

A hierarchical Bayesian model to incorporate uncertainty into methods for diversity partitioning

ZACHARY H. MARION ², JAMES A. FORDYCE, AND BENJAMIN M. FITZPATRICK

Department of Ecology & Evolutionary Biology, 569 Dabney Hall, University of Tennessee, Knoxville, Tennessee 37996 USA

Abstract. Recently there have been major theoretical advances in the quantification and partitioning of diversity within and among communities, regions, and ecosystems. However, applying those advances to real data remains a challenge. Ecologists often end up describing their samples rather than estimating the diversity components of an underlying study system, and existing approaches do not easily provide statistical frameworks for testing ecological questions. Here we offer one avenue to do all of the above using a hierarchical Bayesian approach. We estimate posterior distributions of the underlying “true” relative abundances of each species within each unit sampled. These posterior estimates of relative abundance can then be used with existing formulae to estimate and partition diversity. The result is a posterior distribution of diversity metrics describing our knowledge (or beliefs) about the study system. This approach intuitively leads to statistical inferences addressing biologically motivated hypotheses via Bayesian model comparison. Using simulations, we demonstrate that our approach does as well or better at approximating the “true” diversity of a community relative to naïve or ad-hoc bias-corrected estimates. Moreover, model comparison correctly distinguishes between alternative hypotheses about the distribution of diversity within and among samples. Finally, we use an empirical ecological dataset to illustrate how the approach can be used to address questions about the makeup and diversities of assemblages at local and regional scales.

Key words: community composition; differentiation; effective diversity; estimation; finite sampling; Hill numbers; Markov chain Monte Carlo; overlap; statistical comparison; turnover.

INTRODUCTION

One cornerstone of ecology is comparing and explaining genomic (e.g., Ihmels et al. 2007), phenotypic (e.g., Marion et al. 2015), or community diversity of natural systems (e.g., MacArthur 1965, Torsvik and Øvreås 2002). This naturally leads to important questions about how diversity should be quantified in theory and estimated in practice. Many classical ecological theories address species richness (MacArthur and Wilson 1967, Pianka 1974). Others incorporate concepts of evenness or probability of identity (i.e., the chance that two individuals are the same species; Simpson 1949, Preston 1962, Hubbell 2001). Key questions have to do with changes in diversity over time (Menge and Sutherland 1976, Stegen et al. 2013), dependence of diversity on area (Whittaker et al. 2001, Kraft et al. 2011), relationships between diversity and environment (Ricklefs 2006, Chao et al. 2016), and how diversity is distributed within and between communities (Whittaker 1960, Anderson et al. 2011). There has been substantial progress regarding theoretical metrics and their decomposition into within- and among community components (Legendre and De Cáceres 2013, Chao et al. 2014). Yet, a general understanding of the challenges and pitfalls of estimating desired quantities from real data has lagged behind. Here we attempt to move the empirical side of the field forward with a Bayesian approach that helps address difficult problems of bias and uncertainty in estimating, partitioning, and comparing diversity.

Recently, there has been a general consensus that metrics based on Hill numbers (Hill 1973, Jost 2006) are intuitive and appropriate for ecology (see Ellison 2010). Hill number diversities (qD) are mathematically related metrics that vary in sensitivity to differential abundances according to the parameter q (Appendix S1). When $q = 0$, only presences or absences matter (0D = average richness or number of species present within a unit). When $q = 1$ (1D = exponential of Shannon entropy), species are weighted by their abundances without emphasizing rare or common species. When $q \geq 2$ (2D = inverse of Simpson's concentration), qD depends increasingly strongly on the most abundant species; rare species have less influence.

Hill numbers convert measures of entropy or the probability of identity into effective numbers: the expected diversity of an idealized sample containing equally abundant species. It is often helpful to visualize the diversities together on a graph as a continuous function of q (Chao et al. 2014). This diversity profile characterizes both the diversity and the species abundance distribution. The steepness of the slope describes the evenness of the species abundance distribution: steeper slopes indicate less even communities. Further, Hill numbers provide independent estimates of alpha diversity (average single-location diversity) and beta diversity (relative change in composition among locations). This enables the direct comparison of diversities across different systems (Jost 2006).

Despite this conceptual progress, estimating and partitioning diversity based on real data remains challenging because of finite sampling. First, it is well established that the diversity of a sample tends to underestimate the true diversity of the underlying unit from which the sample was taken (Gotelli and Colwell 2001, Paninski 2003, Hausser

Manuscript received 13 April 2017; revised 3 January 2018; accepted 22 January 2018. Corresponding Editor: Tom E. X. Miller.

² E-mail: zmarion@vols.utk.edu

and Strimmer 2009). This arises in part because rare species are often missed unless all individuals within a patch or community are found and identified. But most diversity measures, including Hill numbers, are still biased even when all species are accounted for because they are nonlinear functions of proportions (Ben-Bassat and Raviv 1978, Dana 2005, Jost 2006). Authors in a variety of fields have proposed bias corrections (Miller 1955, Hausser and Strimmer 2009, Chao et al. 2013), but estimating the number and abundances of unseen species remains a challenge despite ongoing progress (Chao and Jost 2015, Chao et al. 2015).

Second, when partitioning diversity among sampling units (such as islands, ponds, or other discrete patches), the among-sample diversity component (β diversity) and measures of overlap or turnover derived from it are dependent on the number of sampling units. This means β diversity of an archipelago of 33 islands cannot be compared with β diversity of an archipelago of 42 islands without first standardizing to a fixed number of islands. Likewise, β diversity of any set of habitat patches (such as an archipelago) cannot be estimated from a subsample of patches. Elsewhere, we have suggested that the average pairwise turnover is the most intuitive way to express the among-unit component of diversity in a manner that can be compared across systems and studies (Marion et al. 2017).

Third, many ecological data sets are comprised of discrete samples (e.g., transects or quadrats) from contiguous systems. For example, the Barro Colorado Island tree data of Condit et al. (2002) consist of 50 plots within a single continuous tropical forest. The partitioning of diversity within and between such plots is peculiar to the sampling design, not intrinsic to the contiguous system. Nonetheless, ecologists often use diversity decomposition estimates from discrete samples to learn about the diversity and heterogeneity of continuous systems (e.g., Kraft et al. 2011).

