RESEARCH ARTICLE

Methods in Ecology and Evolution   BRITISH ECOLOGICAL SOCIETY

Functional traits and community composition: A comparison among community-weighted means, weighted correlations, and multilevel models

Jesse E. D. Miller [ID] | **Ellen I. Damschen** | **Anthony R. Ives** [ID]

Department of Integrative Biology, University of Wisconsin-Madison, Madison, Wisconsin

Correspondence
Jesse E. D. Miller
Email: jedmiller@ucdavis.edu

Present address
Jesse E. D. Miller, Department of Biology, Stanford University, Stanford, California

Abstract

1. Of the several approaches that are used to analyse functional trait–environment relationships, the most popular is community-weighted mean regressions (CWMr) in which species trait values are averaged at the site level and then regressed against environmental variables. Other approaches include model-based methods and weighted correlations of different metrics of trait–environment associations, the best known of which is the fourth-corner correlation method.

2. We investigated these three general statistical approaches for trait–environment associations: CWMr, five weighted correlation metrics (Peres-Neto, Dray, & ter Braak, *Ecography*, 40, 806–816, 2017), and two multilevel models (MLM) using four different methods for computing *p*-values. We first compared the methods applied to a plant community dataset. To determine the validity of the statistical conclusions, we then performed a simulation study.

3. CWMr gave highly significant associations for both traits, whereas the other methods gave a mix of support. CWMr had inflated type I errors for some simulation scenarios, implying that the significant results for the data could be spurious. The weighted correlation methods had generally good type I error control but had low power. One of the multilevel models, that from Jamil, Ozinga, Kleyer, and ter Braak (*Journal of Vegetation Science*, 24, 988–1000, 2013) had both good type I error control and high power when an appropriate method was used to obtain *p*-values. In particular, if there was no correlation among species in their abundances among sites, a parametric bootstrap likelihood ratio test (LRT) gave the best power. When there was correlation among species in their abundances, a conditional parametric LRT had correct type I errors but had lower power.

4. There is no overall best method for identifying trait–environment associations. For the simple task of testing associations between single environmental variables and single traits, the weighted correlations with permutation tests all had good type I error control, and their ease of implementation is an advantage. For the more complex task of multivariate analyses and model fitting, and when high statistical power is needed, we recommend MLM2 (Jamil et al., 2013). However, care must be taken to ensure against inflated type I errors for both weighted correlations and MLM2. Because CWMr exhibited highly inflated type I error rates, it should always be avoided.

1 | INTRODUCTION

Analysing plant functional traits is an increasingly popular approach for understanding plant community assembly (Cornwell & Ackerly, 2015; Funk et al., 2017). Functional traits are relatively easy-to-measure characteristics of plants representing key life-history processes that are difficult to measure directly (Lavorel & Garnier, 2002; Reich, 2014). Functional traits may elucidate mechanisms by which community composition responds to environmental gradients and thus provide inference beyond species-focused investigations of community assembly processes (Funk et al., 2017; Lavorel & Garnier, 2002). However, testing how community-level functional traits vary in response to ecological gradients poses a statistical challenge, because environmental gradients are measured at the site (or plot) level, whereas functional traits are measured at the species level. This means that values of functional traits for many species must be integrated and compared to single environmental variable values. Several statistical approaches to this challenge have been developed, but there has been no comprehensive analysis of the qualitatively different methods to compare their statistical strengths and weaknesses. In particular, methods should be compared for standard statistical properties such as type I error control and statistical power.

One of the most common approaches for analysing trait–environment relationships uses the community-weighted mean (Lavorel et al., 2008; Ricotta & Moretti, 2011), which is the average of trait values for plants at each site weighted by species abundance. The community-weighted mean is then regressed upon environmental variables among sites; thus, we will call this approach CWMr. A recent "trait-based ecology" review highlights the common use of the CWMr (Funk et al., 2017). Indeed, we found 306 papers in the peer-reviewed literature (Web of Knowledge, 25 October 2018, papers that contain the term "community-weighted mean*" as topics). One appeal of CWMr is that trait–environment relationships can be analysed using simple regression. However, CWMr reduces a large amount of data to a single trait value at each site, which raises concerns because multiple sites contain the same species (ter Braak, Peres-Neto, & Dray, 2017). Therefore, CWMr values for different sites along the gradient are not independent, and the loss of species-level information makes it impossible to account for this nonindependence in statistical analyses. Typical CWMr analyses ignore this nonindependence, which can cause problems with type I error control (Ives & Zhu, 2011).

A second approach involves computing correlations between traits and environmental variables weighted by the abundances of species within sites, and testing the significance of these correlations using permutation tests (Peres-Neto, Dray, & ter Braak, 2017).

Different weightings of traits and environmental variables lead to different measures of trait–environment relationships; we will refer to them here collectively as weighted correlation methods. For example, community-weighted mean correlations (CWMc) can be performed by computing the correlation between mean trait values in a site weighted by species densities and the environmental variables at the site; statistical tests under the null hypotheses that either traits or environmental variables do not affect community composition can be performed by permuting either species among sites or sites among species. Peres-Neto et al. (2017) show the mathematical relationships among three different weighted correlation metrics: CWMc, species niche centroids (SNCc; Cavender-Bares, Kitajima, & Bazzaz, 2004), and fourth-corner correlations (FCCc; Dray & Legendre, 2008). They also compare the weighted correlation methods, concluding that the fourth-corner method has greater precision and statistical power than the others. To avoid type I errors, Peres-Neto et al. (2017) follow ter Braak, Cormont, and Dray (2012) and perform two permutation tests (species-based and sites-based), with the overall test based on the larger of the two resulting *p*-values. Because this approach uses the least powerful permutation test, it might be expected to have low overall power.

A third approach to analysing trait–environment associations uses statistical multilevel models (MLMs) that attempt to describe the full structure of the data (Brown et al., 2014; Cormont, Vos, Van Turnhout, Foppen, & ter Braak, 2011; Jamil, Ozinga, Kleyer, and ter Braak, 2013; Pollock, Morris, & Vesk, 2012; Warton, Shipley, & Hastie, 2015). For MLMs, the occurrence or abundance of each of n species in each of m sites is used as the response variable. Each species-site combination is assigned a species-level trait and a site-level environmental variable. The test for the association between functional traits and environmental gradients involves the trait-by-environment interaction coefficient. Because the $n \times m$ species-site data points are not independent (because species occur at multiple sites), the MLMs include effects for species and/or site to allow species to have unique responses to the environmental gradients. Thus, unlike CWMr, MLMs explicitly account for some types of nonindependence of data among sites. Models can be constructed to include different possible patterns in the data. For example, it is possible to include an effect of a trait (such as fecundity) on the overall abundance of a species even if the trait does not affect a species response to the environmental variables. Similarly, a model could include an environmental variable (such as water availability) that affects the number and abundance of species in a site. The greater detail that can be built into statistical tests should give greater statistical power to detect trait–environment associations. Nonetheless, standard statistical approaches for hypothesis testing, while accounting for

correlations caused by species occurring in different sites, generally do not account for more-complex correlation structures (Brown et al., 2014). For example, phylogenetically related species may be more likely to occur in the site due to factors that are not explained by the trait values under investigation, and this can lead to inflated type I errors (Li & Ives, 2017).

