Accounting for underlying complexities identifies simple hierarchy of trait–environment relationships in Wisconsin forest understory communities

Andrés G. Rolhauser[1, 2, 3]*, Donald M. Waller[4], Caroline M. Tucker[1,5]

[1]Department of Biology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

[2]Departamento de Métodos Cuantitativos y Sistemas de Información, Facultad de Agronomía, Universidad de Buenos Aires, Buenos Aires, Argentina

[3]IFEVA, Universidad de Buenos Aires, CONICET, Facultad de Agronomía, Buenos Aires, Argentina

[4]Department of Botany, University of Wisconsin-Madison, Madison, WI, USA

[5]Environment, Ecology and Energy Program, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

*corresponding author: rolhauser@agro.uba.ar

Statement of authorship AGR and CMT conceived the ideas and designed the study, with input from DMW, who also provided the data. AGR analyzed the data and led the writing. All authors contributed critically to the drafts and gave final approval for publication.

Abstract

Adaptive relationships between traits and the environment are often inferred from observational data by regressing community-weighted mean (CWM) traits on environmental gradients. However, trait–environment relationships are better understood as the outcome of trait–abundance and environment–abundance relationships, and the interaction between traits and the environment. Accounting for this functional structure and for interrelationships among traits should improve our ability to accurately describe general trait–environment relationships. Using forest understory communities in Wisconsin, we applied a generalized mixed model (GLMM) incorporating this structure. We identified a simple hierarchy of trait–environment relationships dominated by a strong positive effect of mean temperature on plant height. Compared to the traditional CWM approach, the GLMM was more conservative in identifying significant trait–environment relationships, and also detected important relationships that CWM regressions overlooked. This work highlights the need to consider the complexity underlying trait–environment relationships in future analyses.

Keywords Community assembly, environmental gradient, generalized linear mixed model, leaf size, leaf mass per area, mean annual temperature, plant height, selection, soil sand content, temperature seasonality.

Introduction

Understanding how environmental gradients drive plant community composition and species distributions is essential for predicting future responses to global change (Thuiller *et al.* 2008; Kearney *et al.* 2010; Scheiter *et al.* 2013; van Bodegom *et al.* 2014). Theory proposes that the distribution of phenotypic plant traits (e.g. leaf size and potential plant height) across co-occurring species should be adaptively associated with local environments, thus providing a mechanistic link between environmental change and community responses (Keddy 1992; Lavorel & Garnier 2002; Violle *et al.* 2007; Shipley 2010). A considerable amount of empirical research has therefore been devoted to characterizing how phenotypic traits relate to environmental gradients, i.e. the so-called trait–environment relationships (Funk *et al.* 2017). However, reported trait–environment relationships vary widely in strength and sign (see reviews in Garnier *et al.* 2016; Funk *et al.* 2017), highlighting the need to improve our understanding of these relationships if we are to make useful predictions. To this end, we should use analytical tools that better represent and capture the complexities that shape trait–environment relationships.

Mechanistically, trait–environment relationships arise because individual fitness or performance depend on the interaction between functional traits and environmental conditions (Laughlin & Messier 2015; Shipley *et al.* 2016; Laughlin *et al.* 2018). In general, we say that traits and environments are related if the relative performance (or adaptive value) of trait values change along an environmental gradient. This is a two-way interaction and can therefore be viewed from two perspectives (ter Braak 2019). From a site-centric perspective, the trait–performance relationship among coexisting individuals would change when environmental conditions change, while a species-centric perspective would show that environment–performance relationships change among species that differ in their trait values (Fig. 1). The corollary is that trait–environment relationships are conceptually three-dimensional and cannot be understood on the information conveyed by trait and environmental axes alone. Rather, detecting trait–environment relationships depends on properly characterizing how both trait values and environmental conditions affect performance. For this, we must acknowledge the different shapes that all these relationships may take.

Species' performances typically show unimodal responses to *environmental* variables

reflecting intermediate optima along the gradient (Curtis 1959; Whittaker 1967; Ter Braak &

Prentice 1988; Austin 1999). Similarly, we expect unimodal performance responses to *traits*

when functional trade-offs lead to optimum trait values within sites (Muscarella & Uriarte 2016;

Rolhauser & Pucheta 2017; Rolhauser *et al.* 2019). Environment–performance relationships can

also be bimodal because of intense negative interspecific interactions (Mueller-Dombois &

Ellenberg 1974; Minchin 1987; Austin 1999), nonlinear trait–environment relationships

(Laughlin & Joshi 2015), or unmeasured environmental factors (Austin *et al.* 1984). Similarly,

trait–performance relationships can be bimodal (or multimodal) when within-site

environmental heterogeneity favors the functional divergence of competitive, dominant species

(Rolhauser & Pucheta 2017; Rolhauser *et al.* 2019). Rectilinear or directional relationships

emerge as particular cases when performance maxima (in optimum relationships) or minima (in

bimodal relationships) occur at the extremes or outside the range of observed trait and

environmental values (Rolhauser *et al.* 2019). Unlike trait–performance and environment–

performance relationships, theoretical discussions on the shape of trait–environment

relationships are rare. Both conceptual (Laughlin & Messier 2015; Garnier *et al.* 2016; Shipley *et*

al. 2016) and empirical studies (e.g. Niinemets 2001; Wright *et al.* 2004, 2017; Moles *et al.* 2009,

2014) largely assume rectilinear relationships. Nonetheless, some empirical work exists to show

that trait–environment relationships can be nonlinear (Laughlin *et al.* 2011; Bello *et al.* 2013).

Studies of trait–environment relationships often resort to various simplifications of the

three-dimensional problem to infer trait–environment relationships. Mechanistic studies that

directly assess how individual performance responds to conditions remain rare and typically

assume rectilinear trait–performance and environment–performance relationships (e.g.

Laughlin *et al.* 2018). Other studies make even stronger assumptions, e.g., by directly regressing

trait values against environmental variables (see reviews in Garnier *et al.* 2016; Funk *et al.* 2017).

