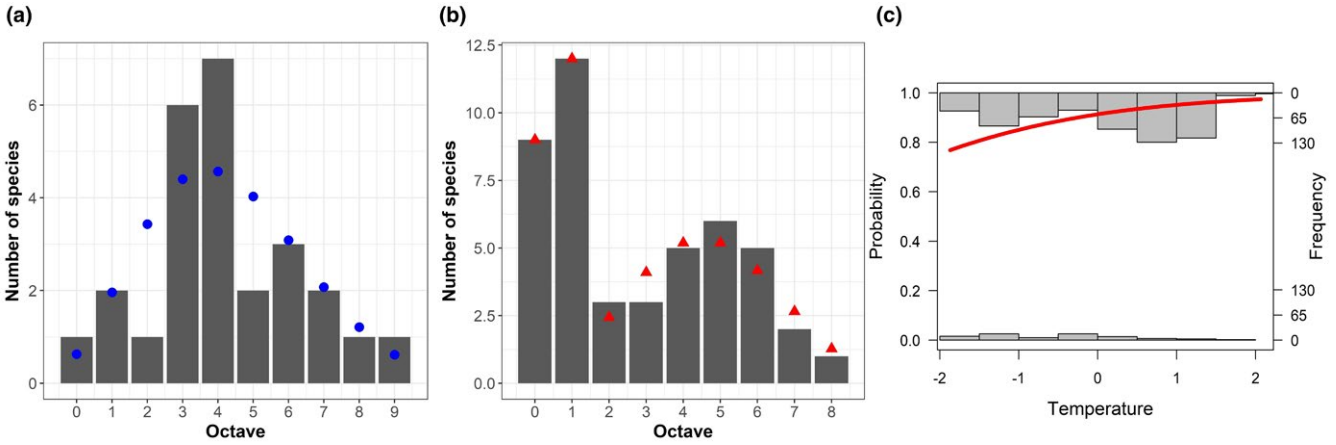

FIGURE 4 An example of a unimodal (a) and bimodal (b) SAD (grey bars), generated using samples of North East American trees, where a sample is a collection of FIA tree plots. In (a), the fit of the one-component gambin model (blue circles), and in (b), the fit of the two-component gambin model (red triangles), are shown. The sample in (a) comprises 26 species and 1,902 individuals, and in (b) 46 species and 1,456 individuals. In (c), the results of a logistic regression are displayed; the red line represents the curve of the predicted values from the model, and the grey bars are the observed data points displayed as a histogram. The predictor variable in the model was (standardized) temperature. The response variable in the model was a binary variable indicating whether a two-component gambin model provided a better fit than a one-component gambin model to a given sample; we used one minus this value for illustrative purposes, and thus the curve shows an increasing probability of a one-component model providing the best fit with increasing temperature. The data are 653 meta-community-scale samples of Eastern North American trees [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 2 Binomial generalized linear model selection results

Rank	AutCov	*n*	Precipitation	Species richness	Temperature	AIC	ΔAIC
1	−0.21 (0.12)	−0.22 (0.14)	–	0.53 (0.17)	−1.05 (0.23)	395.52	0.00
2	−0.22 (0.12)	–	–	0.51 (0.17)	−0.91 (0.20)	396.21	0.69
3	−0.19 (0.12)	−0.24 (0.14)	0.19 (0.21)	0.50 (0.17)	−1.21 (0.29)	396.74	1.22
WoE	1.00	.60	.34	.98	1.00		

Note. Models were compared using the Akaike information criterion (AIC). All models with a ΔAIC ≤ 2 are shown. The response variable in all models was a binary variable describing whether a two-component gambin model provided a better fit than a one-component gambin model. The five predictors were as follows: temperature, precipitation, number of individuals in a sample (*n*), a spatial autocovariate (AutCov) and species richness. All predictors were standardized to have a mean of zero and *SD* of one. The data were pooled samples of North American tree monitoring plots (*n* = 653). For each model, the variable coefficient estimate is given with the *SE* in parentheses. The AIC and ΔAIC of each model are also provided.

to be inflated. Given that only one of our samples in the main analysis had fewer species than 10 (nine), we are confident that our results are not an artefact of low richness in our samples.