In this study, we compare the performance of three general approaches to analysing trait–environment relationships: CWMr, weighted correlations, and model-based methods. We consider five weighted correlations compared by Peres-Neto et al. (2017): community-weighted means (CWMc), weighted community-weighted means (wCWMc), unweighted and weighted species niche centroids (SNCc and wSNCc), and the Chessel fourth-corner (FCCc) correlations. Note that CWMr and CWMc differ, because CWMr performs statistical tests using regression, whereas CWMc uses permutation tests. For model-based methods, we include two MLMs: MLM1 that was proposed by Pollock et al. (2012), and MLM2 that includes a fixed effect for traits (Jamil et al., 2013). For inference from the MLMs, we use standard Wald tests, two parametric bootstrap Likelihood Ratio Tests, and the GLM approach of Brown et al. (2014). We first apply each method to a plant community dataset, and then use a simulation study to investigate type I errors and statistical power. In the simulation study, we assume that there are no species-species correlations, or site-site correlations, in the residual errors. To investigate the possible effects of species-species and site-site correlations, we perform a randomization study that is designed to maintain the species-species or site-site correlations in the plant community dataset. Our overall goal is to assess which methods provide statistically robust and powerful tests of trait–environment associations.

2 | MATERIALS AND METHODS

2.1 | Dataset

To test the ability of the different methods to find trait–environment associations, we used data from a revisit to the low-elevation, non-serpentine of Robert Whittaker's historical plant community study sites in the Siskiyou Mountains of Southwest Oregon (Damschen, Harrison, & Grace, 2010; Whittaker, 1960). Vegetation consisted primarily of forest dominated by Douglas fir *Pseudotsuga menziesii*, with an assortment of conifer and evergreen hardwood trees and a diverse herbaceous understorey. Community surveys were conducted in the summer of 2007 following Whittaker's original sampling methods (Damschen et al., 2010; Whittaker, 1960). Whittaker chose sites to represent the full range of topography, with sites distributed across each of ten topographic moisture levels. A single 0.1 ha study plot was established at each study site, and 25 1 × 1-m quadrats were established along a 50-m transect running down the centre of each plot. Species abundances were calculated from the number of 100 quadrat corners each species intercepted. For our analyses, we included only plants that intercepted at least one quadrat corner so that all plants had an abundance ranking.

For the example analysis, we chose one functional trait, leaf carbon-to-nitrogen ratio (C:N), and one environmental variable, Whittaker's topographic moisture gradient (TMG). The C:N ratio is often considered a surrogate of competitive ability; plants with lower C:N may grow faster but have lower stress tolerance (Cornelissen et al., 2003; Poorter & Bongers, 2006). The topographic moisture gradient ranks each site on a scale from 1 to 10, where sites on more mesic, north-facing slopes receive lower numbers, and sites on warmer, south-facing slopes receive higher numbers. We used a standardized version of the topographic moisture gradient following Damschen et al. (2010). Ecologically, we expected that leaf C:N ratios to have a greater effect on species abundances in sites with high TMG scores, that is, in drier and warmer sites, leading to a positive trait–environment association.

2.2 | Community-weighted mean regression (CWMr)

We calculated the community-weighted means for C:N at each site by multiplying C:N values for the species that occur at the site by their abundances in that site, adding these values, and then dividing by the sum of species abundance values (Lavorel et al., 2008). Community-weighted means were then regressed against the environmental variable (TMG) for each site.

2.3 | Weighted correlation methods (CWMc, wCWMc, SNCc, wSNCc, FCCc)

Peres-Neto et al. (2017) give an excellent summary of weighted correlation metrics, and we give a brief summary in Supporting Information Appendix S1. Significance tests for the relationship between trait and environment values were performed using permutation tests following ter Braak et al. (2012). Two permutations were performed, giving two *p*-values. The first permutes trait values among species, thereby testing the hypothesis that the correlation (measured by CWMc, wCWMc, SNCc, wSNCc, or FCCc) is independent of the distribution of traits among species. The second permutes environment values among sites, thereby testing the hypothesis that the correlation is independent of the environmental gradient. Relying on either one of these tests alone leads to inflated type I errors (false positives), and therefore ter Braak et al. (2012) recommend computing both *p*-values and selecting the larger of the two. Thus, a significant correlation is only identified if the correlation depends significantly on both the distribution of trait values among species and the distribution of environment values among sites. We performed the weighted correlation analyses using the R code provided in Peres-Neto et al. (2017).

2.4 | Multilevel model approaches (MLM1, MLM2)

We used two multilevel models. MLM1 was proposed by Pollock et al. (2012) which we formulated for binomial data, with the abundance of each species taking a value from 0 to 100 in each site.

MLM1 is a logit-normal binomial generalized linear mixed model (GLMM) with the form

$$Y_i = \text{binomial}(100, p)$$

$$\text{logit}(p) = \alpha + a_{\text{spp}[i]} + \beta_{12}\text{env}_{\text{site}[i]} \times \text{trait}_{\text{spp}[i]} + (\beta_1 + c_{\text{spp}[i]})\text{env}_{\text{site}[i]} + e_i$$

$$a, c \sim \text{Gaussian}(0, \sigma_a^2, \sigma_c^2, \rho_{ac}) \tag{1}$$

$$e \sim \text{Gaussian}(0, \sigma_e^2)$$

Here, Y_i (Y_i = 0, ..., 100) is the observed abundance for each of the $i = n \times m$ species-site data rows. Following the convention of multilevel models (Gelman & Hill, 2007), the functions spp[i] and site[i] map row i onto the corresponding species and sites, so that $\text{trait}_{\text{spp}[i]}$ gives the trait (C:N) of the species in row i, and $\text{env}_{\text{site}[i]}$ gives the TMG score. The fixed effect α gives the overall average abundance of species among sites, and the fixed effect β_1 gives the mean response of species to the environmental gradient. Random effect $a_{\text{spp}[i]}$ allows different species to have different overall abundance, and random effect $c_{\text{spp}[i]}$ allows different species to have different responses to the TMG; $a_{\text{spp}[i]}$ and $c_{\text{spp}[i]}$ have means zero and variances σ_a^2 and σ_c^2, with ρ_{ac} denoting the correlation between them. Random effect e_i gives observation-level variance σ_e^2 which accounts for greater-than-binomial variance in the observations; this term is generally required in GLMMs for discrete distributions such as the binomial distribution (except in the case of only two outcomes, 0 and 1) and the Poisson distribution to allow for overdispersion (Harrison, 2014). The trait–environment association is given by the fixed effect β_{12}, and this is the target of statistical testing.

