Incorporating dominant species as proxies for biotic interactions strengthens plant community models

Peter C. le Roux[1, 2]†, Loïc Pellissier[3]†, Mary S. Wisz[3, 4, 5] and Miska Luoto[1]

[1] Department of Geosciences and Geography, University of Helsinki, FI-00014 Helsinki, Finland.

[2] Department of Plant Science, University of Pretoria, 0028, Pretoria, South Africa.

[3] Arctic Research Centre, Department of Bioscience, Aarhus University, DK-4000 Roskilde, Denmark.

[4] Greenland Climate Research Centre, Greenland Institute of Natural Resources, Nuuk, Greenland.

[5] Department of Ecology & Environment, DHI Water & Environment, 2970 Hørsholm, Denmark.

† authors contributed equally to the manuscript

*Correspondence author. E-mail: peter.leroux@up.ac.za.

Running title: Biotic interactions and plant community models

Summary

1. Biotic interactions exert considerable influence on the distribution of individual species and should, thus, strongly impact communities. Implementing biotic interactions in spatial models of community assembly is therefore essential for accurately modelling assemblage properties. However, this remains a challenge due to the difficulty of detecting the role of species interactions and because accurate paired community and environment datasets are required to disentangle biotic influences from abiotic effects.

2. Here, we incorporate data from three dominant species into community-level models as a proxy for the frequency and intensity of their interactions with other species and predict emergent assemblage properties for the co-occurring sub-dominant species. By analysing plant community and field-quantified environmental data from specially designed and spatially replicated monitoring grids, we provide a robust *in vivo* test of community models.

3. Considering this well-defined and easily quantified surrogate for biotic interactions consistently improved realism in all aspects of community models (community composition, species richness and functional structure), irrespective of modelling methodology.

4. Dominant species reduced community richness locally and favoured species with similar leaf dry matter content. This effect was most pronounced under conditions of high plant biomass and cover, where stronger competitive impacts are expected. Analysis of leaf dry matter content suggests that this effect may occur through efficient resource sequestration.

5. *Synthesis.* We demonstrate the strong role of dominant species in shaping multiple plant community attributes, and thus the need to explicitly include interspecific interactions to achieve robust predictions of assemblage properties. Incorporating information on biotic interactions strengthens our capacity not only to predict the richness and composition of communities, but also how their structure and function will be modified in the face of global change.

Key-words: abiotic gradients, biotic interactions, community composition, determinants of plant community diversity and structure, functional structure, leaf dry matter content, species distribution modelling, species richness

Introduction

Predicting the structure and function of species assemblages along environmental gradients is a fundamental goal of ecology and conservation science, but is a complex challenge (Mokany & Ferrier 2011; Nogués-Bravo & Rahbek 2011). Species are not independent entities, and can interact positively (e.g. facilitation) and/or negatively (e.g. competitive exclusion) with other species, and the search for principles explaining the local assembly of communities should therefore integrate the effect of these interactions (Lortie *et al*. 2004; Weiher *et al*. 2011). In the same way that abiotic conditions can limit species distributions, biotic interactions may contribute to determining individual species ranges and, as a result, impact on spatial variation in species assemblages (Wisz *et al*. 2013). Thus, a modelling framework that accounts for both abiotic conditions and biotic interactions in species assembly processes could address some critical gaps in our understanding of biodiversity patterns. Indeed, improved biodiversity forecasts from such models would be particularly valuable for conservation efforts, specifically given the rapid shifts observed in species assemblages under recent global environmental changes (Lenoir *et al*. 2008, Elmendorf *et al*. 2012). However, to date, the problem of accurately and reliably detecting the influence of species interactions in community assembly has limited our ability to predict species assemblages in space (Götzenberger *et al*. 2012).

Community assembly has been viewed as being guided by assembly rules (i.e. biotic restrictions on observed co-occurrence patterns in species assemblages), but such biotic assembly rules have been difficult to define in statistical models (Götzenberger *et al.* 2012). One approach, originally advocated to improve the ecological realism of predictions of individual species distributions, is the incorporation of biotic variables as surrogates for key

interspecific interactions (i.e. as proxy variables for biotic constraints) into statistical models that link species occurrence to environmental conditions (reviewed in Wisz *et al.* 2013). These species-distribution models (SDMs) can generate useful predictions of individual species occurrences across a region (Guisan & Thuiller 2005), and tend to be improved by the inclusion of variables representing important biotic interactions (Heikkinen *et al.* 2007; le Roux *et al.* 2013b). Summing predictions from individual species models (i.e. "stacking") then enables the prediction of entire community assemblages from the species-level up (Guisan & Rahbek 2011). The accuracy of predictions from stacked species distribution models (SSDMs) for species assemblages tend to vary along ecological gradients (Pellissier *et al.* 2012; Pottier *et al.* 2013), but the inclusion of biotic variables may be expected to provide superior estimates of community-level ecological properties under all conditions.

The traditional concept of assembly rules reflects the notion that species do not occur randomly but are restricted in their co-occurrence by interspecific competition, with the outcome of competitive interactions being determined by species functional traits (Keddy 1992; de Bello *et al.* 2012; Götzenberger *et al.* 2012). As a result, statistical models of community assembly may integrate species traits based on functional assembly theory (Keddy 1992). Trait-based models can predict community attributes by comparing traits of individual species to modelled patterns of traits aggregated at the community level (Shipley *et al.* 2006; Laughlin *et al.* 2012). However, those models are generally focused on reproducing abiotic filtering processes and do not explicitly account for positive or negative biotic interactions (although see Schöb *et al.* 2012). Nonetheless, measuring functional traits that reflect resource acquisition strategies may help to detect and account for the influence of biotic processes (Lavorel & Garnier 2002). Indeed, in plant communities competition may drive directional functional shifts toward convergent values of traits conferring better

resource acquisition or sequestration abilities (Kunstler *et al*. 2012). In low resource

environments, including arctic-alpine tundra, leaf dry matter content (LDMC) has been

viewed as reflecting resource sequestration abilities (Kurokawa *et al*. 2010; Lavorel *et al.*

2007). Thus, assuming that this trait is related to competitive ability, including biotic

constraints into community models should improve predictions of community-level patterns

of LDMC due to dominant species favouring the occurrence of species with similar LDMC

values (Pellissier *et al*. 2010), thereby confirming our understanding of this ecological

process.

