RESEARCH ARTICLE

New robust weighted averaging- and model-based methods for assessing trait–environment relationships

Cajo J. F. ter Braak

Biometris, Wageningen University & Research, Wageningen, the Netherlands

Abstract

1. Statistical analysis of trait–environment association is challenging owing to the lack of a common observation unit: Community-weighted mean regression (CWMr) uses site points, multilevel models focus on species points, and the fourth-corner correlation uses all species-site combinations. This situation invites the development of new methods capable of using all observation levels. To this end, new multilevel and weighted averaging-based regression methods are proposed.

2. Compared to existing methods, the new multilevel method, called MLM3, has additional site-related random effects; they represent the unknowns in the environment that interact with the trait. The new weighted averaging method combines site-level CWMr with a species-level regression of Species Niche Centroids on to the trait. Because species can vary enormously in frequency and abundance giving diversity variation among sites, the regressions are weighted by Hill’s effective number ($N_2$) of occurrences of each species and the $N_2$-diversity of a site, and are subsequently combined in a sequential test procedure known as the max test.

3. Using the test statistics of these new methods, the permutation-based max test provides strong statistical evidence for trait–environment association in a plant community dataset, where existing methods show weak evidence. In simulations, the existing multilevel model showed bias and type I error inflation, whereas MLM3 did not. Out of the weighted averaging-based regression methods, the $N_2$-weighted version best controlled the type I error rate. MLM3 was superior to the weighted averaging-based methods with up to 30% more power.

4. Both methods can be extended (a) to account for phylogeny and spatial autocorrelation and (b) to select functional traits and environmental variables from a greater set of variables.

KEYWORDS
community assembly, community-weighted means, fourth-corner approach, generalized linear mixed models, species niche centroid, trait-based ecology, weighted averaging, Whittaker Siskiyou Mountains data
Functional traits are useful in going beyond species-based understanding of community assembly processes (McGill, Enquist, Weiher, & Westoby, 2006; Shipley, 2010a). They allow predicting species’ responses to environmental and trait variation, even if the trait–environment relationship is derived from data in which the species of interest were rare or even absent (Funk et al., 2017; Lavorel & Garnier, 2002). However, the selection and testing of relevant traits and environmental variables from data is statistically challenging, because they lack a common observation unit: traits are observed on species, environmental variables on sites, and the mediating abundances on species-site combinations. The typical data format is shown in Figure 1.

A main issue in all statistical approaches to assess trait–environment association is the number of statistical units. Figure 2 contrasts three standard methods: Community-weighted mean regression (CWMr), the multilevel model, and the fourth-corner correlation. In the popular CWMr, there are as many points as there are sites ($n$) and species, white for sites and grey for species–site combinations) is proportional to the value

![Diagram](image)

**FIGURE 1** The data needed to assess trait–environment association between a single trait $t$ and a single environmental variable $e$ with $Y$ the table with abundance values of species across sites. Derived statistics, which are important in the WA-based methods, are the community weighted means, which are site-specific trait means, and the species niche centroids, which are species-specific environmental means. The size of circles (black for species, white for sites and grey for species-site combinations) is proportional to the value

(Figure 2a). In the model-based analyses proposed by Pollock, Morris, and Vesk (2012) and Jamil, Ozinga, Kleyer, and ter Braak (2013), that is the first and second multilevel models (MLM1 and MLM2) in Table 1, there are as many points as there are species ($m$). MLM1 and MLM2 focus on the variability in the species response to the environmental variable. This is illustrated in Figure 2b which plots species-specific regression coefficients against the trait. Finally, the fourth-corner correlation, a weighted correlation between trait and environmental variable, can be displayed as a plot of $n \times m$ points corresponding to all species-site combinations (Figure 2c). It is worth noting that all three methods use all the available data, but in some way or another, the points not represented are either averaged out or, in the case of the fourth-corner correlation, not yet calculated. For instance MLM1 and MLM2 allow for the prediction and simulation of abundances from the model in order to construct a graph like Figure 2a. However, the $n$ points so calculated do not play a role in the statistical assessment of the trait–environment association.

Peres-Neto, Dray, and ter Braak (2017) and Miller, Damschen, and Ives (2019) showed that CWMr with its $n$ points (Figure 2a) detects traits to be associated with environmental variables far too often in the null situation of no association (see Zeleny (2018) for a review). The generalized linear model (GLM) model proposed by Brown et al. (2014) and Warton, Shipley, and Hastie (2015), denoted as MLM2(glm) in Table 1, essentially uses $n$ points and is too liberal as well (ter Braak, Peres-Neto, & Dray, 2017; Miller et al., 2019). Of course, MLM2(glm) uses all data, but statistical inference is based on bootstrapping the residuals of $n$ sites. For the maximum entropy approach (maxent) that is related to MLM2(glm) (Warton, Shipley, et al., 2015), Shipley (2010b) proposed inference via permutation of trait values (i.e. of species). Later, ter Braak et al. (2017) argued that valid detection of trait–environment association by MLM2(glm) (i.e. a model without random effects) requires both site- and species-level tests to make up for the missing randomness. Each test results in a $p$-value and both $p$-values are then combined by taking the maximum, which is known as the max test (ter Braak, Cormont, & Dray, 2012).