These often involve regressing community-weighted means (CWM) on environmental

variables, termed CWM regression (CWMr) (Miller *et al.* 2019), but we can also find regressions

of site-specific species trait averages (e.g. Dong *et al.* 2020) and individual trait values (Laughlin

et al. 2012). These trait–environment regression approaches assume that the abundance or

presence of a species at a given site indicates their performance. They also have two other

important limitations. First, because they do not explicitly model abundance, they cannot assess

plausible differences in trait–abundance and environment–abundance relationships. For

example, CWMs represent the adaptive value of a trait at a given site only when the most

abundant phenotype happens to be close to the trait mean (Muscarella & Uriarte 2016),

although this is a special case (Rolhauser & Pucheta 2017). Second, they tend to evaluate traits

one at a time using separate regressions, implicitly assuming independence between traits and

their responses to environments. However, traits are inherently interrelated and evolutionary

biologists have long recognized the importance of using multivariate regressions to tease apart

the adaptive significance of correlated traits (Lande & Arnold 1983).

Here, evaluated trait–environment relationships across forest understory communities

in Wisconsin using a generalized linear mixed model (GLMM) (ter Braak 2019) that

incorporated the underlying complexity discussed above. We modeled nonlinearities in all

three functional relationships: trait–abundance, environment–abundance, and trait–

environment (Figs. 1, S3). We also evaluated multiple traits and trait–environment relationships

simultaneously. We used a large dataset (Waller *et al.* 2012) that included: abundances of 185

species distributed among 189 sites; ten functional traits characterizing tissues (e.g. leaf mass

per area, or LMA), organs (e.g. leaf size, or LS), and whole plants (e.g. maximum vegetative

height, or VH); and fourteen environmental variables reflecting local environmental abiotic (e.g.

mean annual temperature and soil sand content) and biotic (tree basal area) conditions.

Functional traits like VH, LS, and LMA have been widely studied as controls of plant

performance and tend to vary along abiotic and biotic gradients (see review in Pérez-

Harguindeguy *et al.* 2013). To demonstrate the usefulness of modelling key underlying

ecological complexities using a single GLMM, we compared the resulting trait–environment

relationships to those found using the traditional CWMr. We show that our approach allowed

us to achieve two disparate goals: explaining a large proportion of the variability in species

abundance across sites while improving the accuracy and reliability of inferring which trait–

environment relationships matter.

Material and methods

Study area

Wisconsin covers 169,639 km^2 yet its elevation ranges only between 177-595 m. Its highly seasonal north temperate climate shows conspicuous north-south gradients in summer and winter temperatures and east-west gradients in seasonality and rainfall that reflect the moderating maritime influence of Lake Michigan to the East. The vegetation is dominated by deciduous forests that can be divided into two floristic provinces: the prairie-forest province to the southwest and the northern hardwoods province to the north, separated by a marked "tension zone" running from the SE to the NW edges of the state (Curtis 1959; see Fig. S1 in Supporting Information).

Vegetation sampling

Data were collected through contemporary resurveys of forests stands (sites) originally surveyed by J.T. Curtis and his students in the 1950s (see Fig. S1 in Supporting information). Only sites that remained less disturbed (i.e. with an intact forest canopy, no nearby edges, and few signs of understory disturbance) were resurveyed. Resurveys in the 2000s employed Curtis-era and similar methods covering similar areas (Waller et al. 2012; see Appendix S1 for details). Briefly, all vascular plants were tallied in each of many replicate (20-403) 1-m^2 quadrats to estimate their abundance (frequency) at each site. In total, 536 species were found at 293 sites. We restricted our attention to the commonest 185 species, for which we had registered trait values. We then selected sites to ensure high trait coverage (80+% of cumulative species occurrences). We also discarded a few sites missing full environmental data, yielding a total of 189 sites.

Environmental variables

Descriptions of the four climatic variables, ten soil variables and one biological variable (tree basal area) appear in Appendix S1. Climatic variables for each site were derived from 10-year

averages for the period 1995-2004. Soil and basal area data derive from samples and surveys at

each site. To limit collinearity among predictors, we first performed pairwise correlations and

principal components analyses to select more independent subsets of variables (all correlations

<0.5) to represent major climatic and soil gradients (Appendix S1). We selected six predictors for

analysis: mean annual temperature (MAT), mean annual precipitation (MAP), temperature

seasonality (i.e. the standard deviation of mean monthly temperatures, abbreviated TSD), soil

nitrogen content (%N), soil sand content (%Sand) and tree basal area (BA). Before analysis,

environmental variables were transformed as needed to reduce the weight of extreme values,

then standardized to zero mean and unit variance (Appendix S1).

164   *Plant traits*

In a similar fashion, we initially analyzed data for ten functional traits in the most common 185

herbaceous species (see Appendix S1 for descriptions of traits and data collection methods).

Briefly, 12+ individuals (at least four plants from each of three sites) were collected in Wisconsin

and processed between 2008 and 2014 following standardized protocols (Pérez-Harguindeguy

et al. 2013) as first described by Amatangelo *et al.* (2014). We then selected four traits for analysis

with low pair-wise correlations (<0.3; Table S3): vegetative height (VH), leaf size (LS, calculated

as the product of length and width), leaf mass per area (LMA), and leaf carbon content (LCC).

Traits were transformed as needed to reduce the weight of extreme values and standardized

(Appendix S1).

175   *Statistical methods*

176   <u>GLMM</u>

The aim of the approach below is at inferring the shape and significance of trait–environment

relationships while modeling meaningful site and species responses to, respectively, trait and

environmental axes (Fig. 1). We used a negative binomial generalized linear mixed model

(GLMM) to model the abundance of species i at site j (A_{ij}) as a function of selected traits and

environmental variables via the natural logarithm link function. A_{ij} is a count variable

calculated as the frequency of quadrats in which species i was present at site j. Negative

binomial models directly estimate an overdispersion parameter from the data (Agresti 2015).

