OIKOS

Research

Integrating correlation between traits improves spatial predictions of plant functional composition

Rafael O. Wüest, Tamara Münkemüller, Sébastien Lavergne, Laura J. Pollock and Wilfried Thuiller

R. O. Wüest (http://orcid.org/0000-0001-6047-1945) (rafael.wueest@gmail.com), T. Münkemüller, S. Lavergne, L. J. Pollock and W. Thuiller (orcid.org/0000-0002-5388-5274), Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, LECA (Laboratoire d'Ecologie Alpine), FR-38000 Grenoble, France. Present address for ROW: Swiss Federal Research Institute WSL, Forest Resources and Managment, Zürcherstrasse 11, CH-8903 Birmensdorf, Switzerland.

Oikos
127: 472–481, 2018
doi: 10.1111/oik.04420

Subject Editor: Erik Aschehoug
Editor-in-Chief: Dries Bonte
Accepted 2 October 2017

Functional trait composition is increasingly recognized as key to better understand and predict community responses to environmental gradients. Predictive approaches traditionally model the weighted mean trait values of communities (CWMs) as a function of environmental gradients. However, most approaches treat traits as independent regardless of known tradeoffs between them, which could lead to spurious predictions. To address this issue, we suggest jointly modeling a suit of functional traits along environmental gradients while accounting for relationships between traits. We use generalized additive mixed effect models to predict the functional composition of alpine grasslands in the Guisane Valley (France). We demonstrate that, compared to traditional approaches, joint trait models explain considerable amounts of variation in CWMs, yield less uncertainty in trait CWM predictions and provide more realistic spatial projections when extrapolating to novel environmental conditions. Modeling traits and their co-variation jointly is an alternative and superior approach to predicting traits independently. Additionally, compared to a 'predict first, assemble later' approach that estimates trait CWMs post hoc based on stacked species distribution models, our 'assemble first, predict later' approach directly models trait-responses along environmental gradients, and does not require data and models on species' distributions, but only mean functional trait values per community plot. This highlights the great potential of joint trait modeling approaches in large-scale mapping applications, such as spatial projections of the functional composition of vegetation and associated ecosystem services as a response to contemporary global change.

Introduction

For decades, community ecology has sought general principles that govern how species vary in space and time and in response to environmental gradients. The failure to arrive at such principles has led to considerable concern, and is caused in part by the inherently complex nature of communities (Lawton 1999, Simberloff 2004). However, this complexity can be tackled by using functional traits of organisms (McGill et al.

2006). Functional traits constitute quantifiable properties that influence individual's performance and as such can help to understand and predict community structure along environmental gradients, an approach which can conveniently be applied to the increasingly large functional datasets as they become available (McGill et al. 2006, Violle et al. 2007). Beyond the link to community assembly and composition, functional traits can also be used to infer ecosystem functions (Garnier et al. 2004) and associated services (Lavorel and Garnier 2002, Lavorel and Grigulis 2012), which makes studying functional traits and/or their distribution along environmental gradients informative and important for environmental change research (Lamarque et al. 2014, Lavorel et al. 2015).

How to analyze and map the functional structure and composition of communities along environmental gradients and how they potentially change across space and time due to global change has been an important area of research. Two alternative approaches have been used so far. One approach consists of first predicting species distributions as a function of environmental variables to produce a stack of species distribution maps that are subsequently used to reconstruct spatially explicit trait compositions ("predict first, assemble later"; Ferrier and Guisan 2006). In a second approach, community-level or grid-based trait composition is directly modeled in relation to environmental predictors, and the model is then used to provide predictions over space and time ("assemble first, predict later"; Ferrier and Guisan 2006). In the former approach, the use of species distribution models (SDMs) in trait-based ecology enables one to project species distributions across space and time (Buisson et al. 2013, Thuiller et al. 2015), but the quantification of trait composition is done post hoc. In other words, structure and composition of functional traits in communities is treated as an emergent property of assembled species and is not modeled per se. An important drawback of this SDM-stacking approach is that it requires a minimal number of observations per species, which can drastically restrict the number of species retained in the analysis and hence bias ad hoc calculation of community weighted mean traits. Another limitation of the 'predict first, assemble later' approach is that species are modeled independently assuming that biotic interactions between species do not feed back on the functional trait structure of communities. Recent approaches such as modeling species simultaneously (Clark et al. 2014, Pollock et al. 2014, Harris 2015) and accounting for multiple interacting traits and environmental gradients (Pollock et al. 2012, Jamil et al. 2013, Brown et al. 2014) have the potential to improve the predictions of community composition beyond independent species models, and in doing so, improve estimates of trait composition across space. However, this new generation of approaches to model species distributions are computationally demanding when large numbers of species are involved and focus on how species, rather than traits, respond to environmental change.

By focusing on traits rather than species, the latter approach ('assemble first, predict later') is more explicit on directly modeling community-level trait characteristics in relation to environmental variables (Kühn et al. 2006). The average trait value of a community, weighted by the relative abundance of the species (community weighted mean, CWM) has been extensively used to study the functional trait structure of communities (Díaz et al. 2007) and how this structure affects ecosystem functioning (Garnier et al. 2004). The ever-increasing availability of trait data (e.g. TRY; < www.try-db.org >, BIEN; < http://bien.nceas.ucsb.edu/bien >) allows us to now study how multiple traits respond to environmental gradients and how they influence ecosystem functioning. However, current approaches are so far limited to estimating statistical relationships of such CWMs along environmental gradients independently for multiple traits (Bernard-Verdier et al. 2012, Widenfalk et al. 2015). While appealing, this practice can be problematic given that traits are not independent of each other but rather exhibit functional relationships (e.g. tradeoffs) within and between species (Díaz et al. 2004, Boucher et al. 2013). One of the most prominent examples is the leaf economic spectrum of plants, which is defined along a gradient of short-lived leaves with low dry mass per area that exhibit a high photosynthetic capacity on one end to long-lived heavy and small leaves that are photosynthetically inefficient on the other end (Wright et al. 2004). Another example is plant height, which is positively correlated with tissue density, as taller plants need mechanical stability of denser tissue to avoid breakage (Niklas 1993). Without this constraint, trees could at the same time get taller and become less dense with increasing temperature because they suffer less from freezing-induced cavitation. Naturally, this is not the case because of the trade-off between height and tissue density. In case tradeoffs scale up, an 'assemble first, predict later' approach that models the CWM of traits independently will not consider trait tradeoffs and could yield erroneous models and predictions. This is analogous to the problem in species distribution modeling, in which independent species models ignore potential interactions between species.