At this point, it is not clear whether MLM1 and MLM2 (Figure 2b) are valid methods for detecting and selecting traits that interact with particular environmental variables. Does the statistical machinery of generalized linear mixed models (GLMM) overarching MLM1 and MLM2 prevent Type I error rate inflation? Is the scatter of the random species-specific effects around the line a sufficient basis for inference? Miller et al. (2019) showed already that MLM1 and MLM2 may give elevated Type I error rates, supporting the idea that current applications of GLMM to test for trait–environment association may not be as robust as assumed.

Miller et al. (2019) were the first to perform a simulation-based comparison of the computationally elaborate multilevel methods and the simple weighted averaging methods, such as CWMr and fourth-corner. Miller et al. (2019) introduced the term weighted correlation methods, but I prefer the term weighted averaging (WA) because of its efficiency in ecological niche models (ter Braak & Barendregt, 1986; ter Braak & Looman, 1986) and its common use.
in the reconstruction of palaeo environments (Birks, Line, Juggins, Stevenson, & ter Braak, 1990; Juggins & Birks, 2012). In the simulations by Miller et al. (2019) the site-level only methods, CWMr and MLM2(glm), had severely inflated Type I error rate, in agreement with recent results (ter Braak et al., 2017; Peres-Neto et al., 2017).

Miller et al. (2019) studied the association between the trait leaf carbon-to-nitrogen ratio (C:N) and the environmental variable topographic moisture gradient (TMG) in a subset of the Whittaker Siskiyou Mountains Revisit data (Damschen, Harrison, & Grace, 2010). In this paper, ‘Revisit data’ refer to this subset. After all their analyses, with the relatively well-behaving methods resulting in p-values just below or above .05, Miller et al. (2019) posed the question ‘is there a C:N-TMG association?’ and their answer was ‘most likely not’.

Stimulated by the paper by Miller et al. (2019), this paper proposes new WA- and model-based methods for analysing trait–environment association that increase robustness over current methods and shows that there is a strong statistical evidence for C:N-TMG association in the Revisit data. The new model-based method (named here MLM3) is a generalization of MLM2. It deals with the lack of random site effects that could be plotted against the environmental variable, so that the scatter can help assess the trait–environment relation, either visually (as in Figure 2a) or statistically. By analogy with the species-specific regression coefficients with respect to the environment variable, such site effects are site-specific regression coefficients with respect to the trait. If MLM2 is viewed as a ‘random intercept, random slope model’, then MLM3 includes two such models, one with respect to the trait and the other with respect to the environment. The new WA-based method is motivated by the observation that species typically vary enormously in frequency and abundance. As in ter Braak, Peres-Neto, and Dray (2018), it combines CWMr and the regression of the Species Niche Centroids (SNC, Figure 1) against the trait (SNCr), but then weights each species and each site in the regression in proportion to the ‘effective’ number of values averaged in the computation of the SNC and CWM. The SNC of a particular species is the mean value of the environmental variable, calculated over sites, when weighted by the relative abundance of that species in each of these sites.

The properties of the proposed new methods are investigated by simulation and applied to the Revisit data. Appendix S1 provides a tutorial in R (R Core Team, 2018) using this data.
2 | MATERIALS AND METHODS

2.1 | Dataset

The data are from the revisit to the low-elevation, non-serpentine sites of Robert Whittaker’s historic plant community study sites in the Siskiyou Mountains of Southwest Oregon (Damschen et al., 2010; Whittaker, 1960). Briefly, the data consist of the abundance of 75 species in 52 sites, calculated from the number of 100 quadrat corners per site that each species intersected (Miller et al., 2019). The environmental variable and functional trait are standardized versions of the original topographic moisture gradient (TMG) and the leaf carbon-to-nitrogen ratio (C:N), respectively (Miller et al., 2019). Note that Miller et al. (2019) were not able to find convincing statistical evidence for a C:N-TMG association in the data based on methods reviewed in this paper.

2.2 | Multilevel model approaches (MLM1, MLM2, MLM3)

This section derives the new multilevel model MLM3 and compares it with the MLM1 and MLM2 models used by Miller et al. (2019). Appendix S2 derives MLM3 from a model with unobserved trait and environmental variables that interact with the observed ones.

The new multilevel model is derived here from a GLMM that models abundances \( y_{ij} \) of \( m \) species in \( n \) sites \((i = 1,...,n; j = 1,...,m)\) as a function of the trait and environmental variable, denoted by \( t \) and \( e \) collectively, and by \( t_i \) and \( e_i \) for values of individual species and sites respectively. As is convention in GLM, \( y_{ij} \sim D(L_{ij}) \), where \( D \) denotes a statistical distribution with its link-function tailored to the kind of data being used, and \( L_{ij} \) denotes the linear predictor (Ovaskainen et al., 2017). The linear predictor is modelled as

\[
L_{ij} = a_0 + b_i^t t_i + c_i^e e_i + d_i + a_j
\]