Irrespective of the methodology employed, there have been few attempts to constrain

community-level modelling by incorporating biotic interactions or well-defined proxies

thereof (Boulangeat *et al.* 2012). The scarcity of community models implementing biotic

assembly processes may reflect the difficulty of detecting both positive and negative biotic

processes acting on plant communities in observational studies (Götzenberger *et al*. 2012)

and/or the detailed abiotic datasets required to avoid detecting spurious correlations due to

shared habitat requirements (Ovaskainen *et al*. 2010; le Roux *et al*. 2013b). Here, we

circumvent these problems by utilizing a unique survey design that is especially suitable for

assessing the accuracy of community-level models, comprising a uniquely detailed dataset of

abiotic parameters (field-quantified soil conditions and topography, see Table S1 in

Supporting Information). Moreover, in the relatively simple arctic-alpine tundra environment

competitive and facilitative effects can be clearly detected and a few dominant species are

known to impact the occurrence of sub-dominant species (Rousset & Lepart 2000; Tybirk *et

al*. 2000; Pajunen *et al*. 2011; le Roux *et al*. 2013b). As the influence of biotic interactions

varies with environmental conditions (Bertness & Callaway 1994), the relative impacts of

dominant species are not expected to be constant across a gradient of environmental severity,

but rather for negative inter-specific interactions (and therefore competitive exclusion) to be strongest under conditions of higher productivity (and visa versa for positive interspecific interactions; Michalet *et al.* 2006; Maestre *et al.* 2009).

In this study we test whether including the cover of three dominant species as biotic constraints into statistical models improves predictions of three assemblage properties of the sub-dominant community (richness, composition and community-weighted mean of LDMC), using both direct modelling and stacked species distribution modelling. We used a robust cross-validation method to analyse detailed field-quantified data from two contrasting study sites on the northern and southern slopes of a mountain massif. Specifically, we test two hypotheses: i) biotic constraints significantly improve the accuracy of predictions of community properties compared to abiotic-only models for both direct and SSDMs modelling approaches, and ii) the improvement in model predictions is greater at the higher productivity site due to stronger competitive interactions (as indicated by higher biomass and vegetation cover; Table S1).

Materials and methods

Sampling design and methods
Vascular plant species cover and environmental characteristics were quantified at two sites on the Saana massif (69° N 21° E; Fig. 1) in north-western Finland. The sites were located on north- (north site hereafter) and south-facing (south site) slopes, and had contrasting ecological condition (differing in mean biomass, vegetation cover and abiotic conditions, see Table S1). At both sites 6 grids, each comprising 160 contiguous 1 m^2 quadrats in an 8 x 20 arrangement, were located within areas of < 2 ha. Grids were separated by 20 - 131 m at each

site, thus minimizing variation in macro-environmental characteristics and species source pools. Grids were placed to sample the full range of environmental conditions at each site, focusing particularly on mesotopography and vegetation cover. Within each quadrat all vascular plants were identified and each species cover visually estimated, assigning species with very low cover a minimum value of 0.25 %. Six abiotic characteristics were recorded, based on previous studies which have demonstrated their importance for the occurrence of arctic-alpine plant species (Gough *et al*. 2000; Körner 2003; Bruun *et al*. 2006); mesotopography, soil moisture, maximum potential solar radiation, rock cover, soil temperature (recorded in each quadrat) and soil pH (recorded at 4 m intervals and interpolated to a 1 m resolution; details in Appendix S1 and in Aalto *et al*. 2013; le Roux *et al*. 2013a). Those field-measured predictors enable the dominant abiotic gradients to be accurately accounted for when analysing species occurrence patterns, reducing the risk of mistakenly identifying spurious correlations between species as the outcomes of true biotic interactions (Ovaskainen *et al*. 2010).

Three biotic characteristics were extracted from the vegetation data for use as predictor variables: the cover of *Empetrum nigrum* ssp. *hermaphroditum* (crowberry; *Empetrum* hereafter), *Betula nana* (dwarf birch; *Betula* hereafter) and *Juniperus communis* ssp. *alpina* (common juniper, *Juniperus* hereafter). These are the only species to have ≥ 5 % mean cover at either (or both) sites, and together comprise 68 and 72 % of the total vascular plant cover in the south and north site respectively. Moreover, these species are known to affect local habitat conditions and the performance of other species (Rousset & Lepart 2000; Tybirk *et al*. 2000; Eskelinen 2010; Pajunen *et al*. 2011; le Roux *et al.* 2013b). The abiotic and biotic predictor variables only exhibited moderate collinearity (across both sites maximum pairwise

$r_s = 0.78$; all other correlations ≤ 0.56), and therefore all were retained for subsequent modelling. Analyses were conducted in R (R Development Core Team 2011).

Leaf dry matter content is a key leaf trait reflecting species resource acquisition strategies and was selected for the analysis of community functional structure because it correlates with competitive ability in low resource environment (Lavorel *et al.* 2007; Pellissier *et al.* 2010). Species LDMC values were collected from the literature and online databases for as many of the species as possible (traits from Kleyer *et al.* 2008; Kattge *et al.* 2012; Pellissier *et al.* 2013b; and supplemented by Bråthen & Odasz-Albrigtsen 2000; Austrheim *et al.* 2005; Cerabolini *et al.* 2010).

Stacked species distribution modelling

When modelling the distribution of individual species, a species had to have a minimum sample size of 8 records for calibrating models during cross-validation (i.e. ≥ 8 records in the grid where it was most common and ≥ 8 records in the other five grids combined) to be included in analyses for one site. A total of 114 vascular plant species were recorded in the 1920 quadrats sampled, with 40 and 51 of these sub-dominant species well enough sampled in the north and south sites (respectively) to include in analyses.

