

Transferability of trait-based species distribution models

Running head: Transferability of trait-SDM

Peter A. Vesk, William K. Morris, Will C. Neal,

Karel Mokany & Laura J. Pollock

January 2020

Author Statement: PV and WM conceived the idea, LP, PV, WM, WN collected data, KM, WM, LP compiled existing data. WN processed samples, WM and PV analysed data. PV and WM wrote the first draft, with revisions from all authors.

Data Accessibility: Should the manuscript be accepted, the data supporting the results will be archived in an appropriate public repository (Figshare) and the data DOI will be included at the end of the article.

Keywords: Functional traits; Eucalyptus; Prediction; Generalisation; Environmental filtering; Community modelling

14 Type of paper: Letter.

15 Manuscript Length: 148 words in the Abstract, 4860 words in the main text, 52
16 references, 6 Figures, 0 tables, 0 text boxes, 4 supplements.

17 Corresponding author: Peter A Vesk, BioSciences 2, School of BioSciences, the
18 University of Melbourne, Parkville, 3010 VIC, Australia. +61-3-83447480 (ph)
19 pvesk@unimelb.edu.au

Abstract

Trait-based species distribution models (trait-SDMs) enable prediction to new species and situations based on traits. However, predictive transferability is unknown. We fit trait-SDMs with specific leaf area (SLA), maximum height and seed mass as species level predictors in generalised linear mixed models with four environmental predictors for 20 species of eucalypt trees in an outlying reference region. Trait-environment interactions included heavy-seeded species increasing in rugged areas and high-SLA species increasing in areas receiving runoff. We predicted occurrences using traits for 82 species across 18 target regions over $> 100,000 \text{ km}^2$ in south-eastern Australia. Median predictive performance for new species in target regions was 0.65 (area under the receiver operating curve) and 1.24 times that of random (area under the precision recall curve). Prediction in target regions did not worsen across geographic, environmental or compositional space. This work provides a path for first-order models of species distribution using traits.

Introduction

Species distributions—where they occur and why—are core elements of fundamental and applied ecology. Species distribution modelling (SDM) has been the focus of substantial research over the past two decades (Araujo & Guisan 2006; Elith & Leathwick 2009). The vast majority of SDM work is correlative, and the basis for transferring such models is uncertain for many reasons relating to links between modelled distributions and the species niche, including chance correlations, biotic interactions, non-equilibrium distributions, and extrapolation (Colwell & Rangel 2009; Briscoe *et al.* 2016). Hierarchical (or multi-level) models can aid the development of correlative models by assuming some commonality of response and sharing strength between species, which is important as most species are rare (Gelfand *et al.* 2006; Dorrough *et al.* 2011; Ovaskainen & Soininen 2011).

Plant functional traits can be used as predictors of species distribution along environmental gradients (Dorrough & Scroggie 2008; Laughlin *et al.* 2012; Pollock *et al.* 2012; Jamil *et al.* 2013; Brown *et al.* 2014; Pollock *et al.* 2018; Miller *et al.* 2019; Warton *et al.* 2015). Trait-based models provide a route for generalised ecological inference—how traits influence species occurrence—but also for prediction to new situations—where species are likely to occur—based only on their traits and the environment. Incorporating traits into hierarchical models can help both establish the functional role of traits and improve correlative models by adding biologically relevant information. That is, a chance correlation of a species distribution with an environmental gradient is less likely to be included in a model if (a) it is strongly dissimilar to other species, and (b) a trait variable can explain why species vary in their

responses. However, it is unclear how robust trait-based models are for species and regions beyond the training set.

To be operationally useful, it is not enough for traits to simply indicate response, they need also to predict response and be transferable (Sequeira *et al.* 2018; Yates *et al.* 2018). For example, can one generalize between regions, by predicting responses in one region with a trait-based model fitted in another region with different species (Thomas & Vesk 2017)? The ability to transfer ecological models from reference to target settings or conditions could provide predictions to facilitate decisions in conservation and natural resource management (Yates *et al.* 2018). Predictive transfer across regions and species is the subject of this paper.

Here we build on past work with trait-based multi-species distribution models (trait-SDM) of eucalypts (*sensu lato*, including the genera *Eucalyptus*, *Corymbia* and *Angophora*), the dominant trees in Australia, with over 800 species dominating the overstorey of over 92 million km². Much SDM work has demonstrated the importance and method of incorporating effects of temperature, rainfall, irradiance and lithology on eucalypt species distributions (Austin *et al.* 1990; Austin 2002; Fithian *et al.* 2015). Pollock *et al.* (2012) incorporated leaf-height-seed (LHS) traits (specific leaf area (SLA), maximum height and seed mass) into generalised linear mixed models (GLMMs) across Gariwerd (the Grampians Ranges), Victoria using seven environmental gradients including climatic, topographic and edaphic variables. Trait-environment relationships captured by those models included: heavier-seeded species being more likely to occur in sandier (cf. clay) soils, low SLA species more likely to occur in sites with greater rock cover (and less exploitable soil volume). Species occurrence across gradients of irradiance and rock cover within the Grampians was predicted utilising SLA and

maximum height, but not evaluated.