MLM2 is similar to MLM1 but follows Jamil et al. (2013) by including an additional fixed effect β_2 for traits on mean species abundance and an additional random effect for site:

$$\text{logit}(p) = \alpha + a_{\text{spp}[i]} + b_{\text{site}[i]} + \beta_{12}\text{env}_{\text{site}[i]} \times \text{trait}_{\text{spp}[i]}$$
$$+ (\beta_1 + c_{\text{spp}[i]})\text{env}_{\text{site}[i]} + \beta_2\text{trait}_{\text{spp}[i]} + e_i \tag{2}$$
$$b \sim \text{Gaussian}(0, \sigma_b^2)$$

These additional terms in MLM2 force less of the variation in species abundance onto the trait-by-environment interaction term β_{12} compared to MLM1. Specifically, the additional fixed effect β_2 allows traits to affect species abundances without doing so through the interaction with the specific environmental gradient considered in the equation. The random effect b allows sites to differ in mean species abundance independently from the environmental gradient. Both MLM1 and MLM2 can be fit as univariate models (one trait and one environmental variable in a model) or multivariate models (one or more traits or environmental variables). Here we only present univariate analyses, because the weighted correlation methods are restricted to the univariate case.

The models were fit using glmer in the R package LME4 (Bates, Maechler, Bolker, & Walker, 2015). For testing the hypothesis that $\beta_{12} = 0$, we used the approximate p-values from asymptotic Wald tests given by LME4; we refer to these tests as MLM1(Wald) and MLM2(Wald). We found, however, that these standard p-values were often too large,

leading to loss of statistical power (see Section 3). Therefore, we also calculated p-values using a parametric bootstrap likelihood ratio test (LRT) in which the distribution of the log likelihood ratio (LLR) is approximated using a parametric bootstrap rather than the standard chi-square approximation. Here, we use the term "parametric bootstrap" to refer to analyses in which simulated datasets are fit using the same model that was used for the simulations; this procedure is a parametric bootstrap, as opposed to a nonparametric bootstrap, because datasets are constructed by simulations under the parametric assumptions of the model, rather than by resampling residuals.

For the parametric bootstrap LRT, we performed the following steps (Ives, 2018): (a) fit the reduced model under the null hypothesis ($\beta_{12} = 0$); (b) simulate 2,000 datasets under the null hypothesis; (c) for each simulated dataset refit the full model (including β_{12}) and reduced model (with $\beta_{12} = 0$) to calculate the LLR; and (iv) use the resulting distribution of 2,000 LLRs as the distribution of test statistic under the null hypothesis. The p-value is given by the fraction of bootstrapped values of LLR that exceed the observed value from the data. These parametric bootstraps can be performed by simulating values for the random effects, which treats species as independent; we refer to this as MLM2(boot). We also performed a parametric bootstrap, MLM(boot.sp), in which the values for the random effects $a_{\text{spp}[i]}$ and $c_{\text{spp}[i]}$ were fixed at their conditional expectations from the fitted model. This is heuristically like treating $a_{\text{spp}[i]}$ and $c_{\text{spp}[i]}$ as if they were fixed-effect coefficients, so that each species has a factor that determines its overall abundance ($a_{\text{spp}[i]}$) and each species has a fixed slope in abundance against the environmental variable ($c_{\text{spp}[i]}$). Thus, MLM2(boot.sp) preserves any correlations among species in their relative abundances ($a_{\text{spp}[i]}$) and responses to the environmental variable ($c_{\text{spp}[i]}$). This approach attempts to overcome the general problem identified by Li and Ives (2017) for the specific case of phylogenetic correlations; they showed that phylogenetic correlations among species contained within $a_{\text{spp}[i]}$ or $c_{\text{spp}[i]}$ (i.e., related species showing similar environmental responses that are not explained by their trait values) will cause inflated type 1 errors if not properly accounted for. Because MLM2 in general performed better than MLM1, we only performed parametric bootstrap LRTs for MLM2.

We used a second method in addition to MLM2(boot.sp) that accommodates correlated responses among species. Brown et al. (2014) use a GLM to fit a model equivalent to MLM2 but treating all random effects as fixed effects, and then performed statistical tests by resampling the residuals in such a way that the correlations between species are preserved. We performed this analysis using the anova.traitglm function with PIT-trap resampling (Warton, Thibaut, & Wang, 2017) in the R statistical package MVABUND (Warton et al., 2015). We refer to this as MLM2(glm).

2.5 | Power simulations

We performed a simulation study to test the statistical power of the different methods. Our goal was to use simulations modelled after the data. Therefore, we simulated data using MLM2 with parameters

fit to the dataset for C:N and TMG. In the simulations, we assumed that trait and environmental values were normally distributed among species and sites respectively; that all coefficients for fixed effects were those from the fitted model; and that all random effects were picked from Gaussian distributions with variances from the fitted model. We selected MLM2 over MLM1 for the simulations, because MLM2 gave a better fit than MLM1 (logLik(MLM2) = −3163.8, logLik(MLM1) = −3196.8, (χ^2_2 = 66, $p \ll 0.0001$).

To dissect the consequences of different patterns in the data on type I errors and power, we considered four simulation scenarios. In scenario (i), species had the same mean abundance across sites (Equation 2, $\sigma_c = \sigma_a = 0$) and there was no dependence on their trait values ($\beta_2 = 0$). Scenario (ii) allows species abundances to depend on their trait values ($\beta_2 \neq 0$), whereas scenario (iii) allows variation among species that does not depend on their trait values ($\sigma_a > 0$, $\sigma_c > 0$). Finally, scenario (iv) is the full MLM2 ($\beta_2 \neq 0$, $\sigma_a > 0$, $\sigma_c > 0$). We chose these simulation scenarios to generate different types of variation among species in their abundances among sites to determine what types of variation could lead to poor statistical tests of the trait–environment association. For each simulation scenario, we fit MLM2 to the dataset under the assumptions (i)–(iv) to obtain parameter values (Table 1). Due to the computation time required for MLM2(glm), we simulated datasets with 20 species distributed among 15 sites. Occasionally datasets were produced where one or more species did not occur in any sites; when this happened, we set an abundance of 1 (out of 100) in the site that had the highest predicted probability of containing that species.