4 | DISCUSSION

The majority of SAD analyses are undertaken at local scales, and thus less is known about SADs at larger meta-community scales. In addition, the use of standardized plot data avoids the biases introduced in many macroecological studies, whereby data from multiple studies that use different sampling methods are combined. We found that our global model (the two climatic variables and species richness) explained around two-thirds of the variation in meta-community SAD shape across ENA. This is a significant amount; a recent global synthetic analysis (i.e. combining multiple independent studies) of SADs was able to explain *c*. 20% of variation in SAD form (Ulrich et al., 2016). We also found some spatial variation in SAD multimodality as a function of temperature, species richness and, to a lesser extent, the number of individuals.

Model selection indicated that temperature was the most important variable driving variation in α; it had the largest effect in all ΔAIC < 2 models (Table 1). Our findings show that, in line with our prediction, the effect of temperature was negative (see Figure 3); increasing temperature resulted in lower α and thus more logseries-like SADs. This finding suggests that temperature is more important than water in structuring ENA tree SADs. Allen et al. (2002) reported a similar finding for the species richness of North American trees. However, although precipitation had a smaller effect than temperature, it was still retained in a model with ΔAIC < 2. It is likely that at some extremes, precipitation has a greater effect on SAD form (e.g. highly arid areas), and that our dataset did not cover enough of these extreme environments (see also Hawkins et al., 2003). It is also possible that other variables (e.g. potential evapotranspiration; Currie et al., 2004) might be more accurate measures of water availability and water deficit than mean precipitation (Anderegg et al., 2015). A more in-depth analysis of the role of different productivity metrics on SAD form would be an interesting future step, although such an analysis is reliant on the availability of suitable data at large scales. It should be noted that in some of the sensitivity tests the effect of precipitation on α was positive (rather than negative as in the main analysis; compare, for example, Figure 3 with Supporting Information Appendix S2, Figure S4). However, the effect was often close to zero, and the *SE* of the effect often overlapped zero (e.g. see Supporting Information Appendix S2, Table S3). In addition, within the same model selection analysis (e.g. Supporting Information Appendix S2, Tables S3 and S7), the effect of precipitation on α was found to switch between positive and negative for different models. Thus, the effect of precipitation on α, based on these data at least, can be considered negligible.

The strong role of temperature and thus energy availability implies that niche processes (e.g. niche division) leave an imprint on the SAD. Previous climate-richness gradient studies have shown that diversity is positively correlated with productivity, owing in part to the fact that more individuals can be supported in productive ecosystems, and the minimum viable population sizes of individual species are often smaller; thus, a larger number of species, with smaller population sizes, can be supported in a given unit of area than in less productive systems (Allen et al., 2002; Brown et al., 2004; Hawkins et al., 2003). For example, the average population densities and population sizes of tree species have both been shown to decrease with increasing temperature (Allen et al., 2002) and decreasing latitude (Currie et al., 2004). A separate but linked idea is the theory that greater niche division in more productive environments enables more species to coexist in a given area (Rosenzweig, 1995; Whittaker et al., 2003). If abundance is linked to niche size (MacArthur, 1972), then greater niche division would result in a higher proportion of relatively rare species being found in more productive environments, which would also explain our findings. The negative effect of species richness on α, although less than the effect of temperature (Table 1), also fits into this rationale; more productive environments, in general, support larger numbers of species (Hawkins et al., 2003). A novel finding of our study is that abundance is distributed across these larger numbers of species in a less even way than in cooler environments. These observations could also be attributable to the filtering out of rarer species (i.e. species with small populations) in colder regions, which in turn could be linked to tropical niche conservatism (Wiens et al., 2010); fewer lineages are adapted to colder temperatures, and thus there is less competition, and a larger proportion of species is able to have higher relative abundance. Another possibility

is that increasing temperature results in a higher rate of speciation (Allen, Gillooly, Savage, & Brown, 2006), leading to a larger number of young species with relatively small population sizes and thus more logseries-like SADs. These different explanations are not necessarily mutually exclusive.