with fixed effects in Greek, and random effects (with non-zero mean in this equation only) in Latin; \( a_0 \) denotes the overall intercept, whereas \( b_i^t \) and \( c_i^e \) denote regression coefficients with respect to the trait and environmental variable that may depend on site and species respectively. Finally, \( d_i \) represents the site-specific error, which accounts for unobserved site-level variables that influence all species in the same way and \( a_j \) accounts for differential abundance among species. If \( b_i^t = \beta_i \) and \( c_i^e = \gamma_i \), then Equation 1 is a main effects only model, but in general the coefficients \( b_i^t \) and \( c_i^e \) may be site- and species-specific. Trait-environment association is introduced by submodels that are linear in \( e \) and \( t \),

\[
b_i^t = \beta_0 + \beta_i^e e_i + b_i \tag{2a}
\]

\[
c_i^e = \gamma_0 + \gamma_i^e t_i + c_i \tag{2b}
\]

where \( \beta_0 \) and \( \gamma_0 \) are intercepts and \( \beta_i^e \) and \( \gamma_i^e \) are slopes with respect to \( (e_i) \) and \( (t_i) \) respectively; \( b_i \) and \( c_i \) are the errors in these submodels, taken to be independent with variances \( \sigma^2_{\beta} \) and \( \sigma^2_{\gamma} \) respectively. It should be noted that this can be generalized to account for spatial and phylogenetic (auto)correlation (Li & Ives, 2017; Ovaskainen et al., 2017). Figure 3 shows these models for the Revisit data. Insertion of these submodels into Equation 1 gives the MLM3 model:

\[
L_{ij} = a_0 + d_i + a_j + (\beta_0 + b_i) t_i + (\gamma_0 + c_i) e_i + \beta_i^e t_i e_i + \gamma_i^e t_i e_i \tag{3}
\]

with \( \beta_i^e = \beta_i^e + \gamma_i^e \) the usual trait-environment interaction coefficient. Table 1 describes the purpose of each term. There is no \( t \cdot e \) association under the null hypothesis \( H_0: \beta_i^e = 0 \), while under the alternative \( H_1: \beta_i^e \neq 0 \), there is a \( t \cdot e \) association. MLM3 was fitted to the Revisit data with the R-package glmmTMB v0.2.2.0 (Brooks et al., 2017) and betabinomial response distribution with logit link by the following R code:

```r

MLM3 <- glmmTMB(formula.MLM3, family = betabinomial, data= Revisit)summary(MLM3)

```

The term \( t \cdot e \) results in the trait main effect \( \beta_0 \), the environmental main effect \( \gamma_0 \) and the interaction effect \( \beta_i^e \); the term \((1\cdot \text{trait}|\text{site})\) gives the site random effects \((d_i)\) and \((b_j)\), their variances and correlation, whereas as the term

![FIGURE 3](image_url)  
**FIGURE 3** MLM3 applied to the Revisit data: site and species coefficients with respect to the C:N ratio and the topographic moisture gradient (TMG) plotted against TMG and C:N ratio (left and right respectively) with fitted lines and parametric and permutation p-values, showing positive C:N-TMG interaction (\( \beta_0 = 0.25, SE = 0.095 \)). The long-dash line is at the fixed effects \( \beta_0 = 1.0 (SE: 0.2) \) and \( \gamma_0 = -0.68 (SE: 0.13) \), showing that abundance increases with the C:N ratio and decreases with TMG (left and right respectively). \( N \) = number of species or sites; bootstrap \( p = \) parametric bootstrap \( p \)-value.
(1+\text{env}|\text{species}) gives the species random effects \( \{a_j\} \) and \( \{c_j\} \), their variances and correlation. The betabinomial distribution is chosen because abundance in the Revisit data is a k out of \( N = 100 \) measurement and may show overdispersion compared to the binomial distribution. In Appendix S7, unrestricted count data are modelled similarly with a Poisson or a negative binomial response distribution with log link.

Compared to MLM1 (Miller et al., 2019; Pollock et al., 2012), MLM3 has the extra terms \( d_j + (b_0 + b_1)_j \), that is the main trait effect and two site-specific terms (Table 1), or equivalently \( \text{trait} + (1 + \text{trait} | \text{site}) \). MLM3 satisfies the good practice in regression analysis of including the main effect of any variable included in an interaction. Moreover, Miller et al. (2019) showed that the main effect of a trait should be included to avoid spurious detection of trait–environment interaction. Similarly, one should consider polynomial terms (or more general nonlinear terms such as splines) of \( t \) and \( e \) if needed, despite the fact that these could in principle be absorbed in the random species- and site-specific random intercepts (Jamil & ter Braak, 2013; Jamil et al., 2013). In particular, data with a clear niche structure require a unimodal model. The model with second-order polynomials is the simplest example, provided that coefficients of the squared terms are negative (ter Braak & Looman, 1986).