For each simulation dataset, we applied 11 methods (CWMr, MLM1(Wald), MLM2(Wald), MLM2(boot), MLM2(boot.sp), MLM2(glm), CWMc, wCWMc, SNCc, wSNCc, FCCc) and recorded the proportion of simulated datasets for which the null hypothesis (no trait–environment association) was rejected at the alpha-significance level of 0.05. Type I errors are given for the case in which β_{12} is set to 0 when simulating the data. To investigate the statistical power of the methods, we set $\beta_{12} = 0$, 0.2, ..., 1.4; provided type I error control is adequate (i.e., 5% of the

simulated datasets are rejected when $\beta_{12} = 0$), a greater proportion of datasets for which the null hypothesis was rejected when $\beta_{12} > 0$ implies greater power. For MLM2(boot), MLM2(boot.sp) and MLM(glm), we performed only 19 bootstraps, although this should not bias the results because p-values are uniformly distributed over the interval from 0 to 1.

All statistical and simulation analyses were performed in R version 3.3.2 (R Core Team, 2016). Example R code including a function for performing the parametric bootstraps is available in Supporting Information Appendix S1, and code for analysing the dataset with all methods is in Supporting Information Appendix S3.

2.6 | Randomizations

The simulation study did not include possible correlations among species in their trait values ($a_{\mathrm{spp}[i]}$) and response to the environmental variable ($c_{\mathrm{spp}[i]}$), or possible correlations among sites in the abundances of species they contain ($b_{\mathrm{site}[i]}$). To incorporate these possible correlations, we performed two randomization schemes of the data, focusing on type I error control. These schemes involved randomizing traits among species (while preserving possible correlations among sites) and randomizing environmental variables among sites (while preserving possible correlations among species); 400 randomizations were performed. These are the same randomizations as used for the permutation tests of the weighted correlations (Peres-Neto et al., 2017).

3 | RESULTS

3.1 | Dataset

For the Siskiyou Mountains dataset, tests of association between the trait (C:N ratio) and environmental factor (TMG) using CWMr was highly significant ($p = 0.0018$; Table 2). In contrast, of the weighted correlations only SNCc was marginally significant. MLM1

TABLE 1 Summary of MLM2 (Equations 1 and 2) regressing species abundances on carbon-to-nitrogen ratio (CN) and the topographic moisture gradient (TMG)

Fixed effects	Coefficient	SE	z-value	p-value (Wald)
(Intercept)	−7.438	0.327	−22.741	<0.001***
TMG	−0.694	0.173	−4.009	<0.001***
CN	1.083	0.308	3.513	<0.001***
TMG:CN	0.226	0.133	1.704	0.088
Random effects				**Variance**
obs (Intercept)				6.371
species (Intercept)				6.233
TMG				0.625
site (Intercept)				0.540

***$p < 0.001$.

TABLE 2 p-values for tests of whether a trait (C:N ratio) predicts species abundances along an environmental gradient (TMG) for 11 statistical methods

Method	p-value
CWMr	**0.0018**
CWMc	0.14
wCWMc	0.13
SNCc	**0.046**
wSNCc	0.059
FCCc	0.059
MLM1(Wald)	0.47
MLM2(Wald)	0.088
MLM2(boot)	**0.0125**
MLM2(boot.sp)	0.12
MLM2(glm)	0.27

p-values < 0.05 are bolded.

and MLM2 did not give significant associations, except for the case of MLM2(boot). These contrasting results among methods raise the question of which is correct. To address this, we assessed the overall properties of the different methods, first using simulations.

3.2 | Power simulations

We simulated data using MLM2 with parameter values fit to the C:N and TMG data. Simulation (i) was the simplest, in which species had the same mean abundances (Equation 3, $\beta_2 = 0$, $\sigma_a = 0$) and the same responses to the environment ($\sigma_c = 0$). In this case, no method had inflated type I errors, and MLM2(boot) had better power than the other methods, followed by MLM2(glm) and MLM2(boot.sp; Figure 1, case i). MLM2(Wald) had power on par with the weighted correlation methods. The low power for MLM2(Wald) is the result of bias in the estimates of β_{12}. As described in more detail in Supporting Information Appendix S4, this bias only occurs in the binomial model when there is overdispersion ($\sigma^2_e > 0$), and for non-overdispersed or continuous data, the power of MLM2(Wald) is the same as MLM2(boot).

Simulation (ii) added variation in the mean abundances of species that depended on their trait values ($\beta_2 > 0$). In this case, MLM1(Wald) had inflated type I errors (Figure 1), because MLM1 does not include a term for a possible effect of traits on the mean abundance of species. Therefore, if there is an effect of traits on mean abundances (Equation 2, $\beta_2 \neq 0$), the resulting variance is forced on the environment-by-trait interaction (β_{12}) because this is the only term in the model that contains a trait effect. In principle, this variation could be absorbed into the random effect a for differences among species in mean abundance that do not depend on trait values, but in practice

it was not. MLM2(glm) also had inflated type I errors, which is consistent with results of Warton et al. (2017) who showed inflated type I errors when the number of species exceeds the number of sites (as in our simulations); ter Braak et al. (2017) also show cases in which this approach gives inflated type I errors. The weighted correlation methods had good type I error control but relatively low power.

Simulation (iii) removed the affect of trait values on mean abundances ($\beta_2 = 0$), but added variation in species overall abundance that is unrelated to trait values ($\sigma_a > 0$) and that depends on the environment in sites ($\sigma_c > 0$). In this case, CWMr and MLM2(glm) had inflated type I errors (Figure 1), possibly because the site-to-site variation in species abundances was attributed to environmental differences among sites, even though in the simulations site-to-site variation was random. MLM1(Wald) and MLM2(Wald) had acceptable type I error rates, and MLM2(boot) had the best power. Although the weighted correlations had correct type I error rates, they had relatively low power, as did MLM2(boot.sp).