In this paper, we argue that the 'assemble first, predict later' approach could be substantially improved by modeling traits jointly and considering the correlation structure between traits. In comparison to 'predict first, assemble later', it offers the possibility of predicting multiple traits at the same time without computational constraints (modeling ten traits is less computationally intensive than modeling 1000 species) and directly links traits to ecosystem functioning. To do so, we suggest the use of multivariate mixed effect models as a powerful approach to model joint responses of traits (or their community weighted means) along environmental gradients. Specifically, we do this by adapting recent approaches of jointly modeling species distributions to jointly model traits along environmental gradients, and compare the performance of models that model traits independently to those that model multiple traits simultaneously, and a joint trait model that explicitly accounts for correlations between traits.

We hypothesized that joint trait models should provide more robust and more ecologically meaningful predictions than independent trait models. Further, we investigate the usefulness and advantages of the joint modeling approaches when projecting functional composition of communities in space. We here hypothesized that joint trait modeling would reduce projection uncertainty, especially when accounting for trait correlations, as this should prevent predicting communities with ecologically unrealistic trait combinations.

Material and methods

Community plots and trait data

We used extensive vegetation survey data from the French National Botanical Alpine Conservatory (Conservatoire Botanique National Alpin; CBNA) that spans the entire French Alps. CBNA plots were surveyed between 1980 and 2015 in homogeneous patches of vegetation with an average area of 100 m^2. Species nomenclature was standardized according to the 'Index synonymique de la flore de France' (Kerguélen 1993). Relative abundance within a survey plot was recorded on an ordinal scale of percentages with cutoffs at 1%, 5%, 25%, 50%, 75% (with cutoff-values being assigned to the lower class), resulting in six abundance classes. We used the mean of each abundance-class to represent the relative abundance of each species in a local plot survey.

Our study area to model trait responses along environmental gradients was the grasslands of the Guisane Valley (Fig. 1), which are situated along steep climatic gradients. The valley is 25 km long and characterized by mean annual temperatures ranging from –8.2°C to 7.8°C. We chose to restrict our study case to the grassland survey-plots of this Valley because a relatively small number of plots ameliorates interpretation of results and cuts computational effort, and because modeling non-linear changes across ecotones, e.g. from grasslands to forest, would not be informative for the aims of our study. Therefore, we included only survey plots characterized as grasslands according to the CORINE land-cover data (European Environment Agency 2013). Consequently, our analyses included only observations from the herbaceous layers of the plots (i.e. shrub and tree saplings were excluded before analysis). For each plot in the Guisane Valley, we extracted relative abundance of all occurring species from the CBNA database.

For each of the species, we extracted individual-level measurements for four traits: plant height (Height, mm), seed mass (Seedm, mg), leaf dry matter content (LDMC, mg g^{-1}), and specific leaf area (SLA, m^2 kg^{-1}). Height, Seedm and SLA were chosen to represent the leaf-height-seed (LHS) plant strategy scheme, which depicts general plant life strategies (Westoby 1998). LHS traits are especially well-suited for our study because the LHS encompasses trait tradeoffs both within and between traits (Westoby 1998), leading to well-documented correlations among LHS traits (Díaz et al. 2015). We included LDMC in our study because LDMC is likely negatively correlated with SLA. Traits were extracted mostly from our own database of trait measurements in the Alps (Thuiller et al. unpubl.), complemented with data from LEDA (Knevel et al. 2003), BioFlor (Kühn et al. 2004), Ecoflora (Fitter and Peat 1994) and CATMINAT (Julve 1998). The individual-level trait measurements with a mean of 4.1 (± 1.5) observations per species were averaged to obtain a mean trait value for each species.

Some species were missing trait data, so we restricted the dataset used in our analyses to only those plots where at least 80% of the plot was covered by species with complete trait information. This is justified by the biomass ratio hypothesis postulating that community level traits and functions are determined by species that dominate the biomass of a community (Grime 1998). Based on this selection, we used 108 plots and 432 species overall. For each plot, we calculated the community weighted mean (CWM) trait value for each of the four traits. CWM averages species level traits weighted by abundance. We first log-transformed CWMs in order to approximate a normal distribution of the data and then standardized the CWMs (each trait separately) by subtracting the

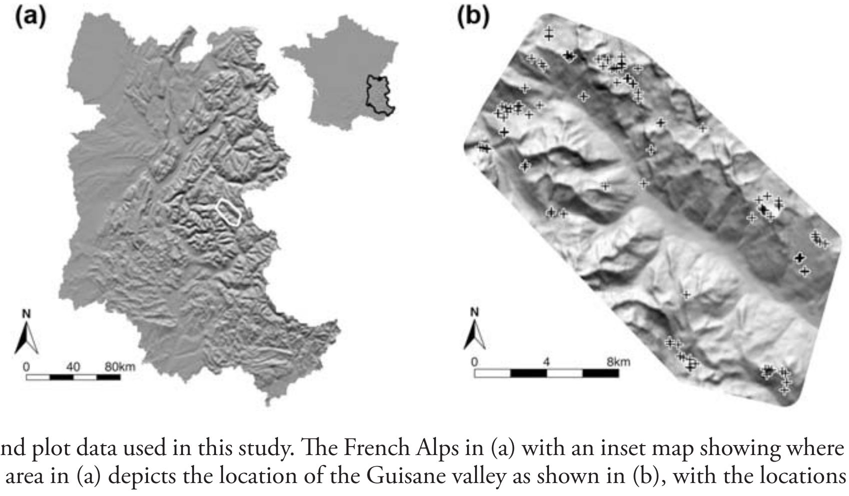

Figure 1. Study regions and plot data used in this study. The French Alps in (a) with an inset map showing where the study region is located within France. The white area in (a) depicts the location of the Guisane valley as shown in (b), with the locations of the 108 grassland plots included our study.

arithmetic mean and dividing by the standard deviation in order to facilitate convergence of the models.