Finally, although multiple samples are often collected, uncertainty associated with sampling is often ignored and only a single point estimate of diversity reported. When error is considered, uncertainty intervals around the point estimate are usually generated through resampling or parametric bootstrapping (Efron 1982). Parametric bootstrapping (using the multinomial distribution, for example) is often within sample. This assumes the total abundance of a site is fixed or known with certainty. Nonparametric bootstrapping usually relies on resampling of sites, assuming samples are fixed or true. In both instances, uncertainty intervals are limited by sample size and effort. Thus, ecologists end up describing the sample diversity rather than diversity of the underlying system of interest. Another challenge is that ecologists are usually interested in comparisons, e.g., differences between islands or mountain ranges. This is a fundamentally statistical problem. Quantifying uncertainty around diversity estimates is a start, but an integrative statistical approach is still lacking.

Here, we present a framework for statistical analysis that addresses these challenges. We advocate hierarchical Bayesian modeling (e.g., Zhang et al. 2014) to explicitly estimate uncertainty in species abundances both within-sample and assemblage-wide. These posterior estimates of relative abundances can then be used post-hoc with any diversity estimation method as derived quantities. Our approach can accommodate problems of finite sampling such as unequal

sample sizes or sample effort. We can then use model comparison through the widely applicable information criterion (WAIC, a Bayesian extension of the Akaike information criterion; Watanabe 2010) or cross-validation (Watanabe 2010, Vehtari et al. 2016b) to assess whether patches/communities/habitats within regions are distinct subassemblages or whether they are “arbitrary” distinctions from one contiguous system, all within one integrated statistical pipeline that propagates the uncertainty from beginning to end.

METHODS

Conceptual overview

Our modeling framework begins with a sample \times species matrix where elements are counts of each species in each unit sampled. We then estimate the species abundances of each sample unit using a Bayesian model (for a comprehensive introduction to Bayesian statistics, see Gelman et al. [2013]). As a general overview:

1. Initial parameter values are proposed: for each sample, a set of multinomial probabilities corresponds to each species. Parameter $\theta_{i,k}$ is the probability an individual found in site i is species k .
2. The likelihood of the data given the parameters is multiplied by the prior probability of the parameters giving a posterior probability estimate of the multinomial θ 's, interpreted as the relative abundances of species for a sample. For each iteration, the parameters are saved or rejected according to the Markov chain algorithm.
3. New parameter values are proposed and the process repeated for a number of iterations.
4. Because the first iterations constitute the adaptive phase of the sampler, these “warm-up” estimates are discarded. The remaining iterations are interpreted as samples from the posterior distribution of relative abundances.
5. For each sample from the posterior, θ 's are used post-hoc to estimate and partition diversity as a derived quantity. In principle, the posterior θ 's can be plugged into formulae for any diversity metric with an ad-hoc bias correction or not. Because we have a posterior distribution of relative abundances, we also have a posterior distribution of diversities that incorporates uncertainty.

The multinomial θ 's are probabilities of occurrence for a species. Although a species may never be observed, that probability estimate will never be zero. Thus, this model assumes we have a comprehensive list of species in the true community, so we assume true richness (0D) is known. The problem of estimating richness has been addressed in several ways (Colwell and Coddington 1994, Gotelli and Colwell 2001, 2010). Many of these can be incorporated with our approach in an ad hoc manner (see *Discussion*), but simultaneous estimation of richness (i.e., the “support size” of a multinomial distribution) and relative abundances (the event probability parameters of the multinomial distribution) requires more assumptions or more information than the framework described here.

Posterior predictive simulations can be used to estimate expected sample richness (or other metrics) given a sampling

scheme. Posterior predictive simulation is often used in Bayesian statistics to check model adequacy by simulating “new” data from the posterior parameter estimates and then comparing the new data to the existing data (Gelman et al. 2013). Here, we generate posterior simulated data as follows:

- 6). For each iteration, we simulate sample data from the posterior distribution of θ 's.

Simulated sample data are saved and then used to estimate diversity as well.

The posterior predictive estimate of diversity inherently describes the sample diversity, while the diversity estimates based on theta describe the diversity of the system. Thus we make a distinction between estimation of the underlying community diversity (i.e., model parameters) and estimation of expected sample diversity (i.e., describing data) given a sampling design. The posterior predictive samples can be used as a form of “Bayesian rarefaction” to assess how the number of observed species is expected to change with sample size and make comparisons between samples based on a standardized sample size.

The Bayesian model

Consider a data set of N samples (e.g., quadrats, transects, or natural units like ponds or islands), with S species from 1, 2, ..., S in total among samples. Let $\mathbf{Y} = [y_{i,k}] \geq 0$ be an $N \times S$ community abundance matrix:

$$\mathbf{Y} = \begin{bmatrix} y_{1,1} & y_{1,2} & \cdots & y_{1,S} \\ y_{2,1} & y_{2,2} & \cdots & y_{2,S} \\ \vdots & \vdots & \ddots & \vdots \\ y_{N,1} & y_{N,2} & \cdots & y_{N,S} \end{bmatrix} \quad (1)$$

where $y_{i,k}$ is the count of species k in the i th sample, and $\sum \bar{y}_i = \sum_{k=1}^S y_{i,k} = n_i$. We assume that the vector of species abundances for sample i follows a multinomial distribution

$$\bar{y}_i \sim \text{Multinomial}(\bar{\theta}_i, n_i) \quad (2)$$

where n_i is the total abundance of individuals for observation i and the parameter vector $\bar{\theta}_i$ describes the probabilities of species occurrences for sample i . More intuitively, we can interpret $\theta_{i,k}$ as the relative abundance of species k in sampling unit i .

For each $\bar{\theta}_i$, we use softmax normal priors with mean μ_k and common standard deviation σ . The softmax transformation takes a vector (α_i) with unconstrained (i.e., $-\infty$ to $+\infty$) values and constrains it to a k -simplex:

$$\begin{aligned} \bar{\theta}_i &= \text{softmax}(\bar{\alpha}_i) \\ &= \frac{\exp(\alpha_{i,k})}{\sum_{k=1}^S \exp(\alpha_{i,k})} \end{aligned} \quad (3)$$

$$\bar{\alpha}_i \sim \text{Normal}(\bar{\mu}, \sigma). \quad (4)$$

We used the softmax function with unconstrained normal priors for two reasons. First, the softmax function is a

multivariate extension of the logit link function, which means that predictor variables could be included in the model as a multinomial regression. Second, this parameterization was much more numerically stable and computationally efficient than an alternative like the traditional conjugate prior of the multinomial distribution, the Dirichlet distribution.