Finally, simulation (iv) most closely follows the data, with all model parameters estimated from the data using MLM2 under the assumption that $\beta_{12} = 0$. The performance of the methods was similar to simulation (iii). This similarity suggests that variation among species in their overall abundance ($\sigma_a > 0$) and trait-independent response to the environmental variable ($\sigma_c > 0$) create the greatest challenge for the methods.

3.3 | Randomizations

The simulations restricted the types of possible patterns in the data to those specified in MLM2: the residual errors among species, and
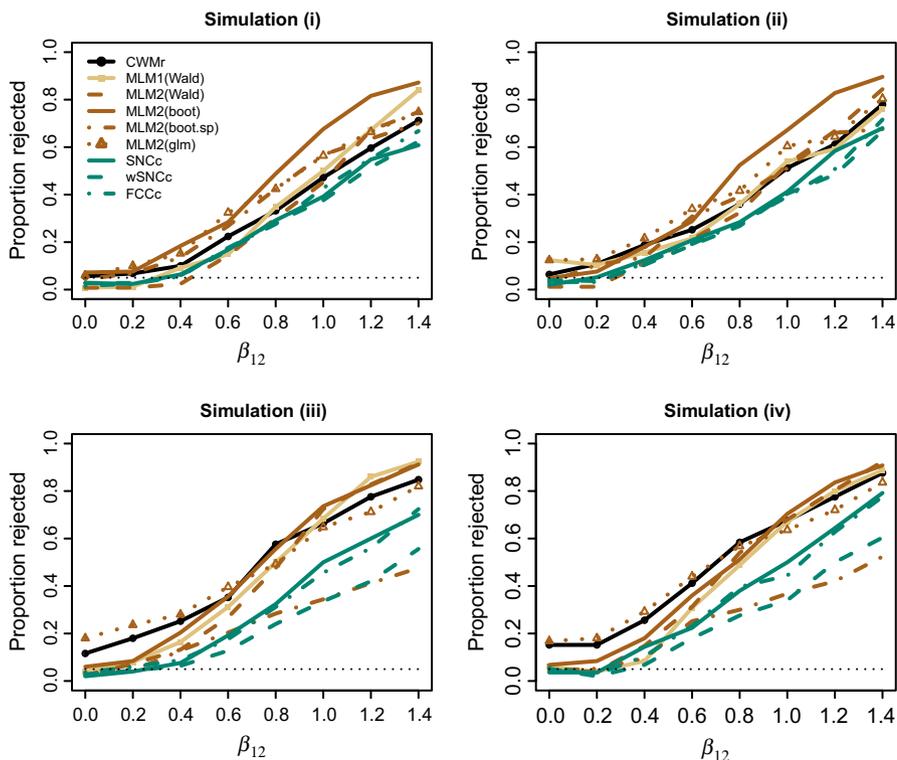

FIGURE 1 Results from simulation studies using MLM2 under scenarios: (i) $\beta_2 = 0$, (ii) $\sigma_a = \sigma_c = 0$, (iii) $\beta_2 = \sigma_a = \sigma_c = 0$, and (iv) $\beta_2 \neq 0$, $\sigma_a \neq 0$, and $\sigma_c \neq 0$. MLM2 was first fit to the dataset under the four scenarios with $\beta_{12} = 0$. The fitted MLM2 was then used to simulate 250 datasets at each of $\beta_{12} = 0, 0.2,..., 1.4$ for 20 species distributed among 15 sites. CWMr, MLM1(Wald), MLM2(Wald), MLM2(boot), MLM2(boot.sp), MLM2(glm), SNCc, wSNCc, and FCCc were fit to each simulated dataset, and the proportion of the 250 datasets for which the null hypothesis of no trait–environment association was rejected is plotted for each scenario. Results for CWMc and wCWMc are not reported, because they are very similar to the results for SNCc and wSNCc respectively

the residual errors among sites, are independent. The possibility of nonindependence among species is the motivation for the resampling strategy used in MLM2(glm) (Brown et al., 2014; Warton et al., 2017) in which species abundance residuals are resampled among sites without disrupting the possible covariances among species. A similar approach for MLM2(boot.sp) implements parametric bootstraps while fixing the random-effects values for species; this leaves the covariances among species intact when simulating residuals.

The Siskiyou Mountains dataset shows greater-than-expected residual error correlations among species. Specifically, the pairwise Kendall correlation coefficients between species in residual errors e_i from MLM2 (Equation 2) show much higher-than-expected values, as do the Kendall correlation coefficients between site residuals (Figure 2). Randomizing the abundance of each species among sites removes the high correlations for species among sites, and greatly reduces the correlations among sites in the abundances of species they contain (Figure 2). Note, however, that randomizations of traits among species will not change these high correlations. Plots of the Kendall correlation coefficients between species in the residuals from MLM2 provide an effective way to detect correlations among species that are not accounted for in the model. It is possible to test for greater-than-expected Kendall correlation coefficients using a Kolmogorov–Smirnov test against their approximate distribution under the null hypothesis of independence; this is given by the distribution of Pearson correlations and is related to a Student's t distribution. For the data (Figure 2), the Kolmogorov–Smirnov test gives $p \ll 10^{-10}$. There is, however, no way of determining diagnostically how strong correlations among species must be to affect the

statistical tests for trait-by-environment interactions. Therefore, if they appear strong, methods that account for these correlations should be used. R code for producing figure 2 and performing Kolmogorov–Smirnov tests are given in Supporting Information Appendix S3.

The randomizations of the dataset are designed to preserve the correlations among species while randomizing the trait values among species or randomizing the environmental variable among sites. Note that the randomizations only change the characteristics of species (trait randomizations) and sites (environment randomizations), not their distributions. Therefore, the randomization study investigates how the nonrandom distribution of species among sites might affect the statistical detection of trait–environment associations when the trait (trait randomizations) or environmental variable (environment randomizations) have no effect on their distributions. In the trait randomization, CWMr had badly inflated type I errors (Table 3). In contrast, all the weighted correlations had good type I error control. This is because the permutation procedure they use mirrors the two randomizations of species and sites: species are permuted to obtain one p-value and sites are permuted to obtain a second p-value, and the highest p-value is used for the overall test.