Modeling traits along environmental gradients

We selected ecologically meaningful topo–climatic variables at a spatial resolution of 100 × 100 m that have previously been shown to be important to explaining the functional structure of the Guisane Valley vegetation (de Bello et al. 2013, Chalmandrier et al. 2015). The set of variables included mean annual precipitation (MAP, mm), coldest temperature (TMIN, °C), relative humidity (RELH, %), and topographic position (TOPO). The climatic variables were derived from downscaling the gridded 1km climate data from MeteoFrance (Benichou and Breton 1987) in the case of MAP and TMIN, and from global CRU data (New et al. 2002) in the case of RELH. Topographic position assesses concavity vs. convexity of a given location with respect to the surrounding landscape, with positive values indicating that the focal pixel is higher than the surrounding (vice versa for negative values). We chose these predictors from a larger set of environmental variables to minimize collinearity (all Pearson's correlations <0.55) and maximize ecological interpretation. We scaled the predictor variables by subtracting the mean and dividing by the standard deviation prior to model-fitting.

We investigate whether joint modeling, and accounting for correlation structure improves the estimation of trait–environment relationships by comparing independent trait models (ITMs) with a multiple trait model (MTM) and a joint trait model (JTM) that explicitly accounts for correlations between traits. While large-scale analyses inform on the general direction of some traits along some of the selected environmental gradients (Wright et al. 2004, Poorter et al. 2009), there are no well-defined theoretical expectations about the exact shape (e.g. linear, curvilinear, exponential) of the response of the four traits to our local environmental gradients. We thus used generalized additive models (GAMs) that do not require predefined response curve shapes along predictor variables (Hastie and Tibshirani 1990). They use a class of equations called 'smoothers' that attempt to generalize data into smooth curves by local fitting to subsections of the data. More technically, we used thin plate regression to estimate the smoothers for the four predictor variables and a tensor product to generate a smoother for an interaction term between MAP and TOPO to account for the fact that effects of precipitation may change with topography due to differential water retention on ridges versus in depressions. The thin plate and tensor product smoothers were estimated using the mgcv package (Wood 2011) in the R statistical environment (ver. 3.3.2; <www.r-project.org>). A potential downside of GAMs is that they sometimes overfit, i.e. the estimated response could get very close to the data and potentially ignores biological realism. We minimized overfitting by allowing for a maximum of five degrees of freedom in the smoothers to avoid overly complex responses. In addition, we also fitted GLMs with linear and quadratic terms for all predictors (and including the interaction between MAP and TOPO) to test whether ITM performance was influenced by the flexibility of response curves in GAMs.

In the ITM approach, we modeled traits independently of each other, as traditionally done in functional ecology (Bernard-Verdier et al. 2012, Widenfalk et al. 2015). In the simplest case of one predictor variable, the model equation for one trait would be as follows:

$$y_i = b + f\left(x_i\right) + \varepsilon_i \tag{1}$$

where each element of the response y_i is the CWM of a trait in plot i, b is an intercept and f represents the smooth term of a predictor variable x. ε_i represents the Gaussian residuals. In the case of several predictors Eq. 1 extends to:

$$y_i = b + f_p\left(x_{ip}\right) + \varepsilon_i \tag{2}$$

where f_p indicates the predictor-specific smooth terms. The model equation for the MTM and JTM further extends Eq. 2 to:

$$y_{ij} = b_j + f_{pj}\left(x_{ipj}\right) + \varepsilon_{ij} \tag{3}$$

where b_j indicates a specific intercept for each trait, f_{pj} indicates that for each predictor p a distinct smooth is fitted for each trait j. Technically, trait-specific responses are achieved by trait-specific random effects on the smooth-term parameters, leading to:

$$y_{ij} = b_j + f\left(x_i\right) + u_{pj} + \varepsilon_{ij} \tag{4}$$

where u_{pj} represents the random effects that enable the model to fit specific smoothers for each trait j along each predictor p. Random effects for each smoother in the MTM are constrained to stem from a normal distribution with expected value zero. The JTM additionally accounts for between-trait correlation by fitting an unstructured correlation structure on u_{pj} with a different parameter for every possible pair of traits.

The 'mgcv' package allows one to fit ITMs, MTMs and JTMs and we include the code to fit these models in the Supplementary material Appendix 2. Note that the correlation structure needs to refer to *plotID*, a factor with a unique ID for each of the *i* plots and its form has to be symmetric (*corSymm*) in order to obtain a pair-wise between-trait correlation structure.

We assessed goodness of fit in both approaches using root mean square error (RMSE) of predicted versus observed trait CWM-values, and R^2 of the regression between predicted and observed values (R^2_{CORR}). For both measures, we calculated a pooled value where the pooled predicted values of all four traits were compared to the pooled observations of all traits and a trait-specific value where we compared predicted versus observed values separately for each trait. Further, as commonly done for SDMs, we applied a repeated split-sampling procedure to assess performance of the models. The split-sampling procedure consisted of a random splitting of original data into two thirds training and one third testing data, with 20 repetitions. Thus, in each repetition the models were fitted on training data, and evaluated against the testing data. Analogous to the full-model comparison, we evaluated model performance in the split-sampling procedure by assessing RMSE and R^2_{CORR} on the hold-out testing data pooled across the four traits and the 20 repetitions. In addition, we compared the correlation structure as estimated in the JTM with the observed correlation between trait CWMs in the Guisane Valley plots in terms of RMSE and bias (mean of predicted minus observed).