The distribution of individual species was modelled using three different statistical techniques to account for potential differences related to methodologies; generalized linear models (GLM), generalized additive models (GAM) and boosted regression trees (BRT; Elith *et al.* 2008). A binomial (or Bernoulli for BRTs) distribution of errors was used to model species occurrence (see le Roux *et al.* 2013b for further details). For all three statistical

techniques, a simple abiotic-only SDM and a more complex full SDM were constructed for each species:

Occurrence = Mesotopography + Soil moisture + Soil temperature + Radiation + Soil pH +
Rock cover   …[simple model]

Occurrence = simple model + *Betula* + *Empetrum* + *Juniperus*

…[full model]

Occurrence predictions from individual species distribution models were then summed across all modelled species to generate species richness predictions (i.e. stacked species distribution models). By stacking the distribution of individual species, we also generated predictions of two other assemblage properties for each quadrat: the community-weighted mean of LDMC and assemblage composition. Predictions of community-weighted mean (CWM) values for LDMC (hereafter $LDMC_{CWM}$) were calculated from the occurrence predictions from both the simple and full models. Species missing trait values were excluded from analyses (for SSDMs $n = 7$ for north and $n = 10$ for south; in total 25 species were missing for the entire dataset but they only comprised 11 % of all occurrences and < 2 % of vascular plant cover). The resulting values were subtracted from observed CWMs, with the residuals from the simple and full models being compared using a paired t-test. The similarity of observed and predicted species composition (i.e. generated from the SSDMs) was calculated using the Sørensen similarity index. The 5th percentile of similarity values and mean similarity were calculated for both the simple and full models, using a paired t-test to test the significance of differences in mean similarity.

Direct modelling

In addition to the SSDM approach, we also modelled species richness and $LDMC_{CWM}$

directly (i.e. using the observed values as response variables). Analyses were again repeated

using all three statistical techniques, assuming a Poisson and a Gaussian distribution of

errors, respectively. Following the methodology described above, two models were

constructed for each analysis:

Richness (or $LDMC_{CWM}$) = Mesotopography + Soil moisture + Soil temperature + Radiation

+ Soil pH + Rock cover   …[simple model]

Richness (or $LDMC_{CWM}$) = simple model + *Betula* + *Empetrum* + *Juniperus*

…[full model]

Validation

Six-fold cross-validation with non-random assignment was used to quantify the predictive

power of all models. In this procedure data from five grids was used to calibrate the model

which was then used to predict the occurrence of the species in the withheld grid, with the

procedure repeated six times until each grid had been withheld once (following Wenger &

Olden's 2012 method of non-random assignment). After predictions had been made for each

withheld grid, the predictions for all 960 quadrats were compared to the observed data. For

occurrence data, this was done by calculating the area under the curve of a receiver operating

characteristic plot (AUC; Fielding & Bell 1997) and the true skill statistic (TSS; Allouche *et*

al. 2006). This cross-validation technique provides a strong test of model transferability and

accounts for overfitting (i.e. estimates of predictive power are not automatically increased

with the inclusion of additional variables; Wenger & Olden 2012). Moreover, cross-validated

estimates of predictive power are fairly robust to the effects of spatial autocorrelation

(Hijmans 2012), particularly when employing non-random cross-validation methods that

create independent training and testing datasets (Wenger & Olden 2012). For species richness

and LDMC$_{CWM}$ the relationship between observed and predicted values was examined using

a bootstrapped Spearman correlation (2000 repeats).

To visualize the relationship between the cover of the three dominant plant species

("dominant plant cover" hereafter) and the richness of all other species, a loess smoother was

fitted to the relationship between species richness and dominant plant cover. In all analyses

Empetrum, *Betula* and *Juniperus* were excluded from the response variable (i.e. species

richness, community composition and LDMC$_{CWM}$ were calculated without the three dominant

species).

Results

Stacked species distribution modelling
The inclusion of the cover of the three dominant species when modelling the occurrence of

individual species significantly increased AUC and TSS values based on independent

evaluation data, irrespective of the statistical method used (Fig. S1). While the relative

importance of predictors differed between the north and south sites, the contribution of the

biotic variables exceeded that of some abiotic variables in all analyses (Fig. S2).

Species richness predictions based on SSDMs generally correlated well with observed

richness, but improved strongly and significantly after the inclusion of the three biotic

variables (Fig. 2a). Predictions of variation in species richness along a gradient of dominant

plant cover showed that the full model gave richness estimates closer to observed values than

the simple models when the combined cover of the three dominant species exceeded c. 20 %

(i.e. approximately half of the sampled cells, except for the GBM model for the north site;

12

Fig. 3a). Below 20 % dominant species cover the simple models provided more accurate or similar estimates of richness as the full model (Fig. 3a). Mean similarity between observed and predicted species composition was significantly higher for full models including biotic interactions than simple models, with marked improvement in the least accurate predictions in the south site (indicated by changes in the 5th percentile of similarity values; Fig. 4).

Predictions of $LDMC_{CWM}$ from models including the cover of the three dominant species were significantly and consistently better correlated with observed values than predictions from the simple models for the higher productivity south site (Fig. 2c). However, this was not the case in the lower productivity north site where predictions from the full models were significantly worse than those of the simple model in four of the six analyses (Fig. 2c). This difference between the two sites was also reflected in the changes in $LDMC_{CWM}$ along the dominant plant cover gradient; at the south site $LDMC_{CWM}$ tended to increase with the combined cover of the three dominant species, especially as dominant cover increase up to 30% (Fig. S3). This trend was better represented by the full model than the baseline model, although the accuracy of the full models depended on the modelling method implemented (with GBM performing worst). By contrast, a hump-shaped response of $LDMC_{CWM}$ to dominant plant cover was observed at the northern site (Fig. S3). This trend was poorly predicted by simple models, with slightly better performance by the full models at the south site (Fig. S3). Nonetheless, with increasing cover of the three dominant species the range of observed $LDMC_{CWM}$ values narrowed at both sites, suggesting with greater dominant cover the convergence of community-wide LDMC to values of *c*. 350, corresponding to the LDMC observed for the most dominant species, *Empetrum hermaphroditum* (Fig. S4).

Direct modelling

13

Predictions of species richness that ignore species identity (i.e. the direct modelling approach) also improved with the addition of the three biotic variables (Fig. 2b). For all three statistical techniques the correlation between observed and predicted richness was significantly stronger for the full models than the simple models (Fig. 2b). The performance of simple and full species richness models varied along a gradient of dominant plant cover, with full models generally performing best in areas where the cover of the three dominant species exceeded 20% (Fig. 3b). Directly modelling $LDMC_{CWM}$ with the full model gave significantly stronger correlations with observed values than the simple model for the south site, but not the north site (Figs 2d; in agreement with the results based on SSDMs). As for the SSDMs, the contribution of predictors differed between the two sites, with soil pH having the strongest relative importance at the south site, compared to soil moisture at the north site (Fig. S5). The three biotic variables had high relative importance for all models, except when modelling LDMC in the northern site (Fig. S5).