None of the model-based methods had inflated type I errors when trait values were randomized among species. However, several had inflated type I errors when the environmental variable was randomized among sites (Table 3). This implicates the higher-than-expected positive correlations among species (Figure 2) in driving inflated type I errors, because these positive correlations are preserved in the randomizations. The inflated type I errors are particularly bad for
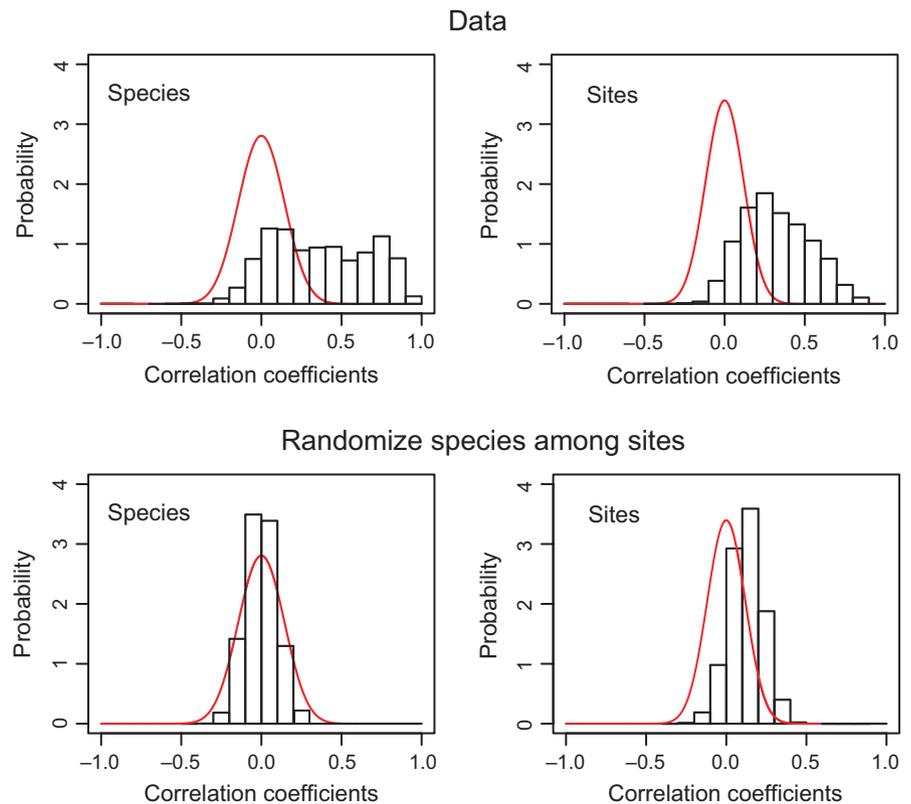

FIGURE 2 Correlations between residual errors from MLM2 (equations 1 and 2, e_i). For each pair of species or sites, Kendall correlation coefficients were computed. The panels in the top row are for the data, and the panels in the bottom row are for an example randomization of species (and their trait values) among sites. The red lines give the t-distributions that are expected under the assumption that the residuals between species or sites are independent. A Kolmogov–Smirnov test showed strong statistical departures of species and site correlations from random ($p \ll 10^{-10}$)

TABLE 3 Proportion of 400 randomization datasets for which each of the 11 methods rejected the null hypothesis at a nominal alpha level of 0.05. Proportions ≥0.07 are in boldface

	Trait	Environment
CWMr	**0.405**	0.040
CWMc	0.0425	0.0125
wCWMc	0.0475	0.0175
SNCc	0.02	0.035
wSNCc	0.035	0.0325
FCCc	0.0375	0.0375
MLM1(Wald)	0.01	0.0575
MLM2(Wald)	0.0025	**0.07**
MLM2(boot)	0.0625	**0.16**
MLM2(boot.sp)	0.0025	0.065
MLM2(glm)	0.0025	**0.13**

MLM2(boot), which in the simulation study had the highest power (Figure 1). Type I errors were controlled by MLM2(boot.sp). This is not a wholly satisfactory solution, however, because MLM2(boot.sp) rejected too few of the trait randomizations, reflecting the lower power of MLM2(boot.sp) found in the simulations (Figure 1). The re-sampling used in MLM2(glm) was designed to account for noninde-pendence among species, although the inflated type I errors in the environment randomizations show that this was not successful.

4 | DISCUSSION

If there are multiple statistical methods that give different results when applied to the same dataset, a researcher might ask which one is correct (Table 2). Unfortunately, for complicated questions such as the role of trait–environment associations in community assembly, there is no "correct" method or "correct" result. The different meth-ods we have considered will be sensitive to different patterns in the data: this means that they test different hypotheses about the role of traits in explaining the distribution of species among sites. Instead of asking which method is correct, it is necessary to ask which method has the best statistical properties for data similar to the real data.

The three different statistical approaches for analysing functional trait responses to environmental gradients—community-weighted means, weighted correlations, and multilevel models—varied sub-stantially in performance (Table 4). Type I error control is an essential statistical property, because if type I errors are inflated, then a re-searcher might conclude that a relationship is significant when in fact it is not. Community-weighted mean regression (CWMr) had poor type I error control in the simulations (Figure 1); a detailed explo-ration using randomizations (Supporting Information Appendix S5) showed that this poor type I error control is caused because CWMr values are not independent among sites. This lack of independence in CWMr values is the result of species occurring in more than one site. If, for example, a species is common (and therefore weights heavily

in CWMr) at several sites that occur at one end of the environmental gradient, then whatever trait value this species has will exert high influence on the slope of CWMr. Because CWMr does not take into account the lack of independence among CWMr values, it will as-cribe greater significance to the regression slope than appropriate. Type I error control was worse when the mean abundance of species among sites varied in a way that was not associated with their trait values (Figure 1, simulations iii and iv). This situation is likely common in community data, arguing against using CWMr.

For the weighted correlation metrics, the dual permutation tests of traits among species or environment values among sites (ter Braak et al., 2012) gave good type I error control under the scenarios we investigated. This comes at a cost, however, of lower power than is possible with the model-based approaches (provided the models have good type I error control). Peres-Neto et al. (2017) used a sim-ulation study to compare the statistical power of the five weighted correlations and found that FCCc had greater power than the oth-ers, whereas in our simulations FCCc was not superior; this could be because in our dataset and simulations, the number of species exceeded the number of sites (S. Dray, pers. comm.). This contrast illustrates that power likely varies depending on the underlying structure of a dataset. The same conclusion is also likely true in com-paring power among model-based methods, or comparing between weighted correlations and model-based methods. Therefore, the de-tails of the comparisons among methods that we present could be different for other simulation/randomization investigations based around other datasets. In particular, when these are strong main ef-fects of trait and environmental variables weighted correlations can give inflated type I errors (Supporting Information Appendix S6).

Of the two multilevel models, MLM2 was better than MLM1. MLM1 (Pollock et al., 2012) exhibited inflated type I error rates when species mean abundances across all sites were not independent of their trait values (Figure 1, simulation ii). Because MLM1 does not contain traits as fixed (or random) effects that could account for this variation in overall mean abundances, it attributes the variation to the only term in the model that contains trait values, the trait-by-environment interaction term.