Trait–environment relationships

We produced partial response curves along the environmental gradients for all traits in order to facilitate identification and interpretation of differences between the modeling approaches. The partial response curve of a trait along a predictor variable describes how a trait changes along that variable and is produced by calculating the trait's predicted values along the focal variable, while keeping all other variables fixed at their mean.

Spatial projections

We used the fitted ITMs and JTMs built from the sampled communities to project the spatial distribution of each trait over the whole Guisane valley at a spatial resolution of 100 × 100 m. We assess the plausibility of predictions by comparing the range of observed trait CWMs with predictions from both the ITM and JTM approaches.

Data deposition

Data available from the Dryad Digital Repository: <http://dx.doi.org/10.5061/dryad.v475g> (Wüest et al. 2017).

Results

Trait–environment relationships – model comparison

All models (ITMs, the MTM and the JTM) explained considerable amounts of variation in CWMs ($R^2_{CORR} \geq 0.31$). ITMs performed best when assessing performance by comparing predicted and observed data across all traits. RMSE across ITMs (0.73) was lower than RMSE of the MTM (0.80) and the JTM (0.83), indicating that ITM predictions best approximated the observed trait values, corroborated by a higher R^2_{CORR} (0.47 across all ITMs versus 0.35 for the MTM and 0.31 for the JTM; Table 1). Differences in trait-specific RMSE of the independent models of the ITM approach indicate that vegetation height (0.63) and seed mass (0.57) are potentially more strongly driven by our selected environmental variables than LDMC (0.82) and SLA (0.86). A complete comparison of trait-specific RMSE and R^2_{CORR} values is provided in Supplementary material Appendix 1 Table A1.

Table 1. Model performance metrics for the ITMs, the MTM, and the JTM that accounts for the correlation structure among traits. RMSE and R^2_{CORR} are calculated across all four traits in all approaches. Percentages in parentheses indicate how much performance drops (i.e. how much RMSE increases and how much R^2_{CORR} decreases) in the split-sample validation. Metrics for the best performing approach in bold.

	Predicted versus observed		Split-sample validation	
	RMSE	R^2_{CORR}	RMSE	R^2_{CORR}
ITM	**0.73**	**0.47**	1.20 (64%)	<0.01 (99%)
MTM	0.80	0.35	0.98 (24%)	**0.11** (69%)
JTM	0.83	0.31	**0.98 (19%)**	0.10 **(68%)**

Assessing the predictive performance using repeated split-sampling revealed that ITM's performance collapsed markedly, where RMSE increased to 1.20 and R^2_{CORR} decreased to virtually zero. In contrast, predictive performance of the joint models did not decrease analogously in the split-sampling validation. The increase in RMSE was at highest 24% (for MTM; 19% for JTM; Table 1) and much lower compared to the ITM (where RMSE increased by 64%). R^2_{CORR} of the JTM approach was reduced by 68% (for JTM; 69% for JTM; Table 1) in the split-sampling procedure but was far from approaching zero, as observed in the ITM approach. Improved predictive performance of a JTM based on GLMMs was comparable to the presented results based on GAMMs and are reported in Supplementary material Appendix 1 Table A2.

We compared the pair-wise correlation between traits as estimated by the JTM approach with the observed correlations of CWMs in the plots of the Guisane valley (Table 2) in order to assess how much of the observed correlation structure is not explained by the predictors and thus mirrors in the residual structure. The low RMSE (0.10) and a bias close to zero (–0.03) indicated that the correlation structure as estimated in the JTM approach was close to the observed among-trait correlations. Directional trends for all trait-correlations were congruent between estimated and observed

Table 2. Pearson's correlation coefficients among traits as observed in plots of the Guisane valley (based on CWMs) and as estimated by the JTM.

	Seed mass	LDMC	SLA	RMSE[1]
Observed				
height	0.53	–0.13	0.28	
seed mass		–0.21	0.32	
LDMC			–0.31	
Estimated (JTM)				0.10
height	0.45	–0.18	0.10	
seed mass		–0.20	0.23	
LDMC			–0.32	

[1] compared to observed correlations.

correlations and a regression between observed and estimated correlations revealed that the intercept and slope do not differ from zero and one, respectively (details in Supplementary material Appendix 1 Fig. A1).

Response curves of the ITM and the MTM/JTM approaches differed in two aspects. First, ITM response curves were in general more complex than MTM and JTM response curves. For example, the response curves of SEEDM along all environmental variables took complex, data-driven shapes in the ITM approach, while except for relative humidity, both the MTM and the JTM approach yielded more linear responses along all gradients. Second, compared to the ITMs, prediction uncertainty was reduced when traits were modeled simultaneously using the JTM approaches. This manifested as reduced confidence intervals around the response curves. For example, confidence intervals in the case of SLA along minimum temperature were larger for the ITM compared to MTM and JTM. Even though less pronounced, the same tendencies also held for the comparison between MTM and JTM, where accounting for correlations between traits in JTM often led to simpler responses with less uncertainty associated to the predictions. Figure 2 illustrates the general findings, while partial response curves for all traits along the four environmental gradients