Discussion

Here we demonstrate that including biotic interactions (inferred from dominant species cover) into assemblage-level models provides better predictions of community properties, including species richness, composition and functional structure. By focusing on a relatively simple ecosystem, arctic-alpine tundra, we could reveal the strong biotic influence of dominant species for community-level models. Incorporating the cover of dominant species consistently improved predictions of community composition and gave better estimates of species richness at sites where dominants had moderate or high cover. The community-level value of a key functional trait, LDMC, was also predicted with significantly better accuracy at one of the two sites, suggesting that model improvements may be contingent on some site

14

characteristics (e.g. biomass). These improvements were consistent across the two prevailing modelling techniques; stacked species distribution models, which consider each component species separately, and direct modelling, which models aggregate community properties. Thus, our study highlights the necessity of including biotic interactions when modelling both individual species and entire communities, irrespective of the approach utilized. Moreover, as biotic interactions are thought to impact more strongly on species assemblages at lower latitudes (Pellissier *et al*. 2013a; and altitudes: Schemske *et al*. 2009), we expect this result to be true in a diversity of habitats.

Interestingly, we found some contrasting results from the north and south sites, with improvements in predictions being stronger and more consistent at the south site when including biotic interactions. The sites differ considerably in ecological conditions that may affect the relative importance of facilitation and competition (and, therefore, of the three dominant species). Specifically, the growing season starts earlier at the south site due to earlier spring melting (*c*. three weeks), contributing to the higher biomass and vascular species cover at this site (see e.g. Litaor *et al.* 2008). As a result, the greater improvement of community property predictions at the south, but not the north, site may reflect the overall greater influence of the dominant species and competitive processes there. Indeed, the three dominant species considered in our models can efficiently sequester resource under low-nutrient conditions (as evidenced by their high LDMC; see Fig. S6) and at the south site limit co-occurring individuals to species with similar sequestration abilities (i.e. comparable LDMC values; Pellissier *et al.* 2010; as suggested by Fig. S4). The same pattern is not observed at the north site (e.g. Fig. 2c, d), possibly due to the relative scarcity of species with low LDMC values (Table S1, Fig. S4) and/or the tendency for biotic interactions generally to have lower importance at that site (Figs S2 and S5). The greater influence of biotic

interactions at the south site is also evident when examining variable importance; results from both SSDMs and direct modelling consistently attribute a greater influence to the three dominant species in the south site than in the north site (Figs. S2 and S5).

The rapid non-linear decrease in observed species richness with increased dominant plant cover was differentially captured by the simple (i.e. abiotic) and full (abiotic + biotic) models (Fig. 3). Both simple and full models correctly reproduced the steep decline of species richness with increasing cover of the dominant species (in agreement with, e.g., Tybirk *et al*. 2000; Pajunen *et al*. 2011). However, when biotic interactions were not considered, richness was overpredicted under high cover of the dominant species. This reflects that relatively similar abiotic conditions occur in cells with high diversity and with high dominant cover, and that the consideration of dominant plant cover is necessary to accurately distinguish which assemblage will develop under those conditions. While models lacking the biotic variables predict only minor changes in mean richness across a dominant plant cover gradient at the south site, at the north site the drop in diversity as dominant cover increases from 0 to c. 20 % is surprisingly well represented by the simple models. This is because on the northern slope the total cover of dominant species is low in species-rich snowbeds (particularly those with temporary flooding; e.g. see Fig. 4 in le Roux & Luoto, 2014); a decline in snow accumulation may therefore simultaneously be correlated with lower species richness and lower dominant cover. Thus, abiotic conditions that accurately predict the total cover of dominant species may be adequate for explaining some competition-related variation in species richness, but only under limited conditions.

Toward more ecological models

Plant assemblage predictions require considering biotic interactions (Wisz *et al.* 2013), which recent research suggests maybe the strongest mechanism explaining differences in realized niches across species ranges (Pellissier *et al.* 2013a). Our study demonstrates both the role of biotic interactions in plant community assembly in a tundra ecosystem and the need to include these interactions to achieve accurate biodiversity forecasts. In this research species cover was used as a proxy for the frequency and intensity of key biotic interactions, an assumption that is reasonable in this ecosystem given the previous experimental demonstration of the strong competitive and facilitative effects associated with the three dominant species examined (Rousset & Lepart 2000; Tybirk *et al.* 2000; Pajunen *et al.* 2011; le Roux *et al.* 2013b). However, if applying this method to other vegetation types it would be crucial to re-assess the validity of this assumption. Additionally, due to the correlative nature of this approach, complementary manipulative studies would be required to confirm the influence of dominant species where such data are lacking.

While some previous studies have used interpolated abiotic variables when modelling species distributions (e.g. Boulangeat *et al.* 2012), in mountain environments spatial heterogeneity may create a mosaic of microhabitats and therefore fine-scale abiotic measurements are needed to account for this heterogeneity and to disentangle biotic and abiotic effects. Moreover, where biotic factors are to be used as predictor variables, it is important that they also be accurately quantified, as species cover may vary greatly over short distances and may be difficult to predict accurately (see e.g. Appendix S2). Our study design based on field-quantified data offers a robust method to demonstrate the role of biotic interactions in community assembly (after accounting for abiotic covariates) and advocates for the further development of community models that are rooted in ecological theory.

So far, two main modelling approaches have been used to model assemblage composition, the first based on species occurrence (Guisan & Rahbek 2011; Pottier *et al*. 2013) and the second on species functional properties (Shipley *et al.* 2006). Our results suggest that both of these approaches to modelling communities are likely lead to unrealistic predictions if biotic interactions are not explicitly accounted for in the modelling process. In our relatively simple study system, a sound ecological understanding about plant-plant interactions (Rousset & Lepart 2000; Tybirk *et al*. 2000; Pellissier *et al*. 2010; Pajunen *et al*. 2011; le Roux *et al*. 2013b) allowed us to select *a priori* the species with the most frequent (and therefore probably most influential) interactions to be included as biotic constraints in the models. However, in many cases, identifying the species which predict (and presumably drive) variation in community composition is not straight forward. Indeed, in systems with complex interaction networks, *a priori* selection of biotic interactions to be included in the models would be very challenging without detailed ecological knowledge of the system. Ideally the simultaneous calculation of competitive strength in pairwise interactions (Kunstler *et al.* 2012) and of the effects of abiotic predictors (Guisan & Thuiller 2005) is required in a unified statistical framework. Further methodological development should focus on simultaneously modelling the hierarchical competitiveness of species (e.g. by considering species functional attributes) and their effect on species distribution patterns.