MLM2 (Jamil et al., 2013) accounts for traits affecting species mean abundances even if there is no trait-by-environment interac-tion, yet two complications must be addressed. First, we found that for overdispersed data, the estimates of the trait-by-environment in-teraction coefficient were biased (Supporting Information Appendix S4), and this reduced power of MLM2(Wald) (Figure 1). This can be corrected with the parametric bootstrap, MLM2(boot). Second, lack of independence among species in the residual errors (Figure 2) can inflate type I errors (Table 3, environmental randomizations). The resampling procedure used in MLM2(glm) was designed to account for nonindependence among species in residual errors (Brown et al., 2014), although it is less successful when the number of species ex-ceeds the number of sites (Warton et al., 2017). The parametric bootstrap in which the random effects for species were fixed at their estimated values, MLM2(boot.sp) did have good type I error control despite nonrandom distributions of species with respect to each other

TABLE 4 Advantages and disadvantages of the three broad approaches to analyse trait–environment associations in community data

Method	Advantages	Disadvantages
Community-weighted mean regression	1. None	1. Inflated type I error rates yield unreliable results
Weighted correlations	1. Give correct type I error	1. The primary information given is only *p*-values
	2. Simple and fast	2. Limited to simple analyses
Model-based approaches	1. Can give highest power	1. Complicated and computationally intensive
	2. Give most information about the data	2. Require diagnostics and possibly parametric bootstraps
	3. Flexible for multivariate analyses	

when there was no relationship between the species abundances and the environmental characteristics of sites (Table 3, environmental randomizations), although it had lower power when species distributions were independent of their traits (Table 3, trait randomizations).

These results were derived from analyses of a single dataset and simulations based around it. The dataset is challenging, because the response variable is binomial (rather than continuous) and has high overdispersion (greater-than-binomial variance), and because there are covariances among the abundances of species that are not explained by trait values (Figure 2). The dataset therefore exposes the difficulties that can be encountered in analyses of trait–environment associations, as we have shown. Nonetheless, our results about the magnitudes of the difficulties, or the performances of the different approaches, will not necessarily generalize to all datasets. Therefore, each dataset must be analysed with caution and knowledge of the difficulties that can be encountered.

4.1 | Suggestions for practitioners

Despite the preceding caveat that each dataset must be investigated cautiously in its own right, three broad generalizations are possible. First, CWMr often showed highly inflated type I errors because it treats community-weighted mean trait values as independent. Therefore, we join Peres-Neto et al. (2017) in recommending against its use. We suggest that practitioners who wish to continue using community-weighted means perform the permutation tests suggested by ter Braak et al. (2012); in other words, use CWMc rather than CWMr.

Second, all weighted correlations showed good type I error control under all of the situations we examined and therefore represent safe analyses. If the goal is only to identify strong trait–environment associations in a dataset, weighted correlations are an excellent choice. A limitation is only being able to test simple trait-by-environment associations. Also, they do not give a fitted model that can be used to explore the data. For example, fitting MLM2 to the dataset (Table 1) showed strong negative effects of species C:N values on species abundances, and strong positive effects of site TMG values, which is information not given by weighted correlation approaches. Furthermore, a fitted model can be used to confirm

the type I error rates of weighted correlations which can be inflated (Supporting Information Appendix S6).

Third, model-based methods can work very well, but their fits to the data must be examined carefully. MLM2 generally should be preferred over MLM1, because MLM2 contains a parameter for an effect of trait values on abundances independently of the environmental variable; this removes a potential source of inflated type I errors. The need for the additional terms in MLM2 for our dataset is very apparent in routine model fitting that should be conducted in any statistical analysis: MLM2 fit the data much better than MLM1 (logLik(MLM2) = −3163.8, logLik(MLM1) = −3196.8, $\chi^2_2 = 66$, $p \ll 0.0001$).

Analysis of MLM2 needs to be accompanied by appropriate diagnostics. If there is overdispersion in the data (the observation-level variance is greater than zero), then a parametric bootstrap should be considered to increase power. The residual errors must also be checked for unexpected positive correlations among species (Figure 2); we present R code for doing this in Supporting Information Appendix S3. If correlations among species are found, a parametric bootstrap test should be performed that preserves these correlations, MLM2(boot.sp); we provide a function, bootMer_LRT(), in Supporting Information Appendix S2 that performs this bootstrap. Performing the randomization tests as we did in Table 3 would also help. Overall, we advise cautious and thorough analyses.

After all our analyses of the Siskiyou Mountains dataset, is there a C:N-TMG association? Most likely not. In Table 2, two approaches that gave "significant" results, CWMr and MLM2(boot), were shown to have inflated type I errors. SNCc showed a barely significant association ($p = 0.046$), but it would be cherry picking to rely on this result. With all of our focus on *p*-values, however, we have ignored maybe the most important conclusion from the data. MLM2 shows that there is a very strong negative main effect of TMG on abundance (−0.69, Table 1), and a very strong positive main effect of C:N (1.08). Furthermore, the trait-by-environment interaction (0.23), whether it is significant or not, is not strong enough to change the sign of these main effects. Our original hypothesis was that C:N ratios should have the greatest positive effects on species abundances in dry sites with high TMG scores. The analyses do not give convincing statistical support for this. Nonetheless, they do show that C:N ratios have a

positive effect on species abundances regardless of TMG scores, and TMG has a negative effect on species abundances regardless of C:N ratios. This biological conclusion is a useful result from a model-based approach (MLM2) regardless of whether the trait-by-environment interaction is significant, illustrating how model-based approaches can be more informative than alternative methods.

AUTHORS' CONTRIBUTIONS

Authorship was determined using a first-last-author-emphasis approach; J.E.D.M. and A.R.I. led conceptualization of the project; E.I.D. provided field data; A.R.I. and J.E.D.M. led the development of the analytical approach and conducted analyses; J.E.D.M. and A.R.I. wrote the first draft of the manuscript; and all authors contributed to manuscript revisions.

DATA ACCESSIBILITY

Data are available from the Dryad Digital Repository https://doi.org/10.5061/dryad.7gj0s3b (Miller, Damschen, & Ives, 2018).

ORCID

Jesse E. D. Miller (iD) http://orcid.org/0000-0002-2316-779X

Anthony R. Ives (iD) http://orcid.org/0000-0001-9375-9523

REFERENCES

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). lme4: Linear mixed-effects models using Eigen and S3. *Journal of Statistical Software*, *67*, 1–48.