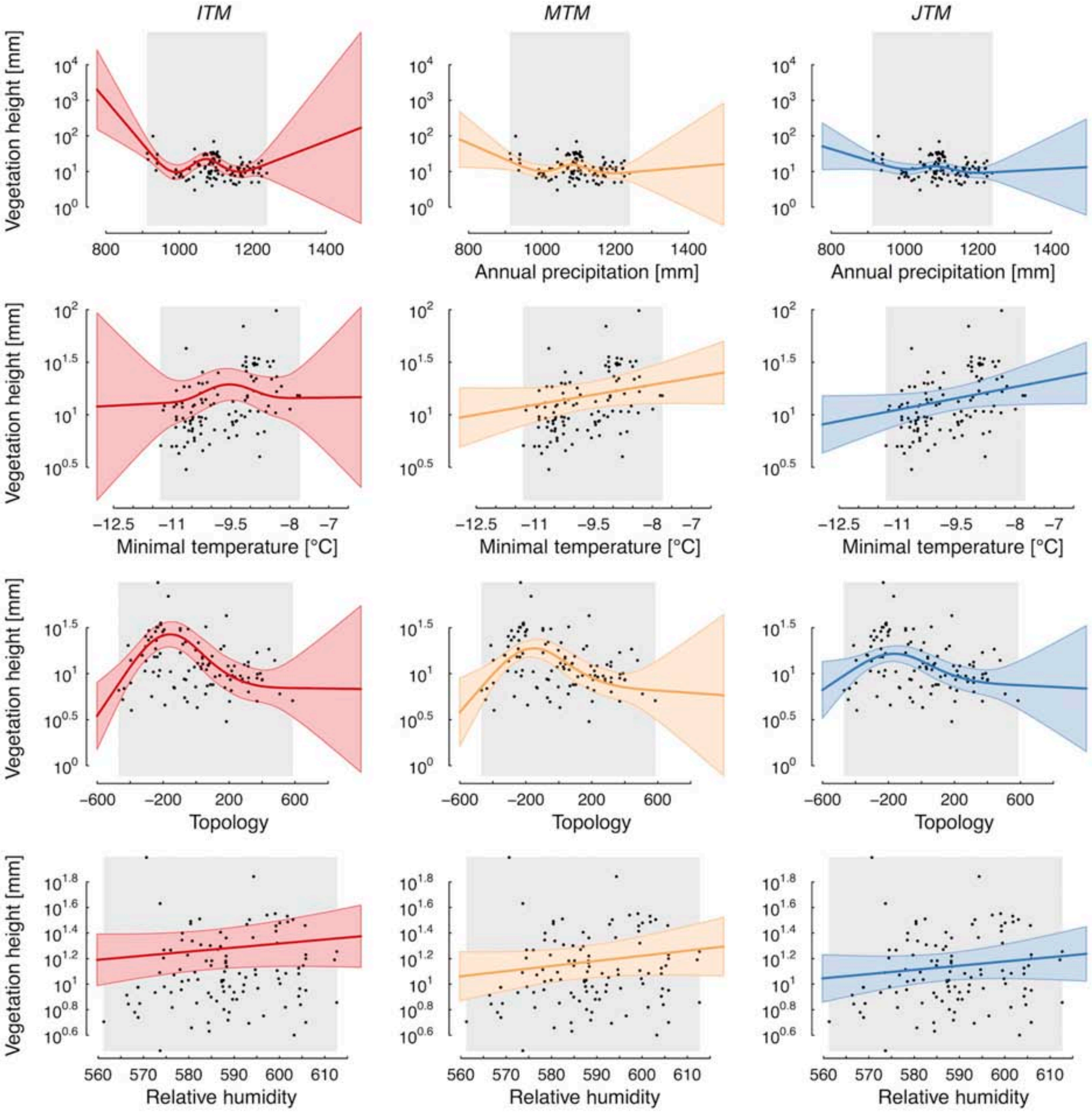

Figure 2. Response curves of ITMs (red), the MTM (orange), and the JTM (blue) along the four environmental predictors for vegetation height. Solid lines represent predictions based on estimates, colored areas include the 95% confidence intervals. The grey background shading indicates environmental conditions that exceed the range spanned by the sample-locations used to fit the models (black dots).

are presented in Supplementary material Appendix 1 Fig. A2–A4.

Spatial projections

Spatial projections of the three modeling approaches showed that the JTM best approximates the range of observed CWMs. While ITMs tended to predict trait values outside the range of observed values that are sometimes unrealistic, the MTM and JTM approaches did not suffer from this problem to the same extent. The example of height (Fig. 3) demonstrates that the over- and under-prediction of ITMs and the MTM appeared mostly but not exclusively in areas of extrapolation, thus in areas with environmental characteristics that were not covered by the data used to train the models (black polygons in Fig. 3a–c). Spatial projections for all traits are presented in Supplementary material Appendix 1 Fig. A5.

Discussion

We show that it is possible to predict the functional composition of alpine grassland communities, and that models that model multiple traits simultaneously outperform independent modeling of traits in several aspects. We further show that accounting for between-trait correlation in joint models further improves predictions, especially when extrapolating to novel environmental conditions. These findings suggest that traits should be modeled jointly rather than in isolation, and that between-trait correlation should be accounted for when modeling functional attributes of communities.

Trait–environment relationships – model comparison

We find that the joint trait model JTM outperforms the multiple trait model MTM and the independent trait models ITMs when predicting functional attributes of plant communities across the Guisane valley. ITMs clearly perform worst and while both approaches that model multiple traits simultaneously increased predictive performance, the difference between the MTM and the JTM that additionally accounts for between-trait correlations is rather marginal. The comparison of partial response curves (Fig. 2) helps to identify reasons for the superior performance of the MTM and JTM approaches over the ITM approach in predicting CWM trait values. ITMs generally yielded complex response curves along all environmental predictors (red curves in Fig. 2), which lead to a break down in ITM cross-validation performance (Table 1). Simultaneous modeling of traits in the MTM and JTM approaches restricts the complexity of response curves