Before describing the full multivariate model, we outline an analysis for one trait and one

environmental variable, both standardized. Given the log link function, the fixed effects in the

GLMM are (cfr. Fig. 1):

$$\ln(A_{ij}) = \beta_0 + \beta_1 T_i + \beta_2 T_i^2 + \beta_3 E_j + \beta_4 E_i^2 + \beta_5 T_i E_j + \beta_6 T_i E_j^2 \qquad (1a)$$

where β_0 is y-intercept of the fitted surface. Since trait T is centered at zero, β_1 and β_2

respectively estimate the mean slope (or "gradient") and the mean curvature of the trait–

abundance relationship (see e.g. Aiken et al. 1991) for the particular case when E_j is set to its

average value (zero). That is, β_1 measures the direction (positive or negative) and the strength of

trait T_i's effect on species abundance. Negative and positive values of β_2 indicate a n- and u-

shaped relationships, respectively. Similarly, β_3 and β_4 reflect the mean slope and curvature of

the environment–abundance relationship (ter Braak 2019). Due to the combination of quadratic

effects (β_2 and β_4) and the log link associated with the negative binomial distribution, equation

1a can reproduce realistic, bell-shaped trait–abundance and environment–abundance

relationships (Figs. 1, S3a). While doing this, we can assess trait–environment relationships by

estimating interactions between T and E, i.e. parameters that quantify how different

environmental conditions modulate the mean effect of the trait on the response variable

(Laughlin et al. 2018). In our model, these environmentally mediated trait effects are estimated

by β_5 and β_6.

Rearranging Eqn. 1a to gather terms for the mean (linear) effect of trait T_i on A_{ij} yields:

$$\ln(A_{ij}) = \beta_0 + T_i(\beta_1 + \beta_5 E_j + \beta_6 E_j^2) + \beta_2 T_i^2 + \beta_3 E_j + \beta_4 E_j^2 \qquad (1b)$$

Then, the environment-dependent mean slope of trait–abundance relationships (denoted φ_j) is:

$\quad \varphi_j = \beta_1 + \beta_5 E_j + \beta_6 E_j^2$ (2)

In this quadratic function, β_5 and β_6 are respectively the mean slope and mean curvature of the

trait–environment relationship (Fig. S3). Indeed, since $T_i E_j^2 = (T_i E_j) E_j$, β_6 estimates how much

the strength of the trait–environment interaction depends on the values of the environment (see

equation 1a), i.e. a quadratic relationship. Importantly, the properties of the fitted surface (e.g.

the position of optimum trait values) change along the environmental gradient depending on β_5

and β_6 (Fig. S3).

$\qquad$ The full GLMM included four traits (VH, LMA, LS, LCC) and six environmental

variables (MAT, MAP, TSD, %Sand, %N and BA). Following equation 1a, the fixed effects

included linear and quadratic terms for all traits and environmental variables as well as all

trait–environment interactions. We formulated random effects after ter Braak (2019) including

random intercepts and slopes for all traits and environmental variables (i.e. MLM3 sensu ter

Braak 2019). The model in R code notation, is thus:

glmmTMB(abundance~(MAP + MAT + TSD + %Sand + %N + BA + I(MAP^2)

+ I(MAT^2) + I(TSD^2) + I(%Sand^2) + I(%N^2) + I(BA^2)) * (VH + LS + LMA + LCC)

+ I(VH^2) + I(LS^2 + I(LMA^2) + I(LCC^2))

+ (1 + VH + LS + LMA + LCC | site)

+ (1 + MAP + MAT + TSD + %Sand + %N + BA | sp)

+ offset(log(quadrats))

$\qquad$ Note that the random slopes with respect to a given trait, T, are estimated for each site

(denoted $b_{(T)j}$ here). Random slopes with respect to environmental variables E are also estimated

for each species ($c_{(E)i}$) (ter Braak 2019). We included the log number of quadrats sampled per site

to measure search-effort, used in our model as an offset (Kéry 2010, pp188-189). We fitted the

GLMM using the R-package glmmTMB (Brooks et al. 2017). We calculated marginal and

conditional R^2 (the proportion of the total variance explained by fixed effects and by both fixed

and random effects, respectively) following the delta method (Nakagawa et al. 2017)

implemented by the r.squaredGLMM function in the R-package MuMIn (Barton 2019). We

visualized uncertainty around estimates of φ_j (for a given *T-E* combination) by plotting overall

site effects calculated as the sum of the fixed (φ_j) and the random slopes associated with the

corresponding trait ($b_{(T)j}$) (ter Braak 2019).

Comparing GLMM and CWMr approaches

We calculated community-weighted means (CWM) as usual:

$$CWM_{(T)j} = \frac{\sum_{i=1}^{n} A_{ij} T_{ij}}{\sum_{i=1}^{n} A_{ij}} \qquad (3)$$

where A_{ij} and T_{ij} are the abundance and the standardized trait value for species *i* in site *j*. We

then related CWM values to standardized environmental variables (E_j) through univariate

regressions (separate for each trait). We included both linear and quadratic terms for E_j, leading

to the following general form

$$CWM_{(T)j} = \gamma_0 + \gamma_1 E_j + \gamma_2 E_j^2 \qquad (4)$$

We fitted CWMr models using the least squares method implemented by the lm function in R.

We evaluated the agreement between GLMM and CWMr approaches by comparing the

statistical significance of homologous model parameters for each trait–environment

combination, i.e. β_5 vs γ_1 and β_6 vs γ_2 (cf. equations 2 and 4). We divided test statistics (*z* and *t* in

the case of GLMM and CWMr, respectively) into those that led to the rejection of the null

hypothesis or not and compared these between approaches (see Appendix S1 for details). There

are several possible outcomes for each trait–environment relationship tested (Table S4; Laughlin

et al. 2018). If both approaches yield non-significant effects, then they agree that the trait has no

adaptive value along that environmental gradient (Outcome #1, or O#1). A trait–environment

relationship found to be not significant with the CWMr model but significant with GLMM

indicates that the trait has likely adaptive value along the gradient, but the CWMr method

lacked sensitivity to detect it, possibly due to confounding abiotic or biotic factors (O#2). If

conversely the trait–environment interaction is significant by CWMr but not by GLMM, the

CWMr result may be spurious or indirect reflecting the highly inflated type I error rates found

in CWMr models (Peres-Neto et al. 2017; Miller et al. 2019) (O#3). Finally, if both methods

identify a significant relationship with the same sign, we can safely conclude that the trait has

adaptive value along that environmental gradient (O#4). It is also possible that the relationship

is significant via both approaches, but with opposite signs (O#5, Table S4).