An alternative explanation is the possible constraining influence of community richness on the shape of the SAD (Locey & White, 2013). A meta-analysis by White et al. (2012), who included in their analysis a subset of the plots we have analysed, has shown that the maximum entropy theory of ecology (see Harte, 2011) can successfully capture a large proportion of the variation in SADs. In addition, Locey and White (2013) found that the form of SADs in many cases is not different from the central tendency of the feasible set of possible distributions. However, our results do not contradict those of White et al. (2012) or Locey and White (2013). Both these studies found that the application of a maximum entropy model and the feasible set framework to empirical data does not capture all of the variation in SAD form, particularly in cases where the SAD is exceptionally even or uneven (see Locey & White, 2013). More importantly, as the authors state, the results of these studies do not imply that ecological processes are unimportant. Instead, ecological processes are likely to be important indirectly, through their impacts on the state variables and constraints considered (e.g. the total number of individuals or the number of species; see Harte, 2011). Our results, alongside the many studies of large-scale richness gradients (e.g. Currie & Paquin, 1987; Field et al., 2009; Hawkins et al., 2003), indicate that temperature, in particular, is a key variable of interest in this regard.

Although the unimodal gambin model provided a better fit, according to BIC, in the majority of cases, the bimodal model was the best-fitting model to a small proportion of samples (10%). Multimodality is also a measure of the shape of a SAD, and our multimodal SAD model selection results (Table 2) provide further evidence in support of the role of temperature in driving variation in SAD shape. However, it should be noted that the amount of variation explained by the best model in this case (pseudo-R^2 = .09) was much lower than in the unimodal gambin α analyses (Table 2), and it is thus hard to draw any general conclusions on the variables driving this pattern. That being said, one interesting result is that the effect of temperature and species richness (in the models in which species richness was included; see Table 2) have opposite signs; temperature has a negative effect (see Figure 4c) and species richness a positive effect on the log odds prevalence of bimodality in the analysed SADs (Table 2). The reasons for contrasting effects of temperature and richness are unclear. One potential explanation might be that an additional variable not included in our analysis covaries with either or both temperature and species richness, such as topographical relief or human disturbance. Another potential explanation for the presence of multimodal SADs more generally is the possible presence of strong fine-scale climatic gradients within some of our coarse-scale grid squares; the fact that our sensitivity analysis using a smaller grid square size provided evidence for a relatively larger (compared with the larger grid square analysis) positive effect of precipitation on the prevalence of multimodality provides evidence supporting this explanation. Further work is needed to explore these possibilities and to identify additional covariates that explain more variation than those analysed here.

Forest management provides another potential confounding factor in our interpretation. Due to the systematic nature of the FIA inventory, some of the plots included in the study dataset were in managed forest (see O'Connell et al., 2017). The inclusion of these forests could potentially have biased our results if there is strong spatial variation in the location of managed forest plots in relationship to the other covariates. For example, there are currently known to be large tracts of managed forests (e.g. aspen and birch, and spruce and fir plantations) in northern USA. However, although management was not explicitly recorded during sampling, each of the plots (and individual trees) included in our analysis has been surveyed at multiple points in time, and when an individual tree was found to have died between time periods, a cause of death was inferred. This allowed us to remove all plots that contained several trees that were listed as having died as a result of management-related activities and to re-run the analyses (which did not affect the overall results; see Supporting Information Appendix S2). Although this is not a perfect metric of management and human disturbance, we are confident that our overall results are not simply an artefact attributable to the inclusion of managed forests.

In a previous study of the SADs of North American trees using a functional trait-based maximum entropy model, Xing, Swenson, Weiser, and Hao (2014) found that broad-scale SADs were important drivers of local-scale abundance, arguing that it was thus necessary to discern the underlying mechanisms of North American tree SADs across broader scales. The results of our study indicate that temperature is a (perhaps the) key environmental driver governing the form of ENA tree SADs at large meta-community scales, which should thus aid in our understanding of the local-scale variation in tree abundance. This, in turn, suggests that niche factors and environmental filtering are important in structuring ENA tree communities.