The softmax function is many-to-one, which can lead to problems of nonidentifiability for the unconstrained α 's. However, only $K-1$ parameters are necessary to parameterize a K -simplex because the K th is one minus the sum of the other $K-1$ parameters. Therefore we mitigate the non-identifiability by fixing the first component of α_i to zero. We assume the prior vector of means $(\bar{\mu})$ follows a normal distribution, and the prior standard deviation σ is half-Cauchy distributed

$$\bar{\mu} \sim \text{Normal}(0, \tau_\mu) \quad (5)$$

$$\sigma \sim \text{Cauchy}^+(0, \tau_\sigma). \quad (6)$$

For computational efficiency, we use a non-centered parameterization of the normal distribution (Papaspiliopoulos et al. 2007), and the priors for both μ and σ are centered on zero with weakly informative scale parameters τ (Gelman et al. 2008). The folded Cauchy distribution was chosen as a prior for scale parameters because it has fatter tails than the normal distribution, which allows occasional sampling of more extreme values while keeping most of the MCMC sampling in the realm of smaller, more plausible values. Alternatively, a folded t-distribution could be used to control the probability mass of the tails.

The μ 's estimated for an assemblage can be converted to relative abundances using the softmax transformation: $\Theta = \text{softmax}(\bar{\mu})$ resulting in the full K -simplex. Θ can be conceptualized as the average assemblage-wide relative abundance for each species.

The main parameters of interest are the multinomial θ 's because these estimate “true” relative abundances for each sampling unit. Because θ is a vector of probabilities, they will never be zero even though a species may truly be absent, i.e., given the multinomial assumption, observed patterns of presence/absence depend on sample size. This poses a problem if one wants to estimate presence/absence-based diversity metrics (e.g., Hill numbers with $q = 0$). We therefore use posterior predictive simulation to simulate sample data given the posterior parameter estimates at each Markov chain iteration. The average abundance of each observation n_i is modeled as Poisson-distributed according to a single λ describing the mean abundance across samples

$$n_i \sim \text{Poisson}(\lambda). \quad (7)$$

The parameter λ is then given a weakly informative half-Cauchy distribution

$$\lambda_i \sim \text{Cauchy}^+(0, \tau_\lambda). \quad (8)$$

For each iteration, we used predictive simulation to generate a “new” total standardized abundance \bar{n} from the posterior estimate of the Poisson λ . This generates a new abundance vector \bar{y}_i

$$\tilde{n} \sim \text{Poisson}(\lambda) \quad (9)$$

$$\tilde{y}_i \sim \text{Multinomial}(\tilde{\theta}_i, \tilde{n}). \quad (10)$$

We use a single λ to model the total abundance for the posterior predictive simulations to standardize sampling effort among samples. Alternatively, one could use a fixed N (e.g., the smallest sample size in the data set) as often done for rarefaction, or one could estimate a separate λ for each sample, depending on the question. In the context of richness, the expected sample richness is estimated as the mean number of species observed in posterior predicted replicates.

For all models described here, posterior probabilities for model parameters given the data (i.e., species abundance matrix \mathbf{Y}) were estimated using the RStan interface (Stan Development Team 2016; Carpenter et al. 2017). We chose prior values of τ (the scale parameters for our priors) that were weakly informative as recommended by Gelman et al. (2008). Specifically, $\tau_\mu = 5$, $\tau_\sigma = 2.5$, and $\tau_\lambda = 25$. An extensive sensitivity analysis (Appendix S2) revealed that our results are robust to prior specification, with little change in the results regardless of how informative the priors were.

We ran four chains with 5,000 iterations each, discarding the first 2,500 iterations per chain as warm-up and resulting in 10,000 posterior samples overall. We used several diagnostic tests to confirm that each model had reached a stationary distribution, including visual examination of the chain history, calculation of effective sample size (ESS), and the Gelman-Rubin convergence diagnostic (\hat{R} ; Gelman and Rubin 1992, Brooks and Gelman 1998). In particular, model convergence was assessed by inspecting the diagnostics of the log-posterior density. All of the code is provided as an online supplement (Data S1).

Statistical inference: are assemblages distinct?

Often ecologists are interested in assessing whether a community consists of a single contiguous assemblage or a patchwork of distinct subassemblages. Bayesian model selection provides a natural extension of statistical inference for addressing such hypotheses by comparing complete-pooling and no-pooling models. A complete-pooling model has a single prior distribution, consistent with the hypothesis that the community of interest is comprised of one contiguous assemblage. A no-pooling model assigns separate prior distributions to each hypothesized subassemblage.

The strengths of different hypotheses can be quantified by comparing models using WAIC, leave-one-out (LOO), or k -fold cross-validation (Vehtari et al. 2016b). Here we use WAIC because its interpretation is similar to AIC, widely used for model comparison in maximum-likelihood-based statistics. WAIC is fully Bayesian and uses the entire posterior distribution. Thus WAIC provides an estimate of uncertainty for comparison of predictive errors between two (or more) models (Vehtari et al. 2016b).

We implemented model comparison by saving the posterior distribution of log-likelihoods from both complete and no-pooling models. We then calculated the difference in WAIC values between models and the predictive errors of the difference using the loo package (Vehtari et al. 2016a).

Like AIC or DIC, there are no firm criteria for determining when one model is unambiguously “better” than another. However, we can use the standard errors that WAIC and LOO provide as strength of the evidence. If the error in predictive difference between models does not overlap zero, we can consider that as strong evidence in favor of one model over the other.

Statistical inference: comparison of assemblage diversities

If model selection suggests that subassemblages should be modeled separately, one might be interested in which is more diverse or heterogeneous. At the within-subassemblage level, alpha, beta (or standardized beta), and gamma estimates can be compared by plotting the diversity profiles of the marginal mean (or median) diversity estimates along with their uncertainty intervals such as the 95% highest density interval (HDI) for a range of q (different weightings of evenness).

At the regional scale, we can estimate posterior α , β , or γ similarly using the Θ 's after softmax transforming the μ priors. Each Θ is the average relative abundance within assemblage, and the posterior diversity estimates at this scale represent regional diversity across subassemblages.

Simulation study of model efficacy

We evaluated the performance of our method by simulating artificial data sets using an approach from Cayuela et al. (2015). For each iteration, we generated a log-normal species abundance distribution. The log-normal distribution was chosen because, depending on parameter values—the distribution is right-skewed (few common species and many rare ones), typical of many ecological communities (Preston 1948). Then we created a “true” community matrix by populating sites/observations with individuals in proportion to the generated species abundance distribution. The total counts of individuals in a site were Poisson distributed around a mean number of individuals in the community (generally $\lambda = 500$), while the total number of sites in the “true” community varied between 30 and 75. We then created an observed community by sampling individuals with replacement from each site with sample effort P , the probability of completely sampling every individual, using a random draw from a binomial distribution.