Conclusion

We demonstrate the strong role of biotic interactions in shaping multiple plant community attributes, and thus the need to account for them when forecasting assemblages in space. Irrespective of the modelling methodology applied, statistical models of communities are likely to be limited in their ability to predict spatial patterns of species assemblages' richness and composition without accounting for biotic interactions. Moreover, incorporating

information on biotic interactions strengthens our capacity not only to predict composition of communities, but also how their structure and function will be modified in the face of global change.

Acknowledgements

References

Aalto, J., le Roux, P. C. & Luoto, M. (2013) Vegetation mediates soil temperature and moisture in arctic-alpine environments. *Arctic, Antarctic, and Alpine Research,* **45,** 1-11.

Allouche, O., Tsoar, A. & Kadmon, R. (2006) Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology,* **43,** 1223-1232.

Austrheim, G., Evju, M. & Mysterud, A. (2005) Herb abundance and life-history traits in two contrasting alpine habitats in southern Norway. *Plant Ecology,* **179,** 217-229.

Bertness, M. D. & Callaway, R. M. (1994) Positive interactions in communities. *Trends in Ecology & Evolution,* **9,** 191-193.

Boulangeat, I., Gravel, D. & Thuiller, W. (2012) Accounting for dispersal and biotic interactions to disentangle the drivers of species distributions and their abundances. *Ecology Letters,* **15,** 584-593.

Bråthen, K. A. & Odasz-Albrigtsen, A. M. (2000) Tolerance of the arctic graminoid *Luzula arcuata* ssp *confusa* to simulated grazing in two nitrogen environments. *Canadian Journal of Botany,* **78,** 1108-1113.

Bruun, H. H., Moen, J., Virtanen, R., Grytnes, J. A., Oksanen, L. & Angerbjörn, A. (2006) Effects of altitude and topography on species richness of vascular plants, bryophytes and lichens in alpine communities. *Journal of Vegetation Science,* **17,** 37-46.

Cerabolini, B. E. L., Brusa, G., Ceriani, R. M., De Andreis, R., Luzzaro, A. & Pierce, S. (2010) Can CSR classification be generally applied outside Britain? *Plant Ecology,* **210,** 253-261.

de Bello, F., Price, J. N., Münkemüller, T., Liira, J., Zobel, M., Thuiller, W. *et al.* (2012) Functional species pool framework to test for biotic effects on community assembly. *Ecology,* **93,** 2263-2273.

Elith, J., Leathwick, J. R. & Hastie, T. (2008) A working guide to boosted regression trees. *Journal of Animal Ecology,* **77,** 802-813.

Elmendorf, S. C., Henry, G. H. R., Hollister, R. D., Björk, R. G., Boulanger-Lapointe, N., Cooper, E. J. *et al.* (2012) Plot-scale evidence of tundra vegetation change and links to recent summer warming. *Nature Climate Change,* **2,** 453-457.

Eskelinen, A. (2010) Resident functional composition mediates the impacts of nutrient enrichment and neighbour removal on plant immigration rates. *Journal of Ecology*, **98,** 540–550.

Fielding, A. H. & Bell, J. F. (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation,* **24,** 38-49.

Götzenberger, L., de Bello, F., Bråthen, K. A., Davison, J., Dubuis, A., Guisan, A. *et al.* (2012) Ecological assembly rules in plant communities-approaches, patterns and prospects. *Biological Reviews,* **87,** 111-127.

Gough, L., Shaver, G. R., Carroll, J., Royer, D. L. & Laundre, J. A. (2000) Vascular plant species richness in Alaskan arctic tundra: the importance of soil pH. *Journal of Ecology,* **88,** 54-66.

Guisan, A. & Rahbek, C. (2011) SESAM – a new framework integrating macroecological and species distribution models for predicting spatio-temporal patterns of species assemblages. *Journal of Biogeography,* **38,** 1433-1444.

Guisan, A. & Thuiller, W. (2005) Predicting species distribution: offering more than simple habitat models. *Ecology Letters,* **8,** 993-1009.

Heikkinen, R. K., Luoto, M., Virkkala, R., Pearson, R. G. & Körber, J.-H. (2007) Biotic interactions improve prediction of boreal bird distributions at macro-scales. *Global Ecology and Biogeography,* **16,** 754-763.

Hijmans, R. J. (2012) Cross-validation of species distribution models: removing spatial sorting bias and calibration with a null model. *Ecology,* **93,** 679-688.

Kattge, J., Díaz, S., Lavorel, S., Prentice, C., Leadley, P., Bönisch, G. *et al.* (2012) TRY - a global database of plant traits. *Global Change Biology,* **17,** 2905-2935.

Keddy, P. A. (1992) Assembly and response rules: two goals for predictive community ecology. *Journal of Vegetation Science,* **3,** 157-164.

Kleyer, M., Bekker, R. M., Knevel, I. C., Bakker, J. P., Thompson, K., Sonnenschein, M. *et al.* (2008) The LEDA Traitbase: a database of life-history traits of the Northwest European flora. *Journal of Ecology,* **96,** 1266-1274.

Körner, C. (2003) *Alpine plant life: functional plant ecology of high mountain ecosystems.* Springer-Verlag, Heidelberg.

Kunstler, G., Lavergne, S., Courbaud, B., Thuiller, W., Vieilledent, G., Zimmermann, N. E. *et al.* (2012) Competitive interactions between forest trees are driven by species' trait hierarchy, not phylogenetic or functional similarity: implications for forest community assembly. *Ecology Letters,* **15,** 831-840.

Kurokawa, H., Peltzer, D. A. & Wardle, D. A. (2010) Plant traits, leaf palatability and litter decomposability for co-occurring woody species differing in invasion status and nitrogen fixation ability. *Functional Ecology,* **24,** 513-523.

Laughlin, D. C., Joshi, C., van Bodegom, P. M., Bastow, Z. A. & Fulé, P. Z. (2012) A predictive model of community assembly that incorporates intraspecific trait variation. *Ecology Letters,* **15,** 1291-1299.