Results

GLMM

Overview

The full model included 113 parameters (overdispersion, 43 random effects, and 69 fixed effects)

to explain abundance across 34,965 species-site combinations. The marginal R^2 of this model

was 0.385 and the conditional R^2 was 0.977. We focus our discussion on trait–environment

relationships, but full model results appear in Appendix S2. Quadratic terms associated with

main trait and environment effects on abundance were negative, indicating optimum

relationships (Table S5). Removing either all quadratic or all interaction terms from the full

model increased AIC>70 units (Table S6), confirming the importance of including these

complexities when analyzing abundance.

Trait–environment relationships

The model identified five significant trait–environment relationships, defined as those

combinations of a trait and an environmental variable where either of both interaction

parameters (β_5 and β_6) showed $p<0.01$ (Table S5). Three relationships were linear, where with

only β_5, not β_6, was significant (shown as blue lines in Fig. 2). Among these, the interaction

between mean annual temperature (MAT) and vegetation height (VH) was the strongest, with

β_5=1.031 (SE=0.146), whereas the remaining significant linear interactions all were below 0.34

(Table S5). Warmer sites (higher MAT) favored taller plants (Fig. 2, top line, second plot and

Fig. 3, a1 and a2). VH also increased linearly with temperature seasonality (TSD) (β_5=0.328,

SE=0.082): more seasonal sites (higher TSD) favored taller plants (Fig. 2, top line, third plot and

Fig. 3, b1 and b2). In addition, leaf size (LS) declined linearly as soil sand content increased (β_5=-

0.266, SE=0.075; Fig. 2, row 2, column 4 and Fig. 3, c1 and c2).

Temperature seasonality (TSD) was negatively associated with both LS (β_5=-0.242,

SE=0.082) and leaf mass per area (LMA) (β_5=-0.239, SE=0.082), and in both cases these decreases

were nonlinear ($\beta_6 \neq 0$) (orange lines, Fig. 2, column 3; Table S5). More seasonal sites (higher TSD)

favored plants whose leaves were smaller and had lower LMA (Fig. 3, b4, b5, b7 and b8). In

these cases, nonlinearity meant that TSD negative effects on LS and LMA were maximal at low

seasonality and tended to disappear as TSD increased.

Other linear trait–environment interactions were marginally significant (0.05<p<0.01, Fig.

2). These included the increase in plant height with soil N (row 1, column 5), the declines in

LMA and leaf carbon content (LCC) with MAT (rows 3 and 4, col 2), and the increase in LCC

with soil sand (row 4, column 4). Basal area (BA) also slightly increased LCC (row 4, column 6).

Surprisingly, annual precipitation (MAP) did not appear to affect any of the evaluated traits

(column 1).

Comparing GLMM and CWMr approaches

The CWMr approach identified 17 significant trait–environment relationships (out of 24), i.e.

single regressions where either of both parameters (γ_1 and γ_2) showed p<0.01 (Table S7). Twelve

were exclusively rectilinear or directional (only $\gamma_1 \neq 0$; e.g. LMA–MAT), four were curved with a

directional component (both $\gamma_1 \neq 0$, $\gamma_2 \neq 0$; e.g. LMA–%N), and one was curved with no significant

directionality (only $\gamma_2 \neq 0$; LCC–MAP) (Table S7).

Test statistics for linear trait–environment parameters (β_5 and γ_1) were quantitatively

similar between the GLMM and CWMr approaches (R=0.61; t=3.572, df=22, p=0.002; Fig. 4a).

Test statistics for the quadratic terms (β_6 and γ_2) were less similar (R=0.41; t=2.095, df=22,

p=0.048; Fig. 4b). Both approaches agreed that six linear terms (including all four interactions

with BA; Outcome #1, Fig. 4a) and 16 quadratic statistics lacked significance (also O#1, Fig. 4b).

In addition, three linear effects of environmental conditions on plant traits (LS–TSD, LS–%Sand,

and VH–MAT) and one quadratic term (LS–TSD^2) emerged as significant in both models (O#4).

Reassuringly, we found no cases where the two models yielded opposite conclusions (O#5).

However, inferences differed conspicuously between approaches in several instances. Two

linear and one quadratic term emerged as significant in the GLMM but not in CWMr model

(O#2). As a result, among the five significant trait–environment relationships in the GLMM

(second column in Fig. 3), two were largely missed by the CWMr approach and both involved

TSD (third column in Fig. 3). Perhaps more worrying were many apparent Type I errors

wherein 13 linear and four quadratic terms that emerged as significant from the CWMr model

lacked significance using the GLMM approach (O#3).

333   Discussion

Trait–environment relationships provide a useful tool to understand and predict community

responses to environmental change, assuming their accuracy is not compromised by using an

inappropriate modelling approach. Unlike community-weighted mean regressions (CWMr),

generalized linear mixed models (GLMM) are compatible with the ecological notions that trait–

environment relationships are inherently three-dimensional and not independent from one

another. Applied to Wisconsin understory vegetation, we found that CWMr inferences granted

significance to far more relationships than the more functionally informed GLMM. Only five

relationships emerged as statistically significant (of 24 evaluated) in the GLMM, and two of

them were largely missed by the CWMr approach. Nonetheless, this handful of trait–

environment relationships contributed to explain nearly 40% of the variation in abundance of

185 plant species distributed among 189 sites.

346   *Prominent trait–environment relationships*

The interaction between mean annual temperature (MAT) and vegetative height (VH) emerged

as the strongest in the GLMM, both in terms of its standardized coefficient (β) and statistical

significance. It was also the only one where the mean slope of the trait–abundance relationship

(φ_i) clearly switched signs over the length of the environmental gradient (Fig. 3), indicating a

clear change in the rank order of species abundances. As such, the VH–MAT relationship would

be the only one considered as "strong" by others (Laughlin et al. 2018). This relationship

predicts increases in plant height as MAT increases, consistent with global patterns (Moles et al.