We then used our Bayesian framework to estimate abundances of sampled communities under two scenarios: low (12%) and high (80%) sampling effort. For the low sampling effort, 12% represents a situation (given our simulation parameters) where the sample size of individuals is similar (on average) to the total number of species in the community, a condition where the problem of sampling bias is known to be severe (Paninski 2003). For each simulation, we partitioned the total diversity (γ) into α and β using the mean of the posterior simulations for $q = 0, 1$, and 2. We did the same for the mean of the multinomial θ 's for $q = 1$ and 2.

For each simulation, estimated diversities for the Bayesian model were compared to the diversities of the “true” community by taking the difference (Δ diversity = estimated – true). We also calculated Δ diversity using two different approaches as benchmarks. For the first, we used a point (“naïve”)

estimate of the diversity from the observed community. The second benchmark was similar to the naïve estimate except we applied a bias correction (Chao and Jost 2015) using the *entropart* R package (Marcon and Herault 2015).

Simulation study for comparing assemblages

To quantify the performance of our approach at identifying distinct subassemblages, we again simulated communities by modifying code from Cayuela et al. (2015) to assess whether simulated communities were better modeled as one contiguous community or two distinct subassemblages. For each simulated community, the total number of observations N for the combined communities varied randomly between 30 and 75 as in the simulations described in *Simulation study of model efficiency*. Total sample size was then partitioned randomly into two subassemblage sizes with the minimum sample size within a subassemblage fixed at 10. A “true” community was then generated as previously described and “sampled” with 80% effort to create a true community.

Within this overall simulation framework, we tested three scenarios (Cayuela et al. 2015).

- 1). One community artificially divided into subassemblages. In this scenario, representing the null expectation that there is a single contiguous community, we generated one log-normal species abundance distribution comprised of 60 species total and then populated sites in proportion to relative abundances of those species. We then randomly divided this contiguous community into two as described *Simulation study of model efficiency*.
- 2). Two assemblages that differ in richness but not evenness. This scenario was similar to the first except that we generated two log-normal species abundance distributions that were identical except that one had 20 and one had 60 species. Note that the richness levels refer to the species abundance distributions from which the subassemblages were drawn. The differences in observed richness among samples varied.
- 3). Two assemblages that differ in evenness but not richness. Here the rank abundance distributions of the two subassemblages had a fixed difference in log standard deviation of 1 unit and ranged randomly between 0.095 and 3.5 on a log scale.

Each simulation scenario was run for 500 iterations each. For each iteration, we calculated the difference in WAIC values between a one-assemblage (complete-pooling) and two-assemblage (no-pooling) model. If the one-assemblage model was more plausible, the WAIC difference were negative; if the two assemblage model was favored, the difference was positive. We then assessed the performance of our approach with three benchmarks from most to least conservative. We asked what proportion of iterations per scenario: (1) unambiguously returned the correct result (i.e., more than one standard error from zero and correct in sign); (2) the point estimate (average WAIC) of the difference was correct but was within one standard error of zero; (3) the point estimate was incorrect but was within one standard error of zero; (4) unambiguously suggested the incorrect model with more than one standard error from zero and incorrect in sign.

RESULTS

Model efficacy simulations

Overall, the model performed well at estimating true diversity (Fig. 1). When sample completeness was low, such that the total abundance is close to the number of species, diversity estimates calculated from posterior predictive data performed about as well as the naïve plug-in estimator. For $q \geq 1$ however, the posterior θ estimates did a much better job, on par with the bias-corrected estimates. As sample completeness increased (i.e., 80%), estimates from the posterior predictive data generally outperformed the naïve estimates due to “information sharing” via the prior. As before, diversities from the posterior θ 's did an excellent job at estimating true diversities for $q \geq 1$ (Note: for clarity we have only included the results for $q \geq 1$. To see the full results with $q = 0$ included, see Appendix S3).

Comparing assemblages

Model selection identified the correct model the vast majority of the time (Table 1). Across the three scenarios, WAIC strongly supported the incorrect model (i.e., standard error of Δ WAIC not overlapping zero) for $\leq 1\%$ of the 500 simulations. The WAIC point estimate was consistent with the correct model for $\approx 97\%$ (one assemblage and different richness) and 98.4% (different evenness). WAIC strongly supported the correct model for 91–94% of simulations that consisted of two different subassemblages, but had a harder time correctly selecting the null for the one-assemblage community simulations (standard error of Δ WAIC not overlapping zero in $\approx 80\%$ of simulations).

Applications: oribatid mite diversity

We applied our Bayesian framework to estimate the diversities of oribatid mites using the data of Borcard et al. (1992) and described more fully in Borcard and Legendre (1994). *Sphagnum* moss cores were sampled from a moss and peat blanket floating in Lac Geai at the Station de Biologie des Laurentides, Quebec, Canada. Samples were stratified by substrate types (*Sphagnum* spp., substratum morphology, shrub cover) and sampled proportionally to substratum surface area (Borcard et al. 1992). The result was a 70 sample \times 35 species community matrix. Borcard et al. (1992) used canonical ordination to partition variation in species diversity into environmental and spatial components. They found that environmental variables explained $\approx 45\%$ of the variation in community structure, and that species data and environmental variables had similar spatial structure, possibly due to similar underlying mechanisms. However, the goal of these papers was to tease out the roles of spatial and environmental variation in structuring communities, not to quantify the diversity of the mite community itself.

Here, we focus on one habitat characteristic, a shrub density gradient on a semi-quantitative scale (no shrubs < few shrubs < many shrubs), to highlight the utility of our approach. We chose this characteristic because environment explained much of the mite community structure, and shrub