Lavorel, S., Díaz, S., Cornelissen, J. H. C., Garnier, E., Harrison, S. P., McIntyre, S. *et al.* (2007) Plant functional types: are we getting any closer to the Holy Grail? *Terrestrial Ecosystems in a Changing World* (eds J. G. Canadell, D. E. Pataki & L. F. Pitelka), pp. 149-164. Springer-Verlag, Berlin.

Lavorel, S. & Garnier, E. (2002) Predicting changes in community composition and ecosystem functioning from plant traits: revisiting the Holy Grail. *Functional Ecology,* **16,** 545-556.

le Roux, P. C., Aalto, J. & Luoto, M. (2013a) Soil moisture's underestimated role in climate change impact modelling in low-energy systems. *Global Change Biology,* **19,** 2965–2975.

le Roux, P. C., Lenoir, J., Pellissier, L., Wisz, M. S. & Luoto, M. (2013b) Horizontal, but not vertical, biotic interactions affect fine-scale plant distribution patterns in a low energy system. *Ecology,* **94,** 671–682.

le Roux, P. C. & Luoto, M. (2014) Earth surface processes drive the richness, composition and occurrence of plant species in an arctic–alpine environment. *Journal of Vegetation Science*, **25**, 45-54.

Lenoir, J., Gegout, J. C., Marquet, P. A., de Ruffray, P. & Brisse, H. (2008) A significant upward shift in plant species optimum elevation during the 20th century. *Science,* **320,** 1768-1771.

Litaor, M. I., Williams, M. & Seastedt, T. R. (2008) Topographic controls on snow distribution, soil moisture, and species diversity of herbaceous alpine vegetation, Niwot Ridge, Colorado. *Journal of Geophysical Research-Biogeosciences,* **113,** G02008.

Lortie, C. J., Brooker, R. W., Choler, P., Kikvidze, Z., Michalet, R., Pugnaire, F. I. *et al.* (2004) Rethinking plant community theory. *Oikos,* **107,** 433-438.

Maestre, F. T., Callaway, R. M., Valladares, F. & Lortie, C. J. (2009) Refining the stress-gradient hypothesis for competition and facilitation in plant communities. *Journal of Ecology,* **97,** 199-205.

Michalet, R., Brooker, R. W., Cavieres, L. A., Kikvidze, Z., Lortie, C. J., Pugnaire, F. I. *et al.* (2006) Do biotic interactions shape both sides of the humped-back model of species richness in plant communities? *Ecology Letters,* **9,** 767-773.

Mokany, K. & Ferrier, S. (2011) Predicting impacts of climate change on biodiversity: a role for semi-mechanistic community-level modelling. *Diversity and Distributions,* **17,** 374-380.

Nogués-Bravo, D. & Rahbek, C. (2011) Communities under climate change. *Science,* **334,** 1070-1071.

Ovaskainen, O., Hottola, J. & Siitonen, J. (2010) Modeling species co-occurrence by multivariate logistic regression generates new hypotheses on fungal interactions. *Ecology,* **91,** 2514-2521.

Pajunen, A. M., Oksanen, J. & Virtanen, R. (2011) Impact of shrub canopies on understorey vegetation in western Eurasian tundra. *Journal of Vegetation Science,* **22,** 837-846.

Pellissier, L., Bråthen, K. A., Pottier, J., Randin, C. F., Vittoz, P., Dubuis, A. *et al.* (2010) Species distribution models reveal apparent competitive and facilitative effects of a dominant species on the distribution of tundra plants. *Ecography,* **33,** 1004-1014.

Pellissier, L., Bråthen, K. A., Vittoz, P., Yoccoz, N. G., Dubuis, A., Meier, E. S. *et al.* (2013a) Thermal niches are more conserved at cold than warm limits in arctic-alpine plant species. *Global Ecology and Biogeography,* **22,** 933-941.

Pellissier, L., Ndiribe, C., Dubuis, A., Pradervand, J.-N., Salamin, N., Guisan, A. *et al.* (2013b) Turnover of plant lineages shapes herbivore phylogenetic beta diversity along ecological gradients. *Ecology Letters,* **16,** 600-608.

Pellissier, L., Pradervand, J. N., Pottier, J., Dubuis, A., Maiorano, L. & Guisan, A. (2012) Climate-based empirical models show biased predictions of butterfly communities along environmental gradients. *Ecography,* **35,** 684-692.

Pottier, J., Dubuis, A., Pellissier, L., Maiorano, L., Rossier, L., Randin, C. F. *et al.* (2013) The accuracy of plant assemblage prediction from species distribution models varies along environmental gradients. *Global Ecology and Biogeography,* **22,** 52-63.

R Development Core Team (2011) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Rousset, O. & Lepart, J. (2000) Positive and negative interactions at different life stages of a colonizing species (*Quercus humilis*). *Journal of Ecology,* **88,** 401-412.

Schemske, D. W., Mittelbach, G. G., Cornell, H. V., Sobel, J. M. & Roy, K. (2009) Is there a latitudinal gradient in the importance of biotic interactions? *Annual Review of Ecology, Evolution, and Systematics,* **40,** 245-269.

Schöb, C., Butterfield, B. J. & Pugnaire, F. I. (2012) Foundation species influence trait-based community assembly. *New Phytologist,* **196,** 824-834.

Shipley, B., Vile, D. & Garnier, É. (2006) From plant traits to plant communities: a statistical mechanistic approach to biodiversity. *Science,* **314,** 812-814.

Tybirk, K., Nilsson, M. C., Michelson, A., Kristensen, H. L., Shevtsova, A., Strandberg, M. T. *et al.* (2000) Nordic *Empetrum* dominated ecosystems: function and susceptibility to environmental changes. *Ambio,* **29,** 90-97.

Weiher, E., Freund, D., Bunton, T., Stefanski, A., Lee, T. & Bentivenga, S. (2011) Advances, challenges and a developing synthesis of ecological community assembly theory. *Philosophical Transactions of the Royal Society of London B - Biological Sciences,* **366,** 2403-2413.

Wenger, S. J. & Olden, J. D. (2012) Assessing transferability of ecological models: an underappreciated aspect of statistical validation. *Methods in Ecology and Evolution,* **3,** 260-267.