2009, 2014). This may be explained by the fact that warmer climates promote taller plants as

they compete for light (Givnish 1982; Falster and Westoby 2003) and/or that shorter plants are

less prone to freeze-embolism in colder climates (Moles et al. 2009).

The fact that leaf size (LS) declined at sites with sandier soils agrees with global patterns

showing that water-limited environments tend to favor plants with smaller leaves (Wright et al.

2017). The standardized coefficient (β_5) for the LS–%Sand interaction was only a quarter the size

of that for the VH–MAT interaction. This may reflect the subordinate role of soil properties on

community assembly at the scale of our observations.

Both leaf size (LS) and leaf mass per area (LMA, the inverse of specific leaf area)

declined with the increase of temperature seasonality (TSD). Temperature seasonality also

strongly increased plant abundance, suggesting that many understory species thrive in more

seasonal environments. This could reflect how Spring and Fall periods of higher light in these

deciduous forests and warmer forest floor conditions favor understory plant growth.

Consequently, species in more seasonal sites would also face more competitive conditions. This

could explain why these sites support taller plant species as height is favored under crowded

conditions (Givnish 1982; Falster and Westoby 2003). Because low-LMA leaves tend to be short-

lived (Poorter et al. 2009), they could be favored at sites with short Spring windows of full light

or shorter seasons (Kikuzawa et al. 2013; Mason and Donovan 2015). Given that leaves often

start to generate carbohydrates at about half their final size (Turgeon 1989), shorter seasons

might also favor smaller leaves for their ability to start exporting carbohydrates sooner than

larger leaves.

Among the environmental factors, only temperature seasonality (TSD) showed

significant nonlinear relationships with plant traits. Specifically, TSD effects on LS and LMA

diminished as seasonality increased. Others that have noted these nonlinearities have

speculated on their theoretical consequences (Laughlin and Joshi 2015). However, we are not

aware of theory to explain the mechanisms behind nonlinear trait–environment relationships.

381 *Missing trait–environment relationships*

Despite many published global-scale studies about how precipitation affects plant traits

(Niinemets 2001; Moles et al. 2009, 2014; Wright et al. 2017), we found no trait–precipitation

interactions in this study. This may reflect the fact that these Wisconsin sites span only 5% of the

0 to 4500 mm/year range spanned at the global scale. In contrast, soil sand varied from 15% to

89% among our sites and we found it to affect leaf size (and leaf carbon content, marginally).

This may imply that sand content better reflects water availability for plants at this scale.

Basal area, our surrogate for understory light, only affected LCC weakly, probably reflecting

how notoriously variable understory light levels are within sites (Canham et al. 1994). The weak

linear effects we found for soil N (favoring taller plant species with flimsier leaves) could also

reflect local variability (e.g., in soil conditions) or perhaps that soil N mainly affects plants

through interactions with other environmental variables. Such interactions were not included in

our GLMM.

Relationships between MAT and LMA have also attracted the attention of ecologists.

The negative LMA–MAT relationship we found was marginally significant, consistent with

other regional-scale studies finding weakly negative or non-significant relationships (Laughlin

et al. 2012, 2018; Mason and Donovan 2015; Rosbakh et al. 2015; Balazs et al. 2020; Dong et al.

2020). Such weak field effects are surprising given the strong negative effects found under

controlled growth conditions (Poorter et al. 2009). We found, however, that LMA actually

responded more to temperature seasonality in the GLMM. In larger, especially global, datasets,

mean temperature and seasonality are so closely correlated (see e.g., Kikuzawa et al. 2013) that

we cannot decouple their separate effects. In our dataset, on the contrary, the MAT–TSD

correlation was mild (-0.37; Table S2) suggesting that the weak MAT effect was not an artifact of

seasonality soaking up species abundance variation. Building on efforts linking leaf longevity to

MAT and seasonality (Kikuzawa et al. 2013) it seems that much remains to be understood

regarding the individual and interactive effects of these variables on LMA.

GLMM vs. CWMr approaches

Community-weighted mean regressions are already known to have inflated type I errors, which

manifest as falsely identifying particular trait–environment associations as significant even in

simulations involving no actual association (Peres-Neto et al. 2017; Miller et al. 2019). In

contrast, GLMM approaches show good balance between type I error control and power (Miller

et al. 2019), particularly when random slopes for both traits and environmental variables are

included in the model (MLM3 sensu ter Braak 2019). Our results support these conclusions and

highlight limitations with CWMr approaches. Many of the terms in the CWMr that emerged as

significant (including commonly reported trait–environment relationships like LMA with MAP,

MAT and %N, LS with MAP and MAT, and VH with MAT and %N) lacked significance in the

GLMM (Outcome #3). Given how frequently these relationships are described in the literature,

we have concerns that some reports may reflect spurious false positives from CWMr

approaches. Conversely, two trait–environment relationships featuring TSD were highly

significant in the GLMM (with VH and LMA) but lacked significance under the CWMr

approach (Outcome #2). Importantly, these hidden relationships emerged in a GLMM able to

control for effects of other trait–environment relationships. This suggests that CWMr

approaches may also generate false negatives, possibly accounting for why this TSD

relationships are under-reported in the literature, particularly in cases were MAT and TSD are

highly correlated. All these inferential differences are particularly striking given that both

approaches are based on exactly the same data.

Strengths, limitations, and future directions

Using a single model enabled us to directly rank trait–environment relationships underlying

species distributions. This cannot be done with separated regressions using single traits as

response variables because they fail to account for unmodeled factors, such as other traits. A

paradigmatic case is that of LMA, which showed extremely significant linear correlations with

three environmental variables (sand and nitrogen content and MAT; see Fig. 4a) that were

deemed nonsignificant in the GLMM. The lack of significant LMA interactions in the GLMM

cannot be ascribed to collinearity with a single competing trait since all correlations with LMA

were below 0.18 (Table S3). Rather, it might have to do with the combination of the many effects

included in the model. Modelling several trait–environment combinations simultaneously

accounts for any possible overlapping effect on species abundance arising from collinearity,

while quadratic effects allow for flexibility in functional relationships. In particular, we

introduced nonlinear trait–environment into the GLMM approach which showed to be

important to characterize temperature seasonality effects. Paradoxically, despite its complexity,

the GLMM points out to a much simpler hierarchy of relationships compared to the

unstructured plethora suggested by the CWMr approach. Here, temperature (mean and

seasonality) appeared to drive plant height variation, while soil texture (in combination with

temperature seasonality) played a subordinate role affecting leaf traits.