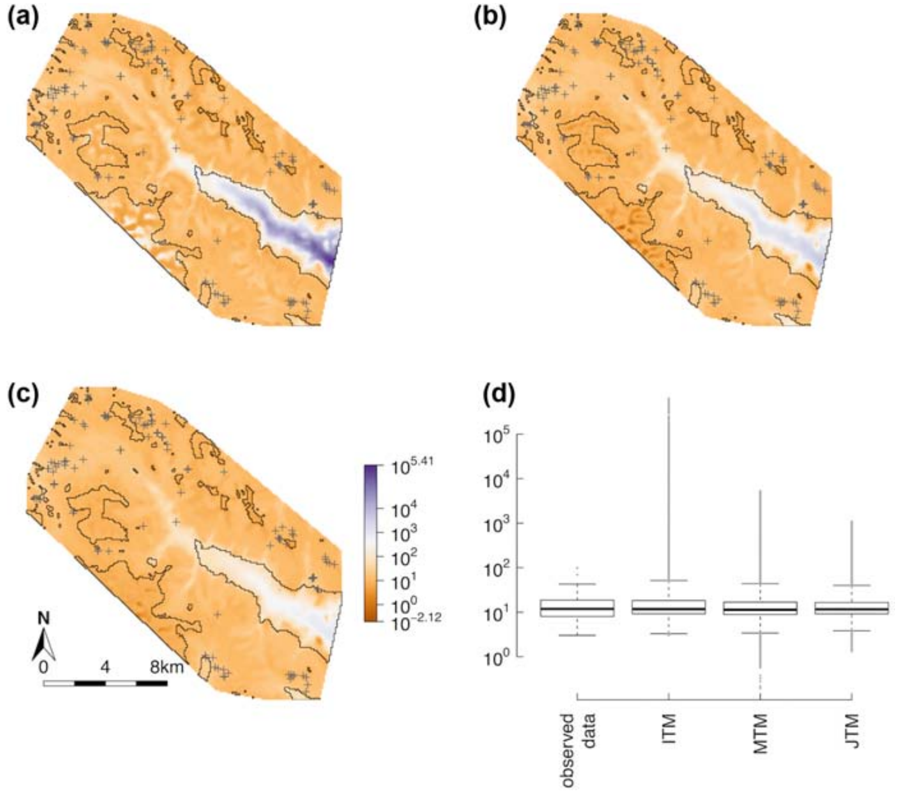

Figure 3. Spatial projection of plant height across the Guisane valley. Projection of ITMs are shown in (a), the MTM projection in (b), and the JTM projection in (c) with colors indicating plant height on a log-scale according to the legend. Grey crosses represent locations of plots used for fitting the models. Black polygons depict areas where at least one of the environmental predictors reaches outside the observed (data-fitting) range. The boxplots in (d) describe the range of observed and predicted CWM values.

and yields more constrained and linear responses (orange and blue curves in Fig. 2), which yield more realistic predictions when extrapolating (Merow et al. 2014). Trait-models based on GLMs showed the same pattern as those based on GAMs: the JTM yielded better predictions in cross-validation than ITMs (Supplementary material Appendix 1 Table A2). Therefore, the complexity of the response is not the only explanation for improved predictions, because GLMs have simpler response curves and are less prone to over-fitting.

Joint modeling may be superior to independent modeling because the modeled trait responses potentially profit from each other, an argument that is put forward in the field of SDMS: rare species are better predicted in JSDMs compared to species-specific SDMs (Ovaskainen and Soininen 2011). Adopted to the case of modeling traits, this could mean that the observed data may not sufficiently well sample the environmental space to effectively approximate the biological reality of a specific trait-response in ITMs, but the response fitted by a joint model may get closer to biological reality in joint models because it is correlated with the responses of other traits. The fact that the JTM that explicitly accounts for correlations between traits performs better than the MTM supports this interpretation. More generally, modeling multiple traits simultaneously likely is superior to ITMs because it accounts for the fact that species' presence and abundance in a given location is not determined by single traits, but rather results from the interplay of multiple trait axes (Muscarella and Uriarte 2016). Indeed, while accounting for trait–environment relationships the JTM estimates a correlation structure that is very similar to the observed between-trait correlations (Supplementary material Appendix 1 Fig. A1). Taken together, this suggests that inherent trait tradeoffs are indeed affecting assembly and functional composition of communities and should by no means be ignored in 'assemble-first, predict later' approaches.

Improved validation-performance of the JTM approach parallels findings from species distribution modeling, where JSDM approaches exceed performance of independent SDMs across a number of multi-species data sets (Clark et al. 2014, Harris 2015). In addition to the reduction in complexity, the JTM approach also yields less uncertainty in parameter-estimates. While reduced confidence intervals (CIs) are visible across the entire range of the environmental variables, differences are greatest when extrapolating beyond the observed data-range (e.g. height along annual precipitation, Fig. 2). In summary, the limited complexity in response curves that are associated with less uncertainty suggests that the modeling approaches that model multiple traits simultaneously, and the JTM in particular, hold much promise for predicting functional attributes of communities and should be favored in any predictive modeling endeavor that involves projecting in space or time.

Spatial projections and extrapolation

Spatial projections of the ITMs, the MTM, and the JTM across the Guisane valley (Fig. 3) illustrate the differences between the three approaches. ITMs often predict extreme trait values that exceed observed values by orders of magnitude in both directions for both height and seed mass. For example, independent trait modeling predicts grasslands grow taller than 100 m (Fig. 2). Seed masses in the same area are predicted to be heavier than 100 g (Supplementary material Appendix 1 Fig. A5); more than 1000 times the maximal seed mass observed for grassland species in the Guisane valley. Spatial projections derived from the MTM and JTM do not suffer from this problem. The predicted ranges of trait values more closely approximate the observed ranges and the models rarely predict unrealistic values (Fig. 3b–d, Supplementary material Appendix 1 Fig. A5). The JTM also predicts more realistic trait CWMs than the MTM approach. For example, MTM predicts grasslands at the bottom of the Guisane Valley to grow as tall as 5 m, while JTM predicts a maximal height of 1 m. Nevertheless, neither of the two approaches that model multiple traits simultaneously avoids all potential issues regarding extrapolation and we urge the need to carefully assess the plausibility of predictions when extrapolating.

It is evident that the extreme predictions are largely restricted to areas where environmental conditions exceed conditions covered by data used to fit the models (black polygons in Fig. 3a–c, Supplementary material Appendix 1 Fig. A5). Responses along most gradients are curvilinear in ITMs (Fig. 2, Supplementary material Appendix 1 Fig. A2–A4), and these flexible responses offer a possible explanation for the extreme predictions. Depending on the shape, extrapolation beyond the observed data range can quickly lead to extreme values outside the observed range (see the partial response of LDMC along minimal temperature in Supplementary material Appendix 1 Fig. A3 for an extreme example). Extrapolation is well known to involve various potential pitfalls and should be avoided whenever possible. However, on-going global change likely causes the appearance of novel climates (Williams et al. 2007). Consequently, the joint modeling of traits along environmental gradients that yields more conservative predictions when extrapolating is, therefore, the preferred approach.