Wisz, M. S., Pottier, J., Kissling, W. D., Pellissier, L., Lenoir, J., Damgaard, C. F. *et al.* (2013) The role of biotic interactions in shaping distributions and realised assemblages of species: implications for species distribution modelling. *Biological Reviews,* **88,** 15-30.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Fig. 1. Location of the two study sites on the Saana Massif in north-western Finland, with the insets showing the relative location and orientation of the 12 study grids.

GLM   GAM   BRT

Richness

a) SSDM

b) Direct

CWM LDMC

c) SSDM

d) Direct

simple   full   simple   full   simple   full
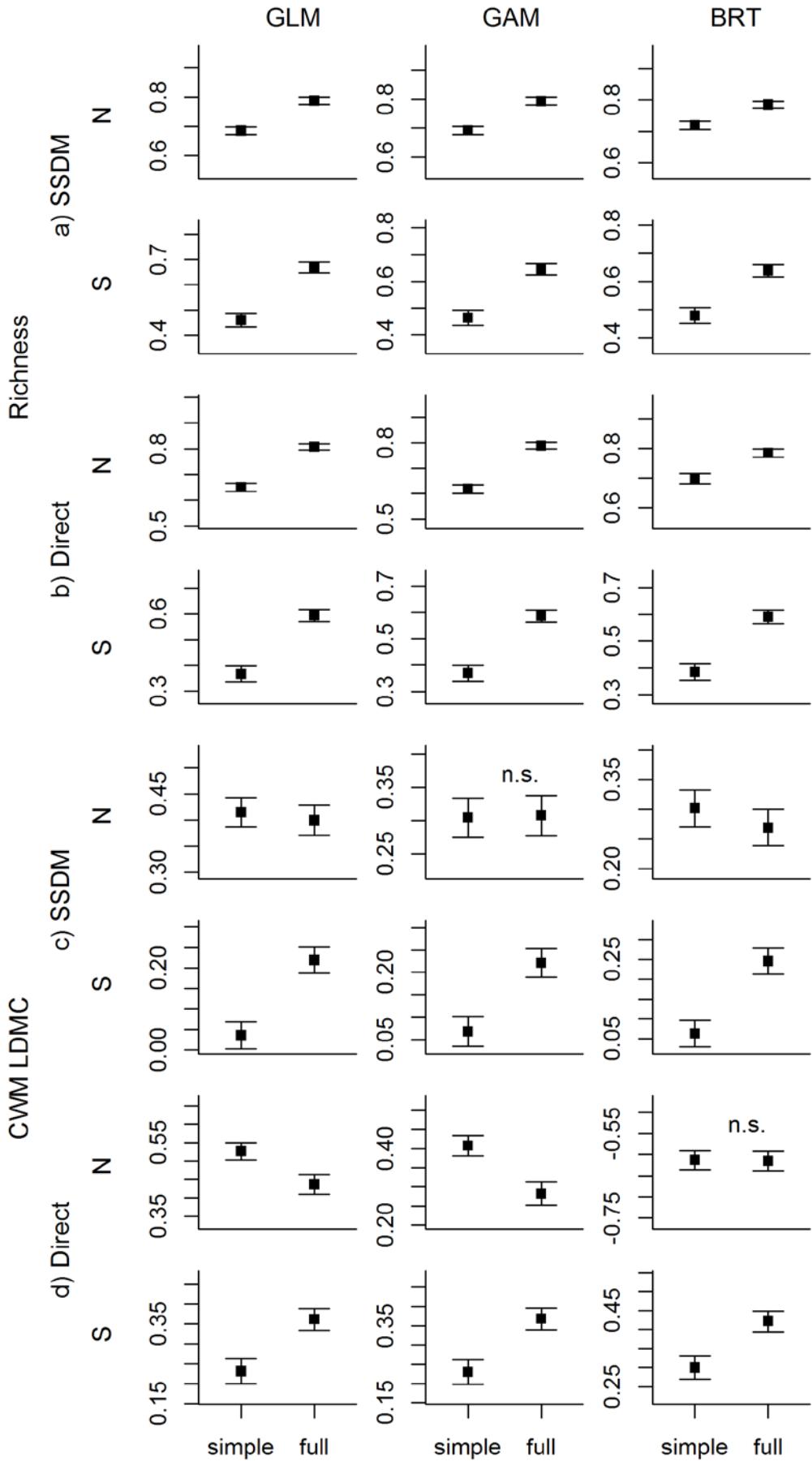
Fig. 2. Bootstrapped Spearman rank correlations between observed and predicted vascular species richness and community-weighted mean leaf dry content matter (CWM LDMC), for simple (six abiotic predictor variables) and full (six abiotic and three biotic predictors) models. Results are presented for the north (n = 40) and south (n = 51 species) study sites. Predicted values either reflect the combination of predictions for individual species (SSDM) or direct modelling (Direct) of the response variable. Three species distribution modelling techniques were implemented: generalized linear models (GLM), generalized additive models (GAM), and boosted regression trees (BRT). All pairs of simple and full models differ significantly in their Spearman correlation values (paired t-test; all $P < 0.001$), except where indicated by "n.s.".
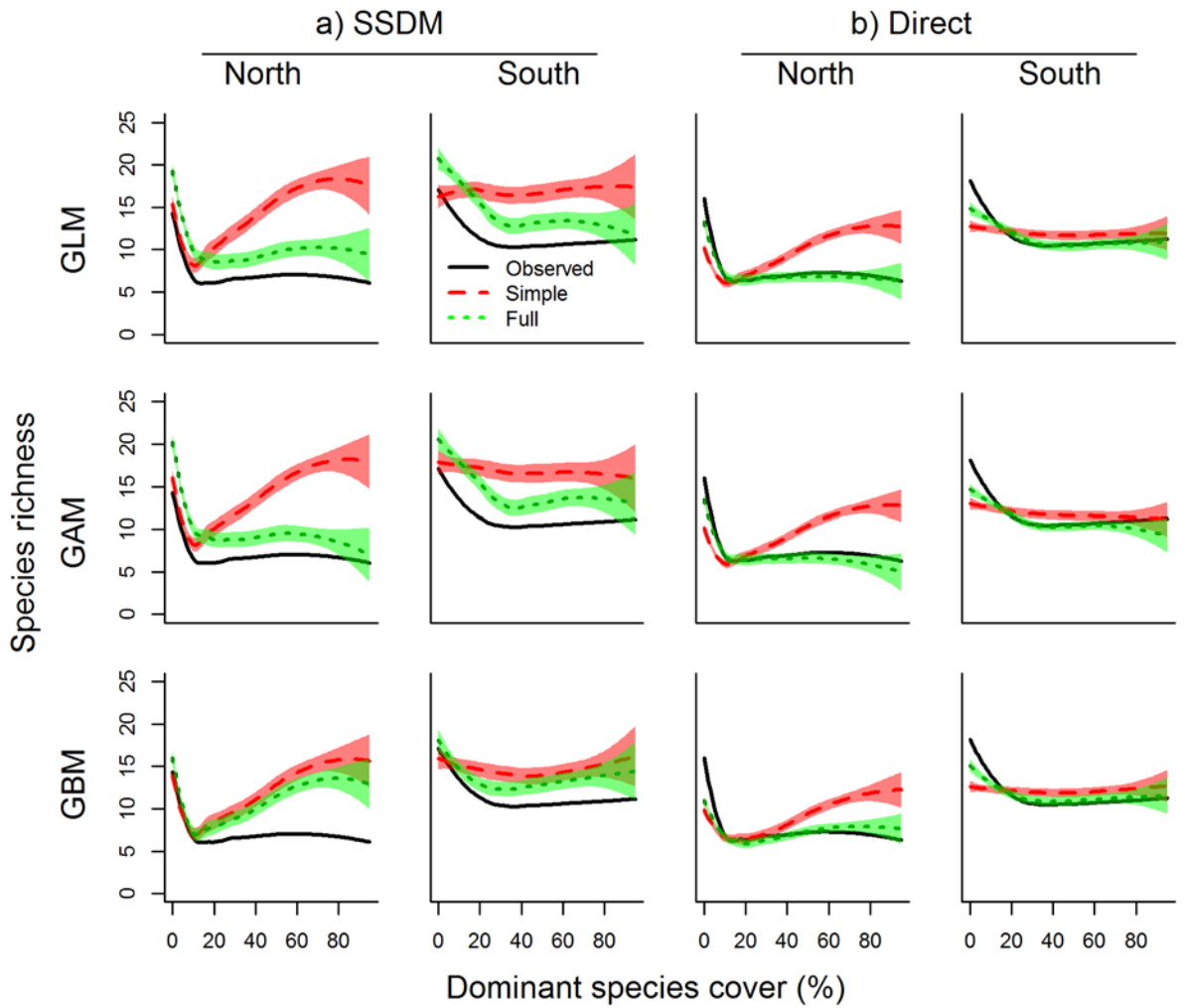