We interpreted φ_j as reflecting how strongly trait values are "favored", "selected" or

"filtered" along particular environmental gradients (Shipley 2010; Rolhauser and Pucheta 2017;

Loranger et al. 2018). We implicitly assumed that trait–environment effects on plant abundance

are not being strikingly confounded by other processes like dispersal and demographic

stochasticity. Yet we acknowledge that stochasticity may play important roles in community

dynamics (Hubbell 2001; Alonso et al. 2006) and could indeed help explain some of the ~60% of

abundance variation unaccounted for by fixed effects in our GLMM. Dispersal limitation now

limits the presence or abundance of some species given extensive habitat fragmentation

particularly in southern Wisconsin forests (Rogers et al. 2009). However, the traits we used here

are more related to local plant performance (see Pérez-Harguindeguy et al. 2013) than to

dispersal among forest patches; for instance, differences in plant height likely affect dispersal

distances, but primarily at fine scales (e.g. Gómez 2007).

Approaches to trait–environment relationships based on the actual distribution of

plants, rather than on individual or population-level performance (such as demographic rates),

have been criticized due to their phenomenological nature (Laughlin and Messier 2015;

Laughlin et al. 2018). Here, we argue that many of the problems associated with

phenomenological approaches arise when we collapse or aggregate the abundance dimension,

as in CWMr. It might thus be useful to evaluate agreement between GLMMs that use

abundance vs. individual or population-level performance as response variables. Ultimately,

the mismatch between these approaches will depend on the link between performance and

abundance, which still remains as a foundation stone in community ecology with loose

empirical support (McGill et al. 2006; Adler et al. 2013; Shipley et al. 2016). We also hope that

our field-based results stimulate much needed theoretical developments on nonlinear trait–

environment relationships, particularly considering temperature seasonality which may be a

key environmental driver of community assembly along regional gradients.

Acknowledgements

474 The vegetation surveys and studies of trait variation were supported by NSF grants DEB

475 023633, DEB 0717315, and a Dimensions of Biodiversity award (DEB-1046355). We also thank J.

476 Beck for providing comments on an early version of the manuscript and Jeremy Ash for his

477 assistance in the compilation of climatic data. The authors declare no competing interests.

References

Adler PB, Fajardo A, Kleinhesselink AR, Kraft NJB (2013) Trait-based tests of coexistence

mechanisms. Ecol Lett 16:1294–1306

Agresti A (2015) Foundations of linear and generalized linear models. John Wiley & Sons

Aiken LS, West SG, Reno RR (1991) Multiple regression: Testing and interpreting interactions.

sage

Alonso D, Etienne RS, McKane AJ (2006) The merits of neutral theory. Trends Ecol Evol 21:451–

457

Amatangelo KL, Johnson SE, Rogers DA, Waller DM (2014) Trait–environment relationships

remain strong despite 50 years of trait compositional change in temperate forests. Ecology

95:1780–1791

Austin MP (1999) A silent clash of paradigms: some inconsistencies in community ecology.

Oikos 170–178

Austin MP, Cunningham RB, Fleming PM (1984) New approaches to direct gradient analysis

using environmental scalars and statistical curve-fitting procedures. Vegetatio 55:11–27

Balazs KR, Kramer AT, Munson SM, et al (2020) The right trait in the right place at the right

time: Matching traits to environment improves restoration outcomes. Ecol Appl 30:e02110

Barton K (2019) MuMIn: multi-model inference. R Packag. version 1.43.6. https//CRAN.R-

project.org/package=MuMIn

Bello F de, Lavorel S, Lavergne S, et al (2013) Hierarchical effects of environmental filters on the

functional structure of plant communities: a case study in the French Alps. Ecography

(Cop) 36:393–402

Brooks ME, Kristensen K, van Benthem KJ, et al (2017) glmmTMB balances speed and flexibility

among packages for zero-inflated generalized linear mixed modeling. R J 9:378–400

Canham CD, Finzi AC, Pacala SW, Burbank DH (1994) Causes and consequences of resource

heterogeneity in forests: interspecific variation in light transmission by canopy trees. Can J

For Res 24:337–349

Curtis JT (1959) The vegetation of Wisconsin: an ordination of plant communities. University of

Wisconsin Pres

Dong N, Prentice IC, Wright IJ, et al (2020) Components of leaf-trait variation along

environmental gradients. New Phytol

Falster DS, Westoby M (2003) Plant height and evolutionary games. Trends Ecol Evol 18:337–

343

Funk JL, Larson JE, Ames GM, et al (2017) Revisiting the Holy Grail: using plant functional

traits to understand ecological processes. Biol Rev 92:1156–1173

Garnier E, Navas M-L, Grigulis K (2016) Plant functional diversity: organism traits, community

structure, and ecosystem properties. Oxford University Press

Givnish TJ (1982) On the adaptive significance of leaf height in forest herbs. Am Nat 120:353–

381

Gómez JM (2007) Dispersal-mediated selection on plant height in an autochorously dispersed

herb. Plant Syst Evol 268:119–130

Hubbell SP (2001) The unified neutral theory of biodiversity and biogeography. Princeton

University Press

Jamil T, Ozinga WA, Kleyer M, ter Braak CJF (2013) Selecting traits that explain species–

environment relationships: a generalized linear mixed model approach. J Veg Sci 24:988–

1000

Kearney MR, Wintle BA, Porter WP (2010) Correlative and mechanistic models of species

distribution provide congruent forecasts under climate change. Conserv Lett 3:203–213