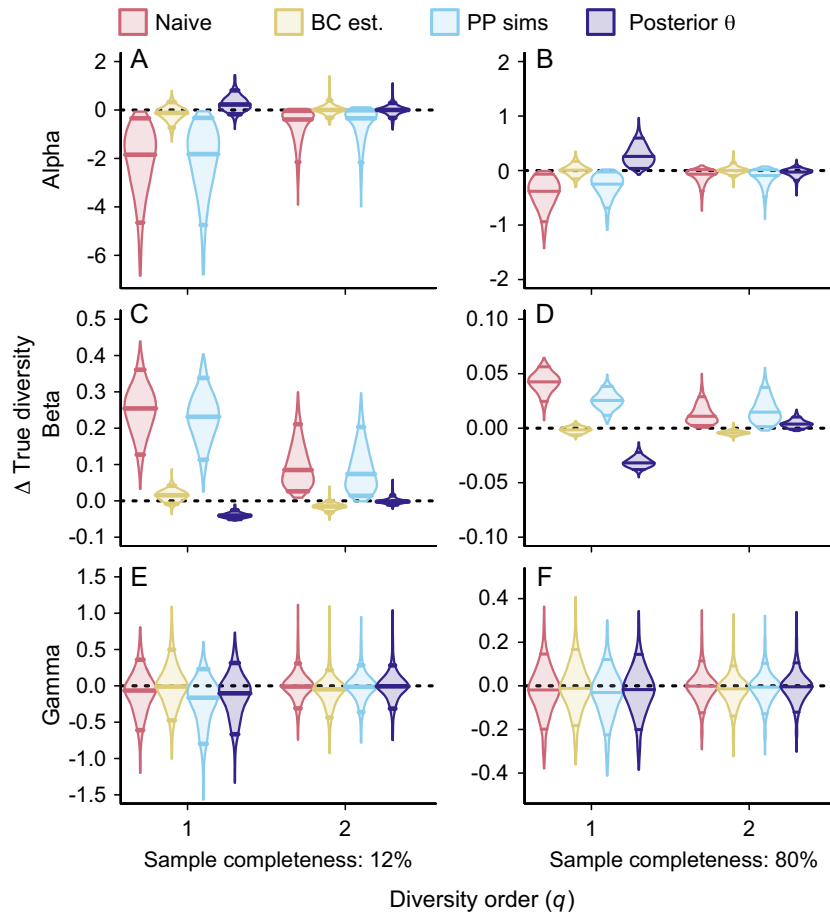


FIG. 1. Distribution of deviations from true diversity (Δ true D) under two levels of sampling completeness for diversity orders (q) = 1–2. For each simulation, mean posterior predictive estimates (PP sims; blue; estimated from data simulated from the posterior parameters) and mean posterior θ diversity estimates (purple; estimated directly from the posterior distributions of relative abundances) were computed and then subtracted from the true diversity value. We also compared diversity estimates from the naive plug-in estimator (red) and an estimate that corrects for unseen species (BC est.; yellow). Δ diversities greater than zero overestimate true diversity, while Δ less than zero underestimate true diversity; values at zero (dotted lines) are correct. A sample completeness of 12% means that, on average for each site, the number of individuals observed was equal to the total number of species in the true community. Violin plot lines indicate the mean and upper and lower 95% uncertainty intervals across simulations.

TABLE 1. Simulation results of WAIC to correctly identify when to model a community as one contiguous assemblage or two distinct subassemblages for three scenarios: (1) a null model of a single community randomly divided into two subcommunities; (2) two assemblages that had different richness but similar evenness; and (3) two assemblages that had similar richness but differed in evenness.

Simulation scenario	Strongly correct, %	Point estimate correct, %	Point estimate incorrect, %	Strongly incorrect, %
Same community	78.4 (392)	18.4 (92)	3.2 (16)	0.00 (0)
Different richness	91.4 (457)	5.2 (26)	2.4 (12)	1.00 (5)
Different evenness	94.4 (472)	4 (20)	1.2 (6)	0.04 (2)

Notes: For each scenario, we quantified performance as the percentage (count in parentheses) out of 500 simulations that were strongly correct or strongly incorrect (SE of Δ WAIC did not overlap zero), or the point estimate was correct or incorrect (the mean Δ WAIC overlapped zero).

density accounted for most of the environmental variation (Borcard et al. 2004). We used WAIC to ask whether to model the community as one assemblage (complete pooling) or different assemblages along a shrub density gradient (no-pooling).

Bayesian model selection supported the three-community model over the single-community model. The posterior difference in WAIC was positive and had a standard error that

did not overlap zero (Δ WAIC 31.92 ± 23.08 ; mean \pm SE). Thus there was strong support for the model that estimated separate parameters for each shrub density category compared to a model that considered the mite community as one contiguous assemblage.

Therefore, we proceeded to partition diversity within and between samples and shrub categories to address the questions of whether mite assemblages differed in diversity and

heterogeneity. We calculated alpha and standardized beta diversity within each shrub-density category using multinomial θ 's and regionally using Θ 's for between 1 and 3. We standardized beta as the average pairwise turnover for comparison among the shrub density levels because of unequal sample sizes (Marion et al. 2017, for caveats, see Baselga 2013).

As shrub density increased, so did within-sample mite diversity (Fig. 2A and B). When posterior estimates of relative abundance were weighted equally ($q = 1$), samples from areas of *Sphagnum* with highest shrub density had approximately 10.5 mite species on average (95% uncertainty interval [UI]: 10.18, 11.06) while samples from areas with few to no shrubs had lower diversity with 8 (7.65, 8.34) and 6.7 (6.26, 7.09) effective species, respectively. At the regional level, samples had 9.68 (8.48, 10.89) effective species on average for $q = 1$. Profile plots decreased rapidly as diversity order increases, suggesting that the mite community is comprised of a few common taxa and several species of low abundance.

Pairwise turnover revealed more interesting patterns of diversity (Fig. 2C and D). The intermediate level of shrub density had almost double the within-habitat heterogeneity

($q = 1$; heterogeneity 0.28 [0.26, 0.29]) compared to the low and high ends of the gradient. Moreover, there was substantial overlap in posterior uncertainty estimates for the no-shrub samples (95% UI: 0.15, 0.2) and high-density samples (95% UI: 0.18, 0.21). Despite the overlap in uncertainty, when considering the joint posterior, there was lower average turnover between no-shrub samples and high-density samples for ~89% of posterior iterations across orders of q . There was also substantial species turnover across the shrub diversity gradient, especially as rarer mite species decreased in importance: up to ~30% (0.24, 0.42) when $q = 3$. This suggests that the identities of the most abundant species (i.e., rank order of relative abundance) differed among shrub density categories.