Limitations and further development

Our results indicate that joint modeling should be preferred over independent modeling of traits along environmental gradients. Nevertheless, our approach is not free from limitations. We argued that the reduction of data dimensionality in a 'assemble first, predict later' approach leads to higher efficiency because of the reduced number of required model-parameters. While this is certainly true for species-rich and trait-poor datasets, the number of model-parameters rapidly increases with the number of traits to be modeled. The increase in parameters will be especially drastic in the case of JTM, where the number of pair-wise trait combinations increases quadratically with the number of traits. We suggest considering the MTM approach instead of JTM for datasets with many traits, or using hierarchical latent variable models to reduce trait dimensionality (Hui 2015, Warton et al. 2015).

While we compare our findings with recent developments in the field of (joint) SDMs, our analysis did not aim to compare our 'assemble first, predict later' approach with the various flavors of 'predict first, assemble later' approaches. Potential future studies comparing both performance and computational efficiency of the various approaches for a range of conditions (i.e. varying numbers of species, traits, and sites) could help to determine the optimal approach.

Our MTM and JTM approaches as currently implemented use GAMMs and are therefore restricted to continuous traits. However, many functional traits, such as growth form of plants or diet of animals are categorical in nature. Categorical (or mixed) traits could potentially be jointly modeled using mixed modeling software that allows for categorical (or mixed) responses, such as MCMCglmm in R (Hadfield 2010) or using the various implementations of BUGS-like languages. While this is outside the scope of our study, we encourage further research to enable inclusion of categorical traits in joint models that follow the 'assemble first, predict later' approach. Another potential avenue of investigations not covered in our initial assessment of the approach is how phylogenetic signal in traits interacts with independent and joint estimation of trait responses, because we know that phylogenetic signal affects the estimation of how functional traits interact with the environment (Li and Ives 2017).

Implications

Traditional approaches to the so-called fourth corner problem (understanding how functional traits mediate species-specific environmental responses) like RLQ analyses do not provide the possibility for spatial projections. While joint species distribution models overcome this limitation (Pollock et al. 2012, Jamil et al. 2013, Brown et al. 2014), they are generally computationally intensive for large datasets (hundreds or thousands of species). While latent variable models improve the capacity to model many species (Warton et al. 2015), these models still require spatial or environmental data for all species, which is not always available. Direct modeling of traits along environmental gradients in a 'assemble first, predict later' manner may be debatable (Clark 2016), but it represents the basis of JTM's efficiency, which has great potential in large-scale applications like mapping ecosystem services in space. Lavorel et al. (2011), for example, relate a set of traits including height and LDMC to the environment to quantify ecosystem services. Such an approach is sensitive to spurious combinations of height-LDMC predictions and would obstruct correct predictions of ecosystem services. Simultaneous modeling of multiple traits, and especially the JTM approach, avoids predicting spurious trait combinations, thus is better suited to inform approaches that predict ecosystem services – an important tool in analyzing and mitigating effects of global change on human well-being (Díaz et al. 2015). Additionally, spatial projections of trait CWMs like vegetation height or leaf traits such as silica content could provide proximal predictors for SDMs of herbivores, ranging from large ungulates to insects. Furthermore, fitted trait–environment relationships could be used to drive trait responses to changing environments in earth system models (Pavlick et al. 2013, Verheijen et al. 2013).

Acknowledgements – Funding – This work was funded by the ANR project Origin-Alps (ANR-16-CE93-0004) and by a grant from Labex OSUG@2020 (Investissements d'avenir – ANR10 LABX56). ROW received funding from the Swiss National Science Foundation (grant no. 147226).

References

Benichou, P. and Breton, O. 1987. Prise en compte de la topographie pour la cartographie des champs pluviométriques statistiques. – Météoorologie 7: 23–34.

Bernard-Verdier, M. et al. 2012. Community assembly along a soil depth gradient: contrasting patterns of plant trait convergence and divergence in a Mediterranean rangeland. – J. Ecol. 100: 1422–1433.

Boucher, F. C. et al. 2013. Unravelling the architecture of functional variability in wild populations of *Polygonum viviparum* L. – Funct. Ecol. 27: 382–391.

Brown, A. M. et al. 2014. The fourth-corner solution – using predictive models to understand how species traits interact with the environment. – Methods Ecol. Evol. 5: 344–352.

Buisson, L. et al. 2013. Toward a loss of functional diversity in stream fish assemblages under climate change. – Global Change Biol. 19: 387–400.

Chalmandrier, L. et al. 2015. Effects of species' similarity and dominance on the functional and phylogenetic structure of a plant meta-community. – Ecology 96: 143–153.

Clark, J. S. 2016. Why species tell more about traits than traits about species: predictive analysis. – Ecology 97: 1979–1993.

Clark, J. S. et al. 2014. More than the sum of the parts: forest climate response from joint species distribution models. – Ecol. Appl. 24: 990–999.

de Bello, F. et al. 2013. Hierarchical effects of environmental filters on the functional structure of plant communities: a case study in the French Alps. – Ecography 36: 393–402.

Díaz, S. et al. 2004. The plant traits that drive ecosystems: evidence from three continents. – J. Veg. Sci. 15: 295–304.

Díaz, S. et al. 2007. Incorporating plant functional diversity effects in ecosystem service assessments. – Proc. Natl Acad. Sci. USA 104: 20684–20689.

Díaz, S. et al. 2015. The global spectrum of plant form and function. – Nature 529: 167–171.

European Environment Agency 2013. Corine Land Cover 2006 raster data. – European Environment Agency, Copenhagen, Denmark.

Ferrier, S. and Guisan, A. 2006. Spatial modelling of biodiversity at the community level. – J. Appl. Ecol. 43: 393–404.

Fitter, A. H. and Peat, H. J. 1994. The Ecological Flora Database. – J. Ecol. 82: 415.