Keddy PA (1992) Assembly and response rules: two goals for predictive community ecology. J

Veg Sci 3:157–164

Kéry M (2010) Introduction to WinBUGS for ecologists: Bayesian approach to regression,

ANOVA, mixed models and related analyses. Academic Press

Kikuzawa K, Onoda Y, Wright IJ, Reich PB (2013) Mechanisms underlying global temperature-

related patterns in leaf longevity. Glob Ecol Biogeogr 22:982–993

Lande R, Arnold SJ (1983) The measurement of selection on correlated characters. Evolution (N

Y) 37:1210–1226

Laughlin DC, Fule PZ, Huffman DW, et al (2011) Climatic constraints on trait-based forest

assembly. J Ecol 99:1489–1499

Laughlin DC, Joshi C (2015) Theoretical consequences of trait-based environmental filtering for

the breadth and shape of the niche: New testable hypotheses generated by the Traitspace

model. Ecol Modell 307:10–21. https://doi.org/10.1016/j.ecolmodel.2015.03.013

Laughlin DC, Joshi C, van Bodegom PM, et al (2012) A predictive model of community

assembly that incorporates intraspecific trait variation. Ecol Lett 15:1291–1299.

https://doi.org/10.1111/j.1461-0248.2012.01852.x

Laughlin DC, Messier J (2015) Fitness of multidimensional phenotypes in dynamic adaptive

landscapes. Trends Ecol Evol 30:487–496

Laughlin DC, Strahan RT, Adler PB, Moore MM (2018) Survival rates indicate that correlations

between community-weighted mean traits and environments can be unreliable estimates

of the adaptive value of traits. Ecol Lett 21:411–421. https://doi.org/10.1111/ele.12914

Lavorel S, Garnier E (2002) Predicting changes in community composition and ecosystem

functioning from plant traits: revisiting the Holy Grail. Funct Ecol 16:545–556

Loranger J, Munoz F, Shipley B, Violle C (2018) What makes trait-abundance relationships when

both environmental filtering and stochastic neutral dynamics are at play? Oikos 127:1735–

1745. https://doi.org/10.1111/oik.05398

Mason CM, Donovan LA (2015) Evolution of the leaf economics spectrum in herbs: evidence

from environmental divergences in leaf physiology across Helianthus (Asteraceae).

Evolution (N Y) 69:2705–2720

McGill BJ, Enquist BJ, Weiher E, Westoby M (2006) Rebuilding community ecology from

functional traits. Trends Ecol Evol 21:178–185

Miller JED, Damschen EI, Ives AR (2019) Functional traits and community composition: A

comparison among community-weighted means, weighted correlations, and multilevel

models. Methods Ecol Evol 10:415–425

Minchin PR (1987) Simulation of multidimensional community patterns: towards a

comprehensive model. Vegetatio 71:145–156

Moles AT, Perkins SE, Laffan SW, et al (2014) Which is a better predictor of plant traits:

temperature or precipitation? J Veg Sci 25:1167–1180

Moles AT, Warton DI, Warman L, et al (2009) Global patterns in plant height. J Ecol 97:923–932

Mueller-Dombois D, Ellenberg H (1974) Aims and methods of vegetation ecology. Wiley New

York, New York, USA

Muscarella R, Uriarte M (2016) Do community-weighted mean functional traits reflect optimal

strategies? Proc R Soc B Biol Sci 283:20152434

Nakagawa S, Johnson PCD, Schielzeth H (2017) The coefficient of determination R 2 and intra-

class correlation coefficient from generalized linear mixed-effects models revisited and

expanded. J R Soc Interface 14:20170213

Niinemets Ü (2001) Global-scale climatic controls of leaf dry mass per area, density, and

thickness in trees and shrubs. Ecology 82:453–469

Peres-Neto PR, Dray S, ter Braak CJF (2017) Linking trait variation to the environment: critical

issues with community-weighted mean correlation resolved by the fourth-corner

approach. Ecography (Cop) 40:806–816

Pérez-Harguindeguy N, Díaz S, Garnier E, et al (2013) New handbook for standardised

measurement of plant functional traits worldwide. Aust J Bot 61:167–234.

https://doi.org/http://dx.doi.org/10.1071/BT12225

Poorter H, Niinemets Ü, Poorter L, et al (2009) Causes and consequences of variation in leaf

mass per area (LMA): a meta-analysis. New Phytol 182:565–588

Rogers DA, Rooney TP, Hawbaker TJ, et al (2009) Paying the extinction debt in southern

Wisconsin forest understories. Conserv Biol 23:1497–1506

Rolhauser AG, Nordenstahl M, Aguiar MR, Pucheta E (2019) Community-level natural selection

modes: A quadratic framework to link multiple functional traits with competitive ability. J

Ecol 107:1457–1468. https://doi.org/10.1111/1365-2745.13094

Rolhauser AG, Pucheta E (2017) Directional, stabilizing, and disruptive trait selection as

alternative mechanisms for plant community assembly. Ecology 98:668–677.

https://doi.org/10.1002/ecy.1713

Rosbakh S, Römermann C, Poschlod P (2015) Specific leaf area correlates with temperature: new

evidence of trait variation at the population, species and community levels. Alp Bot

125:79–86

Scheiter S, Langan L, Higgins SI (2013) Next-generation dynamic global vegetation models:

learning from community ecology. New Phytol 198:957–969

Shipley B (2010) From plant traits to vegetation structure: chance and selection in the assembly

of ecological communities. Cambridge University Press, Cambridge, UK

Shipley B, De Bello F, Cornelissen JHC, et al (2016) Reinforcing loose foundation stones in trait-

based plant ecology. Oecologia 180:923–931

ter Braak CJF (2019) New robust weighted averaging-and model-based methods for assessing

trait-environment relationships. Methods Ecol Evol. https://doi.org/10.1111/2041-

210X.13278

Ter Braak CJF, Prentice IC (1988) A theory of gradient analysis. In: Advances in ecological

research. Elsevier, pp 271–317

Thuiller W, Albert C, Araújo MB, et al (2008) Predicting global change impacts on plant species'

distributions: future challenges. Perspect plant Ecol Evol Syst 9:137–152

Turgeon R (1989) The sink-source transition in leaves. Annu Rev Plant Biol 40:119–138

van Bodegom PM, Douma JC, Verheijen LM (2014) A fully traits-based approach to modeling

global vegetation distribution. Proc Natl Acad Sci 111:13733–13738

Violle C, Navas ML, Vile D, et al (2007) Let the concept of trait be functional! Oikos 116:882–892

Waller DM, Amatangelo KL, Johnson S, Rogers DA (2012) Wisconsin Vegetation Database–

plant community survey and resurvey data from the Wisconsin Plant Ecology Laboratory.