ACKNOWLEDGMENTS

Francois Rigal provided advice on spatial logistic regression. This is paper number 33 of the Birmingham Institute of Forest Research.

AUTHOR CONTRIBUTIONS

T.J.M. designed the study; C.W.W. and T.A.M.P. provided data; T.J.M. ran the analyses; T.J.M. wrote the manuscript with the help of J.P.S.; all authors commented on the manuscript.

DATA ACCESSIBILITY

Our analyses were based on publicly available plot-level data produced by the U.S. Department of Agriculture, Forest Service's Forest Inventory and Analysis Program (http://fia.fs.fed.us/).

ORCID

Thomas J. Matthews https://orcid.org/0000-0002-7624-244X

Yasuhiro Kubota https://orcid.org/0000-0002-5723-4962

REFERENCES

Allen, A. P., Brown, J. H., & Gillooly, J. F. (2002). Global biodiversity, biochemical kinetics, and the energetic-equivalence rule. *Science*, *297*, 1545–1548. https://doi.org/10.1126/science.1072380

Allen, A. P., Gillooly, J. F., Savage, V. M., & Brown, J. H. (2006). Kinetic effects of temperature on rates of genetic divergence and speciation. *Proceedings of the National Academy of Sciences of the USA*, *103*, 9130–9135. https://doi.org/10.1073/pnas.0603587103

Anderegg, W. R. L., Flint, A., Huang, C.-Y., Flint, L., Berry, J. A., Davis, F. W., … Field, C. B. (2015). Tree mortality predicted from drought-induced vascular damage. *Nature Geoscience*, *8*, 367. https://doi.org/10.1038/ngeo2400

Antão, L. H., Connolly, S. R., Magurran, A. E., Soares, A., & Dornelas, M. (2017). Prevalence of multimodal species abundance distributions is linked to spatial and taxonomic breadth. *Global Ecology and Biogeography*, *26*, 203–215. https://doi.org/10.1111/geb.12532

Arellano, G., Umaña, M. N., Macía, M. J., Loza, M. I., Fuentes, A., Cala, V., & Jørgensen, P. M. (2017). The role of niche overlap, environmental heterogeneity, landscape roughness and productivity in shaping species abundance distributions along the Amazon–Andes gradient. *Global Ecology and Biogeography*, *26*, 191–202.

Bartoń, K. (2012). *MuMIn: multi-model inference (R package version 1.2)*. Retrieved from https://cran.r-project.org/web/packages/MuMIn/index.html

Bazzaz, F. A. (1975). Plant species diversity in old-field successional ecosystems in Southern Illinois. *Ecology*, *56*, 485–488. https://doi.org/10.2307/1934981

Bivand, R. (2017). *spdep: spatial dependence: weighting schemes, statistics and models (R package version 0.6-13)*. Retrieved from https://cran.r-project.org/web/packages/spdep/index.html

Brown, J. H., Gillooly, J. F., Allen, A. P., Savage, V. M., & West, G. B. (2004). Toward a metabolic theory of ecology. *Ecology*, *85*, 1771–1789. https://doi.org/10.1890/03-9000

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multi-model inference: A practical information-theoretic approach* (2nd ed.). New-York, NY: Springer.