It makes sense that alpha diversity would increase with shrub density. As the number of shrubs increases, so might habitat complexity, resource input through leaf litter, and opportunities for multitrophic interactions: each of which might increase diversity (MacArthur 1965, Menge and Sutherland 1976). Moreover, the mite community's species-abundance distribution is dominated by a few common species and many rare ones, as are many ecological communities (Preston 1948). As for beta diversity, the substantial amount

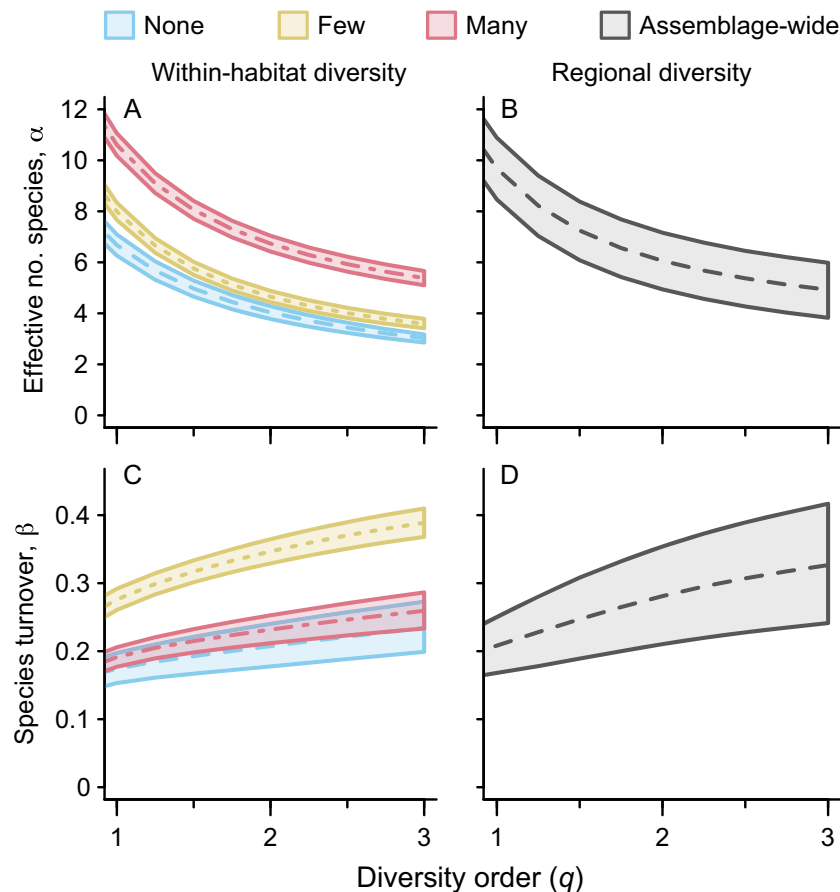


FIG. 2. Profile plots of α and β diversity of oribatid mites (Borcard et al. 1992, Borcard and Legendre 1994) for $q = 1$ –3. Average effective number of species (top) and average turnover in species composition (bottom) are shown at two scales: within-habitat (left) and regionally (right; grey). Mites were sampled across a semi-qualitative density gradient: none (blue) < few (yellow) < many (red). Broken lines represent the posterior median diversity estimate, and shaded regions encompass the 95% uncertainty interval. Regional beta diversity describes the average turnover in mite composition across the shrub density gradient.

of turnover among samples within the intermediate shrub density relative to the low or high densities could be environmentally caused, but we suspect it is simply a consequence of the semiquantitative scale used along the density gradient. The “few” category might encompass a larger range of shrub densities than the “none” or “many” category.

DISCUSSION

Over the last decade, there have been major advances in conceptualizing and quantifying diversity within and among communities, regions, and ecosystems. There has been less progress in the practical application of these advances to estimate diversity, account for uncertainty, and statistically test ecological questions based on sample estimates. Here we offer one avenue to do all of the above under the umbrella of hierarchical Bayesian modeling. The model does as well or better at approximating the “true” diversity of a community relative to naïve or ad-hoc bias-corrected estimates, especially for $q \geq 1$ (Fig. 1). Our framework can be used to answer questions and make comparisons about the makeup of assemblages at local and regional scales, as we demonstrated using both simulations (Table 1) and an empirical example using the oribatid mite dataset (Fig. 2; Borcard et al. 1992).

Strengths

This approach offers a step forward by providing a cohesive framework for partitioning diversity and testing related ecologically relevant hypotheses using sample data. A major strength of Bayesian statistics is that information is shared among samples through the prior. This has three important benefits. First, the posterior θ 's (relative abundance of each species in each sample) and Θ 's (average assemblage-wide relative abundance of each species) describe the relative abundances for the true underlying assemblage rather than describing the data. This is an important distinction as the data are peculiar to the sampling design and not basic attributes of contiguous systems.

Second, this information sharing implicitly offers a solution to unequal sample size and sampling effort. This means that rare species that might have been missed in one sample are accounted for through the prior from other samples. Such information sharing at least partially mitigates problems associated with finite sampling uncertainty.

A third benefit is that the method provides intuitive and natural error envelopes through the Bayesian posterior rather than relying on bootstrapping, which is fundamentally limited by the sample and challenging to implement hierarchically. Marion et al. (2015) developed a multilevel strategy for diversity partitioning that incorporated hierarchical bootstrapping, but error did not propagate across levels and thus information was lost. In contrast, uncertainty propagation is implicit for Bayesian models.

The value of this integrated approach is illustrated in our analysis of the mite data. We were able to statistically distinguish mite communities associated with different levels of shrub density, recapitulating a key result of the original studies, which had used ordination and randomization tests. In addition, we were able to address two novel questions: (1)

Did the different mite communities vary in α diversity (effective number of species per sample)? (2) Did mite assemblages associated with different habitat categories vary in heterogeneity (pairwise β diversity or average turnover between samples)? The answer to both questions was “yes,” and posterior distributions of the diversity components showed that the most heterogeneous (highest β) habitat category had neither the most nor least α diversity.

Remaining challenges

First, we want to be explicit that our contribution here is intuitive estimation of uncertainty for relative abundances, which can then be used downstream for any estimators or metrics of diversity, not just Hill numbers. We are not inventing new theory; we are providing a comparative statistical framework for ecologists working with samples. And our approach does have a few limitations, notably the inability to estimate richness. To estimate probabilities of occurrence as parameters of a multinomial distribution, our model takes the “support size” (species richness) as given, and no θ parameter is ever exactly zero. Estimating asymptotic species richness is a long-standing problem in ecology (Chao and Jost 2015, Chao et al. 2016). Current approaches require additional assumptions about the distribution of abundances or more information in the form of replication that may or may not be possible for many ecological study systems. In the future, we hope it will be feasible to simultaneously estimate richness and relative abundances in a Bayesian framework, but the applicability of such an approach might be fundamentally limited to datasets with extraordinary sampling effort.

One solution is to use posterior predictive simulation to estimate a posterior distribution of sample characteristics, such as expected sample richness, similar in spirit to traditional rarefaction. “Expected number of species in sample of size n ” might be valuable for comparing habitats or treatments. However, diversities calculated from this approach are estimates describing the sample and not necessarily the diversity of the real underlying system, which is ultimately what is of interest. The problem of unseen species is not unique to our approach, nor to ecology in general (Gotelli and Colwell 2001, Paninski 2003, Hausser and Strimmer 2009, Valiant and Valiant 2011). Until an integrated framework for estimating true richness is found, we envision three ad hoc solutions.