Compared to MLM2 (Jamil et al., 2013; Miller et al., 2019), MLM3 has the extra random term \( b_j f_j \) (Table 1) or equivalently \( (\text{trait} | \text{site}) \). This term accounts for any interaction that the observed trait \( t \) has with any unobserved (latent) environmental gradient beyond its possible interaction with the observed environmental variable \( e \) (Appendix S2). The random coefficient \( b_j \) is the error term in Equation 2a. Note that Warton, Blanchet, et al. (2015) and Ovaskainen (2017) implicitly model latent traits and environmental variables via a factorial structure for the residual covariance among species. This can be viewed as an unsupervised way of searching for a term like \( b_j f_j \). Their model requires specialized software, whereas MLM3 Equation 3 does not. In comparison with their model, MLM3 lacks terms for residual covariance (e.g. due to species interactions) that is not captured as a linear function of traits via \( (\text{trait} | \text{site}) \). I investigate by simulation how robust MLM3 is to additional residual covariance. In the remainder of this paper, I focus on MLM2 and MLM3 as Miller et al. (2019) showed already that MLM1 performs worse with the Revisit data.

The multisite maxent approach to community assembly by trait selection (CATS) (Shipley, 2010a, 2010b) is a special case of Equation 1 in which \( a_j \) is a known offset that accounts for the abundances in the local species pool, \( c_j = 0 \) for each \( j \), and \( b_j \) and \( d_j \) are treated as a fixed effects. This maxent approach thus lacks explicit environmental variables, but, as MLM3, includes site-specific trait effects. Warton, Shipley, et al. (2015) extended CATS to include environmental variables resulting in MLM(glm).

### 2.3 Weighted averaging-based methods (\( N_2 \)-weighted regressions)

The best known WA-based method is CWMr, a site-level method which detects trait–environment association far too often when no such association exists (Miller et al., 2019; Peres-Neto et al., 2017). This defect can be ameliorated (ter Braak, Peres-Neto, et al., 2018) by combining CWMr with a species-level method in which SNC is regressed against the trait (SNCr). The SNC is a species-specific environmental mean (Figure 1); it averages the environmental variable for each species across sites, weighted by the abundances in the sites, in the same way as the community weighted mean (CWM) averages the trait for each site (Peres-Neto et al., 2017). In these regressions, each species and each site receives unit weight (Appendix S3.2).

The new WA-based method changes these unit weights to new weights motivated by the observation that species usually differ greatly in their abundance and number of occurrence and many species occur only once. Down-weighting abundant species or removing rare species are not attractive options either as weighting and removal criteria are a source of arbitrariness. Removal would reduce the advantage of trait-based analysis of being able to analyse all data, while species-based analysis cannot. In order to account for this variation, the new weights are based on the ‘effective’ number of occurrences. The motivation for this is as follows. A SNC is a mean (of environmental values) so that its accuracy is likely to increase with the number of values that is averaged and the same applies to a CWM. In the new method, CWMr and SNCr are modified by weighting each site point in CWMr by the effective number of species in the site and each species point in SNCr by the effective number of occurrences of the species. For the latter, suppose a particular species, say species \( j \), occurs three times with abundance values 50, 1 and 1. Then its SNC averages three environmental values, weighted by these abundance values. Effectively, however, this SNC is an average of almost one value because of the dominant 50. The Hill number \( N_2 \) expresses this notion (Hill, 1973; ter Braak & Verdonschot, 1995). In formula, the \( N_2 \)-number of species \( j \) is

\[
N_2 = \frac{1}{\sum_i \left( \frac{y_{ij}}{y_j} \right)^2} \tag{4}
\]

with \( y_{ij} = \sum_i y_{ij} \) the species total. In the example, \( N_2 = 1.08 \). When all non-zero abundances of a species are equal, \( N_2 \) is simply the number of non-zero values. Hill numbers of order other than 2 appear less relevant to weighted averages. The \( N_2 \)-number for sites is the inverse of the Simpson index (Hill, 1973). When all species present in a site have equal abundance, its \( N_2 \)-number equals richness. Note that the \( N_2 \)-weights are used in the regressions (CWMr and SNCr) only; the weights used in the weighted averages CWM and SNC themselves are the abundances.

Because weighted parametric tests can show some Type I error rate inflation (ter Braak, Peres-Neto, et al., 2018), each test is carried out by permutation using the ANOVA F-value of the regression as the test statistic. Subsequently, the maximum of the two p-values is taken (max test). The permutation-based unweighted and weighted methods are referred to as lm CWM-SNC and \( N_2 \)-weighted lm respectively. Note that the permutation-based max test using the squared fourth-corner correlation is precisely
equivalent to the permutation-based max test using the F-value of CWMr and SNCr weighted by the sites and species totals, respectively. Appendix S3 provides existing and novel motivation for WA-based methods. In particular, CWM is approximately proportional to the maximum likelihood estimator of the parameter $b^*$ in Equation 1 when specialized to a Poisson log-linear model and, similarly, SNC is approximately proportional to the estimator of the parameter $c^*$ in Equation 1, when both are small (Appendix S3.4).

The correlations that follow directly from CWMr and SNCr are not suited for measuring trait–environment association; they can be overly variable and unstable and also differ between the two regressions (Peres-Neto et al., 2017). The fourth-corner correlation is a better measure but is, in terms of weights in CWMr and SNCr, restricted to site and species totals. Appendix S3.5 generalizes the fourth-corner correlation to a general weighting system. In particular, it proposes a definition for the $N_2$-weighted and unweighted fourth-corner correlations.