Biodivers Ecol 4:255–264

Whittaker RH (1967) Gradient analysis of vegetation. Biol Rev 42:207–264

Wright IJ, Dong N, Maire V, et al (2017) Global climatic drivers of leaf size. Science (80-)

357:917–921

Wright IJ, Reich PB, Westoby M, et al (2004) The worldwide leaf economics spectrum. Nature

428:821–827

Supporting Information

621   Appendix S1 – Supplementary Material and Methods

622   Appendix S2 – Supplementary Results

Figure legends

Figure 1. Fixed-effect structure of our GLMM approach, were we measure species i

performance in terms of its abundance in site j (A_{ij}). Conceptually, A_{ij} (white to black surface) is

modeled as an exponential function (through a log-link) of a trait (T) and an environmental

variable (E), as shown by the equation at the top. Parameters β_1 and β_2 are respectively linear

and quadratic effects of T on A, β_3 and β_4 are linear and quadratic effects of E on A, and β_5 and

β_6 are linear and quadratic effects of E on T (see details in Statistical methods). Combining the

log link with negative quadratic terms (β_2 and β_4 set to -1 and -0.5, respectively) this model

reproduces unimodal trait–abundance (orange and blue) and environment–abundance (green

and red) relationships. In this example, the result is a rectilinear negative T-E relationship (β_5

and β_6 set to -1 and 0, respectively). Dotted lines are modeled relationships at different T or E

levels while continuous lines are their counterparts projected onto 2-D planes.

Figure 2. Trait–environment relationships as estimated by the GLMM (Equation 2).

Relationships between traits (y-axes) and environmental variables (x-axes) are plotted based on

independent estimates of their the linear (β_5) and quadratic (β_6) coefficients (Table S5, ignoring

statistical significance). The observed range of traits and environmental variables are shown

below axes labels. Orange lines show significant nonlinear relationships (i.e. $\beta_6 \neq 0$), while blue

lines show relationships where only linear effects are significant. Dotted lines show marginal

643 significance (0.10<p<0.05) while grey lines show nonsignificant ones. "Interpretations" describe

644 overall patterns (with marginally significant patterns parenthesized).

646 **Figure 3**. Details of significant interactions ($p<0.01$) from the GLMM visualized for (a)

647 temperature seasonality, (b) mean annual temperature, and (c) soil sand content. Surfaces in the

648 first column show species abundance fitted as a function of trait values and environmental

649 condition (holding other variables constant at their mean value). Plots in the second column

650 show average slopes (*B* in Equation 2) of the given trait–abundance relationship (green line). *B*

651 depends on β_5 and β_6 which respectively estimate the linear and quadratic components of the

652 trait–environment relationships. Each green dot represents the overall effect of site *j*, i.e. the

653 sum of the fixed (φ_j) and the random slope associated with the corresponding trait ($b_{(T)j}$). Plots in

654 the third column show traditional community-weighted mean regressions (CWMr's) between

655 observed CWM trait values (grey dots) and observed environmental variables (Equation 3; dark

656 grey line fitted by Equation 4). Significance levels are shown for γ_1 and γ_2 (the linear and

657 quadratic terms) and overall R^2. The dotted horizontal lines indicate y=0. (Surfaces are colored

658 to enhance the perception of relief.) [Note: figure in next page]

660 **Figure 4**. Scatterplots relating test statistics from univariate CWM regressions to statistics for

661 homologous terms from the GLMM. Plot (a) compares linear terms (γ_1 and β_5) while plot (b)

662 compares quadratic terms (γ_2 and β_6). Each data point corresponds to one of the 24 trait–

663 environment relationships. The purple line is fitted to all 24 points with corresponding

664 Pearson's correlations (R) shown. Green and red rectangles show the range of test statistic

665 values failing to reject the null hypothesis under the CWM and GLMM models, respectively.

666 Codes shown in blue refer to the five outcomes presented in Material and Methods and Table

667 S4. Environmental variable and trait codes as in Fig. 2, except for p = mean annual precipitation,

668 t = mean annual temperature, s = soil sand content, n = soil nitrogen content.

670 **Figure 1**

$$A_{ij} = exp\big(\beta_0 + \beta_1 T_i + \beta_2 T_i^2 + \beta_3 E_j + \beta_4 E_i^2 + \beta_5 T_i E_j + \beta_6 T_i E_i^2\big)$$

674   **Figure 2**

Interpretations	MAP did not seem to affect trait–abundance relationships	Increasing MAT favored taller species (with cheaper leaves)	TSD favored taller species with smaller, cheaper leaves, but these effects tended to disappear as TSD increased	%Sand favored species with smaller (and tougher) leaves	%N marginally favored taller species with cheaper leaves	BA marginally favored with cheaper leaves
Vegetative Height (VH) 5-286 cm						
Leaf Size (LS) 0.05-2473 cm²						
Leaf Mass per Area (LMA) 8.8-97.3 g/m²						
Leaf Carbon Content (LCC) 37.4-48.3 %						
	Mean Annual Precipitation (MAP) 766-1004 mm/year	**Mean Annual Temperature (MAT)** 4.7-9.2 ºC	**Temperature Seasonality (TSD)** 10.2-12.1 ºC	**Soil Sand Content (%Sand)** 15-89 %	**Soil Nitrogen Content (%N)** 0.05-1.38 %	**Tree Basal Area (BA)** 7.2-51.7 m²/ha

677 **Figure 3**

679 **Figure 4**

(a)

(b)