One is to use an asymptotic richness estimate to adjust the size of the community matrix (i.e., add all-zero columns) used as input to our estimation framework. We used such an approach (Appendix S2) in extensive simulations. We found that, at least for $q \geq 1$, the differences between the naïve (observed) species abundance matrix and using one appended with unseen species estimated using an asymptotic estimator (e.g., Chao estimator; Chao 1984, Chiu et al. 2014) was negligible. We admit that the benefit of this approach might be peculiar to particular sampling designs. A second is to apply bias corrections (e.g., Chao and Shen 2003, Chao et al. 2013, Marcon et al. 2015) to the output of our Bayesian model. A third is to use posterior simulations or rarefaction to describe expected observed richness given a sampling design. Posterior predictive simulation of “new” data, as

mentioned above, describes sample properties, i.e., the mean number of species observed in posterior predictive replicates. One could use this as a Bayesian rarefaction to explore how the effective diversity of species changes with sampling intensity. By fixing the sample size to a single value or a range (e.g., $N = 10$ individuals per site), one could make meaningful comparisons across systems or sampling efforts.

FUTURE DIRECTIONS

There is no reason this method could not be used to estimate phylogenetic or functional diversity (reviewed in Chao et al. 2014) in addition to taxonomic diversity because diversity estimation occurs on the posterior abundance estimates from the Bayesian model. Incorporating phylogenetic or functional trait information would be quite tractable without much computational burden.

The models here are not necessarily limited to either complete or no pooling. Depending on the question or situation, partial pooling models could be implemented. In partial pooling models, the priors for each subassemblage themselves share a common prior resulting in “shrinkage.” This means that, unless there is an abundance of information suggesting subassemblages are distinct, their estimates will be pulled towards each other through the shared hyperprior. It also usually results in improved estimates for subassemblages with low sample sizes. However, partial pooling would require another modeling level and may be more computationally challenging depending on sample sizes and the number of subassemblages modeled. As with the challenge of estimating richness, fitting more complex models often suffers from nonidentifiability without levels of replication rarely achievable in ecology.

Last, our Bayesian model relies on a multinomial likelihood function that currently limits analyses to discrete data (counts). We hope to develop an extension for continuous data in the near future. This would facilitate the analysis of communities when biomass or percent cover are used as measurements rather than individuals. This would also allow the method to be applied to phenotypic data such as multivariate chemical defenses, transcriptomes, or ecological metabolomics (reviewed in Marion et al. 2015).

Nonetheless, our model represents a useful advance in addressing biologically motivated questions about the diversities of communities under realistic sampling regimes.

ACKNOWLEDGMENTS

We thank N. Sanders, S. Campagna, C. Peterson, and two anonymous reviewers for helpful discussion and comments. This research was funded by the National Science Foundation (DDIG award no. DEB-1405887 to ZHM and DEB-0614223, 1050947, and 1638922 to J. A. Fordyce) and the Department of Ecology & Evolutionary Biology, University of Tennessee (Z. H. Marion). All authors contributed equally to this work.

LITERATURE CITED

- Anderson, M. J., et al. 2011. Navigating the multiple meanings of beta diversity: a roadmap for the practicing ecologist. *Ecology Letters* 14:19–28.
- Baselga, A. 2013. Multiple site dissimilarity quantifies compositional heterogeneity among several sites, while average pairwise dissimilarity may be misleading. *Ecography* 36:124–128.
- Ben-Bassat, M., and J. Raviv. 1978. Renyi's entropy and the probability of error. *IEEE Transactions on Information Theory* 24: 324–331.
- Borcard, D., and P. Legendre. 1994. Environmental control and spatial structure in ecological communities: an example using oribatid mites (Acari, Oribatei). *Environmental and Ecological Statistics* 1:37–61.
- Borcard, D., P. Legendre, and P. Drapeau. 1992. Partialling out the spatial component of ecological variation. *Ecology* 73:1045–1055.
- Borcard, D., P. Legendre, C. Avois-Jacquet, and H. Tuomisto. 2004. Dissecting the spatial structure of ecological data at multiple scales. *Ecology* 85:1826–1832.
- Brooks, S. P., and A. Gelman. 1998. General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics* 7:434–455.
- Carpenter, B., A. Gelman, M. D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, and A. Riddell. 2017. Stan: a probabilistic programming language. *Journal of Statistical Software* 76:1–32.
- Cayuela, L., N. J. Gotelli, and R. K. Colwell. 2015. Ecological and biogeographic null hypotheses for comparing rarefaction curves. *Ecological Monographs* 85:437–455.
- Chao, A. 1984. Nonparametric estimation of the number of classes in a population. *Scandinavian Journal of Statistics* 11:265–270.
- Chao, A., and L. Jost. 2015. Estimating diversity and entropy profiles via discovery rates of new species. *Methods in Ecology and Evolution* 6:873–882.
- Chao, A., and T.-J. Shen. 2003. Nonparametric estimation of Shannon's index of diversity when there are unseen species in sample. *Environmental and Ecological Statistics* 10:429–443.
- Chao, A., Y. T. Wang, and L. Jost. 2013. Entropy and the species accumulation curve: a novel entropy estimator via discovery rates of new species. *Methods in Ecology and Evolution* 4:1091–1100.
- Chao, A., C.-H. Chiu, and L. Jost. 2014. Unifying species diversity, phylogenetic diversity, functional diversity, and related similarity and differentiation measures through Hill numbers. *Annual Review of Ecology, Evolution, and Systematics* 45:297–324.
- Chao, A., T. C. Hsieh, R. L. Chazdon, R. K. Colwell, and N. J. Gotelli. 2015. Unveiling the species-rank abundance distribution by generalizing the Good-Turing sample coverage theory. *Ecology* 96:1189–1201.
- Chao, A., C.-H. Chiu, and L. Jost. 2016. Statistical challenges of evaluating diversity patterns across environmental gradients in mega-diverse communities. *Journal of Vegetation Science* 27: 437–438.
- Chiu, C.-H., Y.-T. Wang, B. A. Walther, and A. Chao. 2014. An improved nonparametric lower bound of species richness via a modified Good-Turing frequency formula. *Biometrics* 70: 671–682.
- Colwell, R. K., and J. A. Coddington. 1994. Estimating terrestrial biodiversity through extrapolation. *Philosophical Transactions of the Royal Society B* 345:101–118.
- Condit, R., et al. 2002. Beta-diversity in tropical forest trees. *Science* 295:666–669.
- Dana, R.-A. 2005. A representation result for concave schur concave functions. *Mathematical Finance* 15:613–634.
- Efron, B. 1982. The Jackknife, the bootstrap and other resampling plans. Volume 38. Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, USA.
- Ellison, A. M. 2010. Partitioning diversity. *Ecology* 91:1962–1963.
- Gelman, A., and D. B. Rubin. 1992. Inference from iterative simulation using multiple sequences. *Statistical Science* 7:457–472.
- Gelman, A., J. Aleks, P. M. Grazia, and S. Yu-Sung. 2008. A weakly informative default prior distribution for logistic and other regression models. *Annals of Applied Statistics* 2:1360–1383.
- Gelman, A., J. Carlin, H. Stern, D. Dunson, A. Vehtari, and D. Rubin. 2013. Bayesian data analysis. Third Edition. Chapman & Hall/CRC Texts in Statistical Science. CRC Press, Boca Raton, Florida, USA.