2.4 | Data analysis and simulations

The trait and environmental variables were standardized to zero mean and unit variance as in Miller et al. (2019) so as to get standardized estimates of $\hat{\beta}_{te}$, $\sigma_b$ and $\sigma_c$. The WA- and model-based methods were applied to the Revisit data, resulting in parametric and permutation-based p-values for testing the trait–environment interaction. In WA-based permutation tests, the TMG values $\{t_i\}$ are randomly permuted among sites in CWMr, and, similarly, the C:N values $\{c_j\}$ are permuted among species in SNCr. In the model-based permutation tests, either the $\{t_i\}$ or the $\{c_j\}$ of the interaction term $t_j c_i$ of the model are permuted to yield a species- and site-level test respectively. Main effects for the permuted trait and environmental variable are added to ensure that the interaction after permutation has a corresponding main effect. For further details, see Appendices S4 and S5. A MLM3-based parametric bootstrap test was also performed. All multilevel tests are Wald tests, unless noted otherwise.

The MLM3 model was used to simulate new datasets with the R code `simulate(MLM3)`, from which a parametric bootstrap confidence interval for $\hat{\beta}_{te}$ was constructed. Subsequently, the MLM3 object was modified by changing the entry for $\hat{\beta}_{te}$ so as to simulate from models without and with varying strengths of trait–environment interaction. Additional simulations were carried out to investigate the robustness of the methods to large site-specific trait effects, both in terms of type I error rate (for $\hat{\beta}_{te} = 0$), bias in $\hat{\beta}_{te}$ and its standard error. For this, the entry for $\hat{\sigma}_b$ was increased to 1.

Appendix S7 contains similar analyses and simulations based on another real dataset using log-linear models for count data. The WA- and model-based methods were applied to the same datasets with permutation tests using identical permutations.

Appendix S1 is a tutorial in R and Appendix S10 contains the R-functions and scripts used in this paper. The code and a package implementing the methods, called TraitEnvMLMLWA, is available at https://doi.org/10.6084/m9.6figs.8152655.

3 | RESULTS

3.1 | Dataset

MLM3 gave a better fit than MLM2 ($\chi^2 = 28, p << 0.0001$); adding squared main effect terms did not. The trait-related standard deviation ($\sigma_c = 0.34$) is of the same order as the environment-related one ($\sigma_b = 0.42$). The over-dispersion parameter of the beta binomial distribution was estimated as 3.35.

The MLM3 estimate of $\hat{\beta}_{te}$ is 0.25 with a standard error of 0.096. Two thousand parametric bootstraps gave an almost normal distribution of $\hat{\beta}_{te}$ with mean 0.24 and standard deviation 0.096. The usual

![FIGURE 4 N2-weighted lm applied to the Revisit data: the weighted mean C:N ratio against the topographic moisture gradient (TMG) (left) and the weighted mean TMG against the C:N ratio (right) with fitted lines and parametric and permutation p-values, showing the positive association between C:N ratio and TMG. The long-dash line is at the $N_2$-weighted mean of the points (1.0 for sites and −0.33 for species), showing that sites have on average mean C:N ratios greater than the C:N overall mean (left) and species distributions concentrate at the less than average TMG value (right). N2 = effective number of species or sites ($N_2$).](image-url)
normality-based confidence interval of 0.25 ± 2 × 0.096 thus agrees with the bootstrap confidence interval.

Figures 3 and 4 present the essential results of the new WA- and model-based methods. Despite considerable scatter in the graphs with site points and the graphs with species points, the figures show a clear positive association between C:N and TMG which is confirmed with permutation-based p-values well below .05 (Table 2). Figure 3 also shows the main effects in MLM3 - a positive main effect of C:N and a negative main effect of TMG. Main effects are represented in Figure 4 for N2-weighted lm as well by above average mean C:N ratios in sites and below average mean TMG of species, where the means are CWMs and SNCs respectively. The fourth-corner correlation is 0.19 and its N2-weighted version 0.17.

Table 2 compares the p-values of existing and new methods. Except for the fourth-corner, all parametric p-values are below .05, but these are not necessarily trustworthy. The p-values of the permutation-based max test using an appropriate test statistic for each method are just above or below .05 for the existing methods and further below .05 for the new methods with N2-weighted lm giving the lowest value (0.008). The MLM3 parametric bootstrap test (499 bootstraps) yielded p = .014, which is close to the permutation-based max p-value (.012).

### 3.2 | Simulations

Figure 5 shows rejection rates (Type I error rate for $\hat{\beta}_0 = 0$ and power for $\beta_0 > 0$) of each method, when reality is the same as for the MLM3 model with parameters from the fit to the Revisit data using a beta binomial response distribution. Only MLM2 had inflated Type I error rates, 0.09 for $\sigma_0 = 0.34$ (the fitted value) and 0.44 for $\sigma_0 = 1$, which explains its vertical jump at effect size 0 in Figure 5. Of the methods that controlled the type I error rate, MLM3 showed the highest power, followed by N2-weighted lm and the other WA-based methods. Appendix S6 compares the performance of WA- and model-based methods in these simulations in more details.