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research*, *33*, 261–304. https://doi.org/10.1177/0049124104268644

Currie, D. J., & Paquin, V. (1987). Large-scale biogeographical patterns of species richness of trees. *Nature*, *329*, 326–327. https://doi.org/10.1038/329326a0

Currie, D. J., Mittelbach, G. G., Cornell, H. V., Field, R., Guégan, J.-F., Hawkins, B. A., … Turner, J. R. G. (2004). Predictions and tests of climate-based hypotheses of broad-scale variation in taxonomic richness. *Ecology Letters*, *7*, 1121–1134. https://doi.org/10.1111/j.1461-0248.2004.00671.x

Dornelas, M., & Connolly, S. R. (2008). Multiple modes in a coral species abundance distribution. *Ecology Letters*, *11*, 1008–1016. https://doi.org/10.1111/j.1461-0248.2008.01208.x

Dornelas, M., Soykan, C. U., & Ugland, K. I. (2011). Biodiversity and disturbance. In A. E. Magurran & B. J. McGill (Eds.), *Biological diversity: Frontiers in measurement and assessment* (pp. 237–251). Oxford, UK: Oxford University Press.

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, *37*, 4302–4315. https://doi.org/10.1002/joc.5086

Field, R., Hawkins, B. A., Cornell, H. V., Currie, D. J., Diniz-Filho, J. A. F., Guégan, J.-F., … Turner, J. R. G. (2009). Spatial species-richness gradients across scales: A meta-analysis. *Journal of Biogeography*, *36*, 132–147. https://doi.org/10.1111/j.1365-2699.2008.01963.x

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. https://doi.org/10.2307/1411

Gaston, K. J., & Blackburn, T. M. (2000). *Patterns and process in macroecology*. Oxford, UK: Blackwell Science.

Giam, X., & Olden, J. D. (2016). Quantifying variable importance in a multimodel inference framework. *Methods in Ecology and Evolution*, *7*, 388–397. https://doi.org/10.1111/2041-210X.12492

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

Hawkins, B. A., Field, R., Cornell, H. V., Currie, D. J., Guégan, J.-F., Kaufman, D. M., … Turner, J. R. G. (2003). Energy, water, and broad-scale geographic patterns of species richness. *Ecology*, *84*, 3105–3117. https://doi.org/10.1890/03-8006

Hijmans, R. J. (2017). *raster: geographic data analysis and modeling (R package version 2.6)*. Retrieved from https://cran.r-project.org/web/packages/raster/index.html

Kubota, Y., Kusumoto, B., Shiono, T., Ulrich, W., & Jabot, F. (2015). Non-neutrality in forest communities: Evolutionary and ecological determinants of tree species abundance distributions. *Oikos*, *125*, 237–244. https://doi.org/10.1111/oik.02232

Locey, K. J., & White, E. P. (2013). How species richness and total abundance constrain the distribution of abundance. *Ecology Letters*, *16*, 1177–1185. https://doi.org/10.1111/ele.12154

Mac Nally, R., Duncan, R. P., Thomson, J. R., & Yen, J. D. L. (2018). Model selection using information criteria, but is the "best" model any good? *Journal of Applied Ecology*, *55*, 1441–1444.

MacArthur, R. H. (1960). On the relative abundance of species. *The American Naturalist*, *94*, 25–36. https://doi.org/10.1086/282106

MacArthur, R. H. (1972). *Geographical ecology*. Princeton, NJ: Princeton University Press.

Magurran, A. E., & Henderson, P. A. (2003). Explaining the excess of rare species in natural species abundance distributions. *Nature*, *422*, 714–716. https://doi.org/10.1038/nature01547

Matthews, T. J., & Whittaker, R. J. (2015). On the species abundance distribution in applied ecology and biodiversity management. *Journal of Applied Ecology*, *52*, 443–454.

Matthews, T. J., Borges, P. A. V., de Azevedo, E. B., & Whittaker, R. J. (2017). A biogeographical perspective on species abundance distributions: Recent advances and opportunities for future research. *Journal of Biogeography*, *44*, 1705–1710. https://doi.org/10.1111/jbi.13008

Matthews, T. J., Borregaard, M. K., Gillespie, C. S., Rigal, F., Ugland, K. I., Krüger, R. F., … Whittaker, R. J. (2018). Extension of the gambin model to multimodal species abundance distributions. *Methods in Ecology and Evolution*, https://doi.org/10.1111/2041-210X.13122

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. https://doi.org/10.1111/ecog.00861

May, R. M. (1975). Patterns of species abundance and diversity. In M. L. Cody & J. M. Diamond (Eds.), *Ecology and evolution of communities* (pp. 81–120). Cambridge, MA: Harvard University Press.