- Gotelli, N. J., and R. K. Colwell. 2001. Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters* 4:379–391.
- Gotelli, N. J., and R. K. Colwell. 2010. Estimating species richness. Pages 39–54 in A. Magurran and B. McGill, editors. *Biological diversity: frontiers in measurement and assessment*. Oxford University Press, Oxford, UK.
- Hausser, J., and K. Strimmer. 2009. Entropy inference and the James-Stein estimator, with application to nonlinear gene association networks. *Journal of Machine Learning Research* 10:1469–1484.
- Hill, M. O. 1973. Diversity and evenness: unifying notation and its consequences. *Ecology* 54:427–432.
- Hubbell, S. P. 2001. *The unified neutral theory of biodiversity and biogeography*. Volume 32. Princeton University Press, Princeton, New Jersey, USA.
- Ihmels, J., S. R. Collins, M. Schuldiner, N. J. Krogan, and J. S. Weissman. 2007. Backup without redundancy: genetic interactions reveal the cost of duplicate gene loss. *Molecular Systems Biology* 3:86.
- Jost, L. 2006. Entropy and diversity. *Oikos* 113:363–375.
- Kraft, N. J., et al. 2011. Disentangling the drivers of β diversity along latitudinal and elevational gradients. *Science* 333:1755–1758.
- Legendre, P., and M. De Cáceres. 2013. Beta diversity as the variance of community data: dissimilarity coefficients and partitioning. *Ecology Letters* 16:951–963.
- MacArthur, R. H. 1965. Patterns of species diversity. *Biological Reviews of the Cambridge Philosophical Society* 40:510–533.
- MacArthur, R., and E. Wilson. 1967. *The theory of island biogeography*. Landmarks in biology series. Princeton University Press, Princeton, New Jersey, USA.
- Marcon, E., and B. Hérault. 2015. entropart: an R package to measure and partition diversity. *Journal of Statistical Software* 67:1–26.
- Marcon, E., Z. Zhang, and B. Hérault. 2015. The decomposition of similarity-based diversity and its bias correction. HAL-01212435.
- Marion, Z. H., J. A. Fordyce, and B. M. Fitzpatrick. 2015. Extending the concept of diversity partitioning to characterize phenotypic complexity. *American Naturalist* 186:348–361.
- Marion, Z. H., J. A. Fordyce, and B. M. Fitzpatrick. 2017. Pairwise beta diversity resolves an underappreciated source of confusion in calculating species turnover. *Ecology* 98:933–939.
- Menge, B. A., and J. P. Sutherland. 1976. Species diversity gradients: synthesis of the roles of predation, competition, and temporal heterogeneity. *American Naturalist* 110:351–369.
- Miller, G. 1955. Note on the bias of information estimates. Pages 95–100 in H. Quastler editor. *Information theory in psychology II-B*. Free Press, Glencoe, IL.
- Paninski, L. 2003. Estimation of entropy and mutual information. *Neural Computation* 15:1191–1253.
- Papaspiliopoulos, O., G. O. Roberts, and M. Skold. 2007. A general framework for the parametrization of hierarchical models. *Statistical Science* 22:59–73.
- Pianka, E. R. 1974. *Evolutionary Ecology*. Harper & Row, New York, New York, USA.
- Preston, F. W. 1948. The commonness, and rarity, of species. *Ecology* 29:254–283.
- Preston, F. W. 1962. The canonical distribution of commonness and rarity: part I. *Ecology* 43:185–215.
- Ricklefs, R. E. 2006. Evolutionary diversification and the origin of the diversity-environment relationship. *Ecology* 87:S3–S13.
- Simpson, E. H. 1949. Measurement of diversity. *Nature* 163:688.
- Stan Development Team. 2016. RStan: the R interface to Stan, version 2.12.1. <http://mc-stan.org/>
- Stegen, J. C., et al. 2013. Stochastic and deterministic drivers of spatial and temporal turnover in breeding bird communities. *Global Ecology and Biogeography* 22:202–212.
- Torsvik, V., and L. Øvreås. 2002. Microbial diversity and function in soil: from genes to ecosystems. *Current Opinion in Microbiology* 5:240–245.
- Valiant, G., and P. Valiant. 2011. Estimating the unseen: an $N/\log(N)$ -sample estimator for entropy and support size, shown optimal via new CLTs. Pages 685–694 in *Proceedings of the Forty-third Annual ACM Symposium on Theory of Computing, STOC 11*. ACM, New York, USA.
- Vehtari, A., A. Gelman, and J. Gabry. 2016a. loo: Efficient leave-one-out cross-validation and WAIC for Bayesian models. R package version 0.1.6. <https://arxiv.org/abs/1507.04544>
- Vehtari, A., A. Gelman, and J. Gabry. 2016b. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing* 27:1–20.
- Watanabe, S. 2010. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research* 11:3571–3594.
- Whittaker, R. H. 1960. Vegetation of the Siskiyou Mountains, Oregon and California. *Ecological Monographs* 30:279–338.
- Whittaker, R. J., K. J. Willis, and R. Field. 2001. Scale and species richness: towards a general, hierarchical theory of species diversity. *Journal of Biogeography* 28:453–470.
- Zhang, J., T. O. Crist, and P. Hou. 2014. Partitioning of alpha and beta diversity using hierarchical [B]ayesian modeling of species distribution and abundance. *Environmental and Ecological Statistics* 21:611–625.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at <http://onlinelibrary.wiley.com/doi/10.1002/ecy.2174/supinfo>