![MLM3 BB (Revisit data)](image)

**FIGURE 5** Rejection rates of methods for data from the MLM3 model with parameters fitted to the Revisit data using a beta binomial response distribution (BB) except that $\beta_0$ was set post-hoc to the value on the abscissa, based on 250 simulated datasets for each value of $\beta_0$ (Wald test for MLM2 and MLM3 and permutation-based max test for the WA-based methods with $F$-value test statistic and 499 site and species permutations). For $\beta_0 = 0$, additional simulations with $\sigma_0 = 1$ were carried out and the lines include this scenario so as to emphasize potential Type I error rate inflation. The horizontal solid line is at the nominal significance threshold; rates above the dashed line (at 0.078) are significantly greater than 0.05.

MLM2 and MLM3 showed little bias in $\hat{\beta}_0$ in simulations with small $\sigma_0$ (results not shown), but for $\sigma_0 = 1$ MLM2 underestimates both $\hat{\beta}_0$ and the standard error, whereas MLM3 continues to do well (Figure 6). In simulations with log-linear models, MLM2 was unbiased for $\hat{\beta}_0$ but downward biased for its standard error (Appendixes S7 and S8). In simulations of logit models with residual correlation, MLM2 and MLM3 showed 30 and 20% downward bias in $\hat{\beta}_0$, respectively, which, for MLM3 only, disappeared without residual correlation; no such bias was found in log-linear models (Appendix S8). In MLM3, the standard error of $\hat{\beta}_0$ increases sharply with the size of the variance components $\sigma_0^2$ and $\sigma_2^2$ but only slightly with the size the structured noise; MLM2 fails to pick up the variance in $\sigma_2^2$, as it does not include this random component (Figure A2 in Appendix A8.3). In simulations with and without residual correlation (Appendix S8), MLM3 controlled the type I error rate, whereas MLM2 did not (showing type I error rates up to 0.52). However, for small number of species or sites (say, less than 20) the MLM3-based Wald test performed less well (Table A5 in Appendix S8). In a small simulation study, the parametric bootstrap test had 15% more power than MLM3-based max test, at the cost of being based on more assumptions (Appendix S5.2).

Whereas the model-based methods showed type I error inflation for small number of species or sites, the WA-based methods showed some inflation for large numbers, when both $\sigma_0$ and $\sigma_2$ are large. Of these methods, N2-weighted lm is the least affected (Appendix S8). If either $\sigma_0$ or $\sigma_2$ is zero, the WA-based max tests do control the type I error (Appendices S5.1 and S8.3).

### TABLE 2

A p-values for tests on trait (C:N ratio)–environment (topographic moisture gradient: TMG) association in the Revisit data for five statistical methods and four ways of obtaining p-values (prmtrc = parametric Wald test for MLM and parametric max test for (weighted) linear regressions; 499 permutations)

| Method                | Fourth Corner | Site | Species | Max |
|-----------------------|---------------|------|---------|-----|
| Im CWM/SNC            | 0.001         | 0.002| 0.052   | 0.052|
| MLM2 BB               | 0.002         | 0.040| 0.006   | 0.040|
| MLM3 BB               | 0.009         | 0.012| 0.008   | 0.012|
| N2-weighted lm        | <0.001        | 0.004| 0.008   | 0.008|
FIGURE 6 Bias (boxplot) and standard error (red violin plot) of $\hat{\beta}_t$ in MLM2 (left) and MLM3 (right) in simulated data as in Figure 1 with $\sigma_t = 1$, that is data simulated from the MLM3 model of the Revisit data with $\sigma_t$ set to 1. The standard deviation of $\hat{\beta}_t$ across simulations is shown as a red diamond.

4 | DISCUSSION

This paper introduces two new methods for statistical analysis of trait–environment association, $N_t$-weighted lm and MLM3. The permutation-based max tests using these methods and the MLM3-based bootstrap test provide strong statistical evidence for a C:N-TMG association in the Revisit data (Table 2). In the simulations, the permutation-based max test using $N_t$-weighted regressions had at least 70% of the power of MLM3 and performed well compared to the other WA-based methods (Appendices S6-8).

This paper argues that MLM3 is the minimal GLMM model to evaluate the importance of trait–environment association, including its statistical significance. MLM2 simply uses the wrong error term to judge significance, resulting in too many false positives. The reason is that the importance of a fitted line must be evaluated in view of the scatter. In the MLM3 model, there are two such fitted lines (Figure 3; Equations 2a and 2b). Nevertheless, no such model has been proposed previously. Even Figure 3 gives an optimistic view of the scatter around the line as the estimated random effects in mixed models are shrunken towards 0. Note also that more terms may be needed: nonlinear main effect terms, and submodels that account for the fact that the $[b]$ and $[c]$ may not be independent because of spatial and phylogenetic correlations among sites and species respectively.