McGill, B. J. (2011). Species abundance distributions. In A. E. Magurran & B. J. McGill (Eds.), *Biological diversity: Frontiers in measurement and assessment* (pp. 105–122). Oxford, UK: Oxford University Press.

McGill, B. J., Etienne, R. S., Gray, J. S., Alonso, D., Anderson, M. J., Benecha, H. K., … White, E. P. (2007). Species abundance distributions: Moving beyond single prediction theories to integration within

an ecological framework. *Ecology Letters*, *10*, 995–1015. https://doi.org/10.1111/j.1461-0248.2007.01094.x

O'Connell, B. M., Conkling, B. L., Wilson, A. M., Burrill, E. A., Turner, J. A., Pugh, S. A., … Menlove, J. (2017). *The forest inventory and analysis database: database description and user guide for phase 2 (version 7.0)*. Retrieved from https://www.fia.fs.fed.us/library/database-documentation/current/ver70/FIADB%20User%20Guide%20P2_7-0_ntc.final.pdf

Preston, F. W. (1948). The commonness, and rarity, of species. *Ecology*, *29*, 254–283. https://doi.org/10.2307/1930989

R Core Team. (2017). *R: A language and environment for statistical computing (Version 3.4.3)*. Vienna, Austria: R Foundation for Statistical Computing.

Rosenzweig, M. L. (1995). *Species diversity in space and time*. Cambridge, UK: Cambridge University Press.

Ugland, K. I., & Gray, J. S. (1982). Lognormal distributions and the concept of community equilibrium. *Oikos*, *39*, 171–178.

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., Kusumoto, B., Shiono, T., & Kubota, Y. (2016). Climatic and geographic correlates of global forest tree species–abundance distributions and community evenness. *Journal of Vegetation Science*, *27*, 295–305. https://doi.org/10.1111/jvs.12346

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. https://doi.org/10.1016/j.ecocom.2018.07.002

Ulrich, W., Ollik, M., & Ugland, K. I. (2010). A meta-analysis of species–abundance distributions. *Oikos*, *119*, 1149–1155. https://doi.org/10.1111/j.1600-0706.2009.18236.x

Vergnon, R., van Nes, E. H., & Scheffer, M. (2012). Emergent neutrality leads to multimodal species abundance distributions. *Nature Communications*, *3*, 663. https://doi.org/10.1038/ncomms1663

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

Ward, M. D., & Gleditsch, K. S. (2008). *Spatial regression models*. Thousand Oaks, CA: Sage.

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

Whittaker, R. J., Willis, K. J., & Field, R. (2003). Climatic–energetic explanations of diversity: A macroscopic perspective. In T. M. Blackburn & K. J. Gaston (Eds.), *Macroecology: Concepts and consequences* (pp. 107–129). Malden, MA: Blackwell.

Wiens, J. J., Ackerly, D. D., Allen, A. P., Anacker, B. L., Buckley, L. B., Cornell, H. V., … Stephens, P. R. (2010). Niche conservatism as an emerging principle in ecology and conservation biology. *Ecology Letters*, *13*, 1310–1324. https://doi.org/10.1111/j.1461-0248.2010.01515.x

Xing, D., Swenson, N. G., Weiser, M. D., & Hao, Z. (2014). Determinants of species abundance for eastern North American trees. *Global Ecology and Biogeography*, *23*, 903–911. https://doi.org/10.1111/geb.12167

BIOSKETCH

Tom Matthews is a Research Fellow at the University of Birmingham, U.K. His research interests span community ecology, macroecology and biogeography. He is particularly focused on investigating global environmental change questions using macroecological methods and theory.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Matthews TJ, Sadler JP, Kubota Y, Woodall CW, Pugh TAM. Systematic variation in North American tree species abundance distributions along macroecological climatic gradients. *Global Ecol Biogeogr.* 2019;28:601–611. https://doi.org/10.1111/geb.12879