Garnier, E. et al. 2004. Plant functional markers capture ecosystem properties during secondary succession. – Ecology 85: 2630–2637.

Grime, J. P. 1998. Benefits of plant diversity to ecosystems: immediate, filter and founder effects. – J. Ecol. 86: 902–910.

Hadfield, J. D. 2010. MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. – J. Stat. Software 33: 1–22.

Harris, D. J. 2015. Generating realistic assemblages with a joint species distribution model. – Methods Ecol. Evol. 6: 465–473.

Hastie, T. J. and Tibshirani, R. J. 1990. Generalized additive models. – Chapman and Hall/CRC.

Hui, F. K. C. 2015. boral – Bayesian ordination and regression analysis of multivariate abundance data in R. – Methods Ecol. Evol. 7: 744–750.

Jamil, T. et al. 2013. Selecting traits that explain species–environment relationships: a generalized linear mixed model approach. – J. Veg. Sci. 24: 988–1000.

Julve, P. 1998. Baseflor. Index botanique, écologique et chorologique de la flore de France. – Inst. Catholique de Lille, Lille, France.

Kerguélen, M. 1993. Index synonymique de la flore de France. – Mus. Natl d'Hist. Nat. Secrétariat de la Faune et de la Flore.

Knevel, I. C. et al. 2003. Life-history traits of the northwest European flora: the LEDA database. – J. Veg. Sci. 14: 611–614.

Kühn, I. et al. 2004. BiolFlor – a new plant-trait database as a tool for plant invasion ecology. – Divers. Distrib. 10: 363–365.

Kühn, I. et al. 2006. Relating geographical variation in pollination types to environmental and spatial factors using novel statistical methods. – New Phytol. 172: 127–139.

Lamarque, P. et al. 2014. Plant trait-based models identify direct and indirect effects of climate change on bundles of grassland ecosystem services. – Proc. Natl Acad. Sci. USA 111: 13751–13756.

Lavorel, S. and Garnier, E. 2002. Predicting changes in community composition and ecosystem functioning from plant traits: revisiting the Holy Grail. – Funct. Ecol. 16: 545–556.

Lavorel, S. and Grigulis, K. 2012. How fundamental plant functional trait relationships scale-up to tradeoffs and synergies in ecosystem services. – J. Ecol. 100: 128–140.

Lavorel, S. et al. 2011. Using plant functional traits to understand the landscape distribution of multiple ecosystem services. – J. Ecol. 99: 135–147.

Lavorel, S. et al. 2015. Ecological mechanisms underpinning climate adaptation services. – Global Change Biol. 21: 12–31.

Lawton, J. H. 1999. Are there general laws in ecology? – Oikos 84: 177–192.

Li, D. and Ives, A. R. 2017. The statistical need to include phylogeny in trait-based analyses of community composition. – Methods Ecol. Evol. 8: 1192–1199.

McGill, B. J. et al. 2006. Rebuilding community ecology from functional traits. – Trends Ecol. Evol. 21: 178–185.

Merow, C. et al. 2014. What do we gain from simplicity versus complexity in species distribution models? – Ecography 37: 1267–1281.

Muscarella, R. and Uriarte, M. 2016. Do community-weighted mean functional traits reflect optimal strategies? – Proc. R. Soc. B 283: 20152434.

Supplementary material (available online as Appendix oik-04420 at <www.oikosjournal.org/appendix/oik-04420>). Appendix 1–2.

New, M. et al. 2002. A high-resolution data set of surface climate over global land areas. – Clim. Res. 21: 1–25.

Niklas, K. 1993. Influence of tissue density-specific mechanical properties on the scaling of plant height. – Ann. Bot. 72: 173–179.

Ovaskainen, O. and Soininen, J. 2011. Making more out of sparse data: hierarchical modeling of species communities. – Ecology 92: 289–295.

Pavlick, R. et al. 2013. The Jena diversity-dynamic global vegetation model (JeDi-DGVM): a diverse approach to representing terrestrial biogeography and biogeochemistry based on plant functional tradeoffs. – Biogeosciences 9: 4137–4177.

Pollock, L. J. et al. 2012. The role of functional traits in species distributions revealed through a hierarchical model. – Ecography 35: 716–725.

Pollock, L. J. et al. 2014. Understanding co-occurrence by modelling species simultaneously with a joint species distribution model (JSDM). – Methods Ecol. Evol. 5: 397–406.

Poorter, H. et al. 2009. Causes and consequences of variation in leaf mass per area (LMA): a meta-analysis. – New Phytol. 182: 565–588.

Simberloff, D. 2004. Community ecology: is it time to move on? – Am. Nat. 163: 787–799.

Thuiller, W. et al. 2015. From species distributions to meta-communities. – Ecol. Lett. 18: 1321–1328.

Verheijen, L. M. et al. 2013. Impacts of trait variation through observed trait-climate relationships on performance of an Earth system model: a conceptual analysis. – Biogeosciences 10: 5497–5515.

Violle, C. et al. 2007. Let the concept of trait be functional! – Oikos 116: 882–892.

Warton, D. I. et al. 2015. So many variables: joint modeling in community ecology. – Trends Ecol. Evol. 30: 766–779.

Westoby, M. 1998. A leaf-height-seed (LHS) plant ecology strategy scheme. – Plant Soil 199: 213–227.

Widenfalk, L. a. et al. 2015. Spatially structured environmental filtering of collembolan traits in late successional salt marsh vegetation. – Oecologia 179: 537–549.

Williams, J. W. et al. 2007. Projected distributions of novel and disappearing climates by 2100 AD. – Proc. Natl Acad. Sci. USA 104: 5738–5742.

Wood, S. N. 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. – J. R. Stat. Soc. 73: 3–36.

Wright, I. J. et al. 2004. The worldwide leaf economics spectrum. – Nature 428: 821–827.

Wüest, R. O. et al. 2017. Data from: Integrating correlation between traits improves spatial predictions of plant functional composition. – Dryad Digital Repository, <http://dx.doi.org/10.5061/dryad.v475g>.