Previous work could be summarized by ‘a valid trait–environment analysis must combine site-level and species-level analyses’ (Peres-Neto, Dray & ter Braak, 2017; ter Braak, 2017; ter Braak, Peres-Neto, & Dray, 2017, 2018). For example ter Braak et al. (2017) complemented the site-level analysis of MLM(glm) with a comparable species-level analysis. The statement can be paraphrased for multilevel models as ‘a valid trait–environment model must contain both site-level and species-level random effects associated with traits and environment respectively’. I recommend to always include the two variance components associated with these random effects, even when non-significant, because they represent inherent sources of variation in trait–environment studies. These random effects avoid the spurious detection of association between the observed trait and environment in the more realistic situation that relations are not perfect (Figures 2 and 3). Spurious detection is most likely to occur (a) when the observed trait interacts with a latent environmental variable that is uncorrelated to the observed one, (b) the observed environment interacts with a latent trait that is uncorrelated to the observed one, or (c) both (Appendices S2 and S8). As the latent variables are uncorrelated with the observed ones, it is not a ‘confounding variable’ problem (McDonald, 2014) but a problem of not accounting for crossed random effects in the appropriate way.

The max test attempts to account for random effects by re-sampling sites and species, even when the test statistic, such as the fourth-corner correlation, does not account for them. Braga, ter Braak, Thuiller, and Dray (2018) proposed a max test using the fourth-corner correlation with randomizations of the trait and environmental variable that conserve phylogenetic and spatial correlations. The model-based approaches of Warton, Blanchet, et al. (2015) and Ovaskainen et al. (2017) can also account for spatial and phylogenetic correlations. These approaches also model residual correlation which may make them robust to the issues found in MLM2.

MLM3 and the WA-based methods give very similar results for the Revisit data. This does not need necessarily hold true for other data, for example data where species have differential niche widths (Jamil, Kruk, & ter Braak, 2014). Miller et al. (2019) showed that the fourth-corner correlation can have severely inflated Type I error rates in a logit model. They demonstrated this in a scenario where main effects are large (more than twice the size in the Revisit data) and the variance components are small (the most important variance component is that of the species-specific response to the environment which was reduced by a factor of >50). The inflation in this scenario emphasises the fact that interaction is link-dependent (ter Braak et al., 2017). For example main effects with a log-linear link are multiplicative on the count scale and thus interact in a model with identity link (advocated by Clark (2016)). For binomial data, the logit and complementary log-log link-functions may differ somewhat. CWM, SNC and the fourth-corner correlation are more closely linked to the Poisson log-linear model than to logit models and models with overdispersion; in such models, the WA-based statistics may become contaminated with main effects, when large and the maximum probability of occurrences is above 0.3 (Appendices S3.3 and S3.4). The max test guarantees that the abundances are trait- and environment-related, but not that the trait and environment interact (Appendix S5.1). Out of the WA-based methods, the $N_t$-weighted version best controlled the type I error rate in simulations and had comparable power (Appendices S6-8).
With highly variable data, an easy approach, warned against by Warton, Lyons, Stoklosa, and Ives (2016), is to transform the data by taking square-roots or logarithms. Such transformations also implicitly change the model, and hence the (link-)scale on which the interaction is tested. The mass ratio hypothesis, the importance of which is emphasized by Funk et al. (2017), also argues against transformation of abundance in CWM and SNC calculations. An issue is that these methods lack a formal way of choosing the transformation, enabling potential cherry-picking of the transformation showing the most significant effect (in the Revisit data this is the quarter-root transformation). $N_{i}^{2}$-weighted lm does not suffer from this danger; its reduces automatically to unweighted CWM/SNC regression if species occur in about equal numbers of sites and sites contain about equal numbers of species.

In the WA-based methods, the abundances appear as weights in the CWM-, SNC- and fourth-corner correlation, whereas they are response variables in the model-based approach. Also, the WA-based methods appear limited to single traits and single environmental variables. But appearances are deceptive (as Peres-Neto et al. (2017) hinted at); the WA-based methods are closely related to Poisson log-linear models (Appendix S3.4) and to double constrained correspondence analysis (dc-CA) (ter Braak, Šmilauer, & Dray, 2018), in which the abundances are response variables and the trait(s) and environmental variable(s) are predictors as in MLM3. CWM_{r} and SNC_{r} generalize to the second and third step in the four-step algorithm of dc-CA (ter Braak, Šmilauer, et al., 2018). A procedure for selecting functional traits and environmental variables from a greater set of variables is also available (ter Braak & Šmilauer, 2018). This paper extended and compared WA- and model-based methods. The extensions to the multi-trait-multi-environment case deserve a similar comparison.

MLM3 and max-test-based WA-methods are superior to MLM2. One may argue that MLM2 gives higher power than the WA-based methods, but this power comes at a cost of (a) lack of protection against unobserved environmental variation that interacts with the trait, (b) sensitivity to residual correlation, at least in logistic models, and (c) much higher computational costs. MLM3 is superior to the WA-based methods: (a) it gives better type I error control, at least for reasonable numbers of species and sites, and (b) had up to 30% higher power in the simulations. These advantages are offset by greater computational complexity. If a practitioner applies both WA-based methods and MLM3 to the same data with sufficient numbers of species and sites (both >20), the results of MLM3 should be leading.

ACKNOWLEDGEMENTS

I thank Henry Ehlers, John Birks, Kevin Mildau, Stéphane Dray and Pedro Peres-Neto for comments on this or an earlier version. Comments by the Editor, and by Tony Ives, Bill Shipley and the third reviewer led to clarifications and improvements. The author is the senior author of the software Canoco.