Fig. 3. Loess smooth (± 95 % CI) fitted to the relationship between species richness (observed and predicted) and the combined cover of the three dominant species (*Betula* + *Empetrum* + *Juniperus*) in the north and south site. Predictions of species richness are calculated by a) summing the occurrence predictions for individual species or b) directly modelling total species richness. Three statistical methods were used; GLM = generalized linear models, GAM = generalized additive models, BRT = boosted regression trees. The "simple" predictions are models using only abiotic predictor variables, while the "full" model comprised both abiotic and biotic predictor variables. Observed and predicted species richness exclude occurrences of the three dominant plant species.
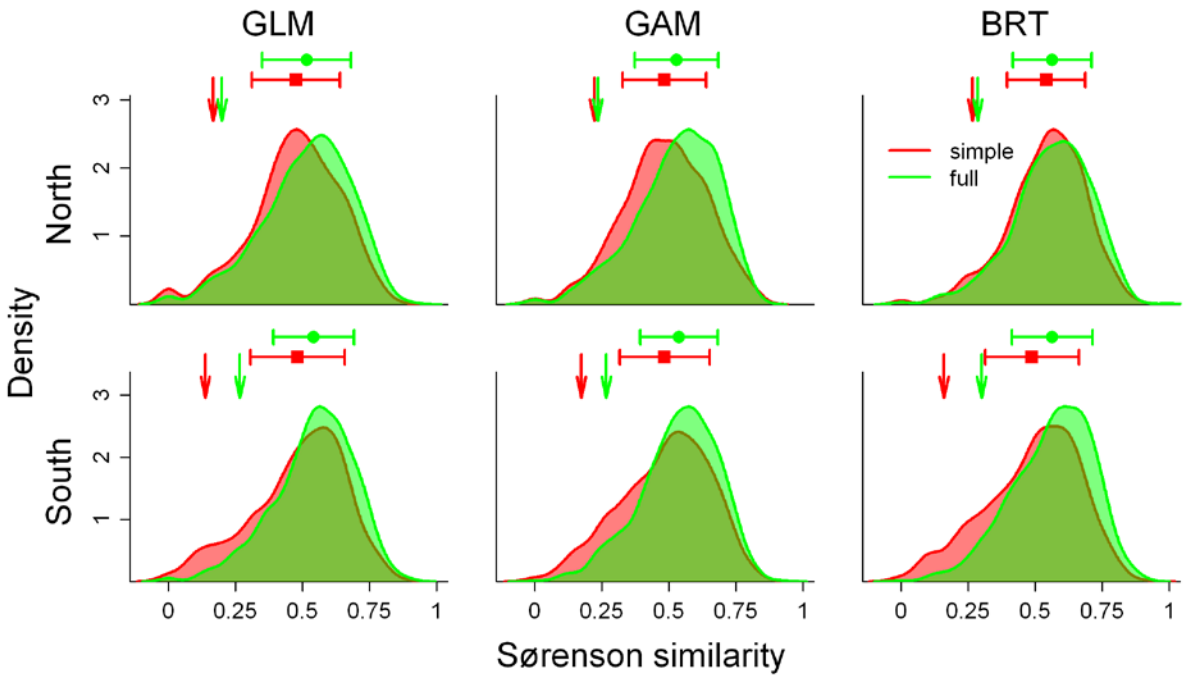
Fig. 4. Density distribution of community similarity values (assessed using the Sørensen index; excluding the three dominant plant species) for the north (n = 40) and south (n = 51 species) study sites, as modelled using simple (six abiotic predictor variables) and full (six abiotic and three biotic predictors) models. Three statistical techniques were implemented: generalized linear models (GLM), generalized additive models (GAM), and boosted regression trees (BRT). Values of 1 indicate perfect predictions of community composition. Arrows indicate the 5th percentile, and horizontal bars indicate mean ± SE similarity values. All full models have significantly higher similarity values than their equivalent simple models (paired t-test; all $P < 0.05$).