Brown, A. M., Warton, D. I., Andrew, N. R., Binns, M., Cassis, G., & Gibb, H. (2014). The fourth-corner solution - using predictive models to understand how species traits interact with the environment. *Methods in Ecology and Evolution*, *5*, 344–352. https://doi.org/10.1111/2041-210X.12163

Cavender-Bares, J., Kitajima, K., & Bazzaz, F. A. (2004). Multiple trait associations in relation to habitat differentiation among 17 Floridian oak species. *Ecological Monographs*, *74*, 635–662. https://doi.org/10.1890/03-4007

Cormont, A., Vos, C. C., Van Turnhout, C. A. M., Foppen, R. P. B., & ter Braak, C. J. F. (2011). Using life-history traits to explain bird population responses to changing weather variability. *Climate Research*, *49*, 59–71. https://doi.org/10.3354/cr01007

Cornelissen, J. H. C., Lavorel, S., Garnier, E., Díaz, S., Buchmann, N., Gurvich, D. E., ... Poorter, H. (2003). A handbook of protocols for standardised and easy measurement of plant functional traits worldwide. *Australian Journal of Botany*, *51*, 335. https://doi.org/10.1071/BT02124

Cornwell, W. K., & Ackerly, D. D. (2015). Community assembly and shifts in plant trait distributions across an environmental gradient in coastal California. *Ecological Monographs*, *79*, 109–126.

Damschen, E. I., Harrison, S., & Grace, J. B. (2010). Climate change effects on an endemic-rich edaphic flora: Resurveying Robert H. Whittaker's Siskiyou sites (Oregon, USA). *Ecology*, *91*, 3609–3619. https://doi.org/10.1890/09-1057.1

Dray, S., & Legendre, P. (2008). Testing the species traits-environment relationships : The fourth-corner problem revisited. *Ecology*, *89*, 3400–3412. https://doi.org/10.1890/08-0349.1

Funk, J. L., Larson, J. E., Ames, G. M., Butterfield, B. J., Cavender-Bares, J., Firn, J., ... Wright, J. (2017). Revisiting the Holy Grail: Using plant functional traits to understand ecological processes. *Biological Reviews*, *92*, 1156–1173. https://doi.org/10.1111/brv.12275

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multi-level/hierarchical models*. New York, NY: Cambridge University Press.

Harrison, X. A. (2014). Using observation-level random effects to model overdispersion in count data in ecology and evolution. *PeerJ*, *2*, e616. https://doi.org/10.7717/peerj.616

Ives, A. R. (2018). *Mixed and phylogenetic models: A conceptual introduction to correlated data*. leanpub.com.

Ives, A. R., & Zhu, J. (2011). Statistics for correlated data: Phylogenies, space, and time. *Ecological Applications*, *16*, 20–32.

Jamil, T., Ozinga, W. A., Kleyer, M., & ter Braak, C. J. F. (2013). Selecting traits that explain species-environment relationships: A generalized linear mixed model approach. *Journal of Vegetation Science*, *24*, 988–1000. https://doi.org/10.1111/j.1654-1103.2012.12036.x

Lavorel, S., & Garnier, E. (2002). Predicting changes in community composition and ecosystem functioning from plant traits: Revisiting the Holy Grail. *Functional Ecology*, *16*, 545–556. https://doi.org/10.1046/j.1365-2435.2002.00664.x

Lavorel, S., Grigulis, K., McIntyre, S., Williams, N. S. G., Garden, D., Dorrough, J., ... Bonis, A. (2008). Assessing functional diversity in the field - methodology matters!. *Functional Ecology*, *22*, 134–147.

Li, D., & Ives, A. R. (2017). The need to include phylogeny in trait-based analyses of community composition. *Methods in Ecology and Evolution*, *8*, 1192–1199. https://doi.org/10.1111/2041-210X.12767

Miller, J. E. D., Damschen, E. I., & Ives, A. R. (2018). Data from: Functional traits and community composition: A comparison among community-weighted means, weighted correlations, and multilevel models. *Dryad Digital Repository*. https://doi.org/10.5061/dryad.7gj0s3b

Peres-Neto, P. R., Dray, S., & ter Braak, C. J. F. (2017). Linking trait variation to the environment: Critical issues with community-weighted mean correlation resolved by the fourth-corner approach. *Ecography*, *40*, 806–816. https://doi.org/10.1111/ecog.02302

Pollock, L. J., Morris, W. K., & Vesk, P. A. (2012). The role of functional traits in species distributions revealed through a hierarchical model. *Ecography*, *35*, 716–725. https://doi.org/10.1111/j.1600-0587.2011.07085.x

Poorter, L., & Bongers, F. (2006). Leaf traits are good predictors of plant performance across 53 rain forest species. *Ecology*, *87*, 1733–1743. https://doi.org/10.1890/0012-9658(2006)87[1733:LTAGPO]2.0.CO;2

R Core Team. (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Reich, P. B. (2014). The world-wide "fast-slow" plant economics spectrum: A traits manifesto. *Journal of Ecology*, *102*, 275–301. https://doi.org/10.1111/1365-2745.12211

Ricotta, C., & Moretti, M. (2011). CWM and RAo's quadratic diversity: A unified framework for functional diversity. *Oecologia*, *167*, 181–188. https://doi.org/10.1007/s00442-011-1965-5

ter Braak, C. J., Cormont, A., & Dray, S. (2012). Improved testing of species traits–environment relationships in the fourth-corner problem CAJO. *Ecological Society of America*, *93*, 1525–1526.

ter Braak, C. J. F., Peres-Neto, P., & Dray, S. (2017). A critical issue in model-based inference for studying trait-based community assembly and a solution. *PeerJ*, *5*, e2885. https://doi.org/10.7717/peerj.2885

Warton, D. I., Shipley, B., & Hastie, T. (2015). CATS regression - a model-based approach to studying trait-based community assembly. *Methods in Ecology and Evolution*, *6*, 389–398. https://doi.org/10.1111/2041-210X.12280

Warton, D. I., Thibaut, L., & Wang, Y. A. (2017). The PIT-trap-A "model-free" bootstrap procedure for inference about regression models with discrete, multivariate responses. *PLoS ONE*, *12*, e0181790. https://doi.org/10.1371/journal.pone.0181790

Whittaker, R. (1960). Vegetation of the siskiyou mountains, Oregon and California. *Ecological Monographs*, *30*, 279–338. https://doi.org/10.2307/1943563

SUPPORTING INFORMATION

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