DATA AVAILABILITY STATEMENT

Data are available from the Dryad Digital Repository https://doi.org/10.5061/dryad.7g0s3b (Miller, Damschen, & Ives, 2018). Data with tutorial and R package TraitEnvMLMWA are archived at https://doi.org/10.6084/m9.figshare.8152655.v5 (ter Braak, 2019)

ORCID

Cajo J. F. Braak https://orcid.org/0000-0002-0414-8745

REFERENCES

Birks, H. J. B., Line, J. M., Juggins, S., Stevenson, A. C., & ter Braak, C. J. F. (1990). Diatoms and pH reconstruction. *Philosophical Transactions of the Royal Society London, Series B*, 327, 263–278. https://doi.org/10.1098/rstb.1990.0062

Braga, J., ter Braak, C. J. F., Thuiller, W., & Dray, S. (2018). Integrating spatial and phylogenetic information in the fourth-corner analysis to test trait-environment relationships. *Ecology*, 99, 2667–2674. https://doi.org/10.1002/ecy.2530

Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., ... Bolker, B. M. (2017). glmnetMB: Balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal*, 9, 378–400. https://doi.org/10.32614/RJ-2017-066

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

Clark, J. S. (2016). Why species tell more about traits than traits about species: Predictive analysis. *Ecology*, 97, 1979–1993. https://doi.org/10.1002/ecy.1453

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

Funk, J. L., Larson, J. E., Ames, G. M., Wright, J., Butterfield, B. J., Cavender-Bares, J., ... Sutton-Grier, A. E. (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

Hill, M. O. (1973). Diversity and evenness: A unifying notation and its consequences. *Ecology*, 54, 427–432. https://doi.org/10.2307/1934352

Jamil, T., Kruk, C., & ter Braak, C. J. F. (2014). A unimodal species response model relating traits to environment with application to phytoplankton communities. *PloS ONE*, 9, e97583. https://doi.org/10.1371/journal.pone.0097583

Jamil, T., Ozinga, W. A., Kleyer, M., & ter Braak, C. J. F. (2013). Selecting 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

Jamil, T., & ter Braak, C. J. F. (2013). Generalized linear mixed models can detect unimodal species-environment relationships. *PeerJ*, 1, e95. https://doi.org/10.7717/peerj.95

Juggins, S., & Birks, H. J. B. (2012). Quantitative environmental reconstructions from biological data. In H. J. B. Birks, A. F. Lotter, S. Juggins, & J. P. Smol (Eds.), *Tracking environmental change using lake sediments* (pp. 431–494). New York: Springer.

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
Li, D., & Ives, A. R. (2017). The statistical 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
McDonald, J. H. (2014). Handbook of biological statistics, 3rd ed. Baltimore, MD: Sparky House Publishing. Retrieved from http://www.biostathandbook.com/confounding.html
McGill, B. J., Enquist, B. J., Weiher, E., & Westoby, M. (2006). Rebuilding community ecology from functional traits. Trends in Ecology & Evolution, 21, 178–185. https://doi.org/10.1016/j.tree.2006.02.002
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.7ej0s3b
Miller, J. E. D., Damschen, E. I., & Ives, A. R. (2019). Functional traits and community composition: A comparison among community-weighted means, weighted correlations, and multilevel models. Methods in Ecology and Evolution, 10, 415–425. https://doi.org/10.1111/2041-210X.13119
Ovaskainen, O., Tikhonov, G., Norberg, A., Guillaume Blanchet, F., Duan, L., Dunson, D., ... Abrego, N. (2017). How to make more out of community data? A conceptual framework and its implementation as models and software. Ecology Letters, 20, 561–576. https://doi.org/10.1111/ele.12757
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
R Core Team. (2018). R: A language and environment for statistical computing, version 3.5.1. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from www.R-project.org
Shipley, B. (2010a). From plant traits to vegetation structure: Chance and selection in the assembly of ecological communities. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511806971
Shipley, B. (2010b). Inferential permutation tests for maximum entropy models in ecology. Ecology, 91, 2794–2805. https://doi.org/10.1890/09-1255.1
ter Braak, C. J. F. (2017). Fourth-corner correlation is a score test statistic in a log-linear trait–environment model that is useful in permutation testing. Environmental and Ecological Statistics, 24, 219–242. https://doi.org/10.1007/s10651-017-0368-0
ter Braak, C. (2019). Tutorial and R-package TraitEnvMLMWA for new robust weighted averaging- and model-based methods for assessing trait-environment relationships. Figshare. https://doi.org/10.6084/m9.figshare.8152655.v5
ter Braak, C. J. F., & Barendregt, L. G. (1986). Weighted averaging of species indicator values: Its efficiency in environmental calibration. Mathematical Biosciences, 78, 57–72. https://doi.org/10.1016/0025-5564(86)90031-3
SUPPORTING INFORMATION
Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: ter Braak CJF. New robust weighted averaging- and model-based methods for assessing trait–environment relationships. Methods Ecol Evol. 2019;10:1962–1971. https://doi.org/10.1111/2041-210X.13278