Here we ask whether such trait-SDMs can be transferred between regions (Morán-Ordóñez *et al.* 2017; Sequeira *et al.* 2018) and between species. Can one build and train a model in the Grampians and transfer it to new target regions and species based only on the modelled trait-environment interactions? This is a severe test—previous tests of model transferability within species between regions (Randin *et al.* 2006) or time periods (Morán-Ordóñez *et al.* 2017; Dobrowski *et al.* 2011), report performance ranging from failure to excellent, but with most less than fair. Internal cross-validation of trait-SDM showed good performance between species, within a dataset (Brown *et al.* 2014). Here we evaluate the capacity to predict not only to different regions, but to different species as well.

We are interested in the capacity to (1) predict the direction of the response (i.e., is a target species likely to increase or decrease in occurrence along a given environmental gradient?) and to (2) predict presences and absences. We fit GLMMs to 20 taxa in the Gariwerd-Grampians region, a small outlier of Australia’s Great Dividing Range, and then evaluate their predictive performance with 85 taxa in 18 target regions across over 100,000 km² along the Great Dividing Range in southeast Australia. Because one would expect degree of extrapolation to affect prediction, we then (3) assess whether the predictive performance is negatively related to distance from the reference region in geographic, environmental and compositional space.

Methods

Study system and datasets

Geography

Our work was conducted in southeastern Australia (Fig. S1.1). We used subregions from the the Interim Biogeographic Regionalisation of Australia (IBRA version 7) (Department of the Environment & Energy 2012); we refer to them from hereon as regions. The Greater Grampians (Gariwerd is the Indigenous name), our reference (or training) region, covers an isolated series of mountain ranges in an area ~ 75 km north-south and 30 km east-west. Rising out of sedimentary plains with high topographic variation, the region includes gently rising scarps, rocky ridges and cliffs, gullies, sandy outwash plains and clay-rich depressions. Our target (or test or transfer) regions cover a roughly triangular area of $\sim 118\,000$ km² spanning 5 degrees of latitude and 5 degrees of longitude. This yielded 18 target regions ranging in area from 400 to 17 300 km², containing 10-72 eucalypt taxa each (Table S1.1)

Occurrence data

For the Grampians we used the plots surveyed by Pollock *et al.* (2012). Briefly, ~ 460 plots were surveyed using a gradient-directed transect design following an environmentally stratified selection of start points. Plots were centred on a tree and included the four nearest trees in cardinal points or extended to a maximum of 20 m, whichever was less.

The southeastern Australian data were compiled from the Victorian Biodiversity Atlas

(The State of Victoria, Department of Environment, Land, Water and Planning 2018) and Southeast forests datasets (Austin *et al.* 1990; Austin *et al.* 1996). These were fixed area, 200–2500 m² (90% of which were 900–1 000 m²) plots with all woody tree species recorded. Subspecies were recognised and so, “taxon” is the more correct term, however we occasionally use the term “species” for simplicity. From all datasets we extracted binary presence-absence data.

Trait data

Traits were measured according to standard protocols (Perez-Harguindeguy *et al.* 2016). Trait collection for the Gariwerd-Grampians is described in Pollock *et al.* (2012). New data for the southeast dataset were collected in a series of fieldtrips across the regions, using our plot data to guide sampling. We sampled trees near to roads and tracks, where canopies were accessible with 4m pole clippers. Occasionally for tall taxa, blown-down branches were used.

For each plant sampled, we chose three young, fully-expanded adult leaves from the outer canopy, lacking obvious indication of herbivore or pathogen attack or other epiphylls. We aimed to select over a range of leaf sizes but avoiding the smallest leaves. Leaves were stored in sealed plastic bags inside an insulated cooler while in the field, and in a refrigerator before measurement. Individual fresh leaves were rubbed dry and leaf area (mm², including petiole) measured with a pre-calibrated (LI-COR LI3000 leaf area meter). Occasionally, if a leaf area meter was not available, leaves were scanned on a flatbed scanner with a scale bar. The area of the leaf was then calculated using the software ImageJ. Leaves were then placed in paper bags and oven-dried at 60°C for at least 72

hours. Once removed from the oven, leaves were immediately weighed on laboratory scales (Mettler Toledo ML104).

We harvested 10-20 mature fruits from each plant sampled. Fruits were placed in paper bags and put into an oven at 60°C for at least 72 hours. This process causes the fruits to dehisce their seeds and was followed by shaking the bag to encourage seeds to fall out. Once removed from the oven, we weighed 10 mature seeds. Maximum height data were extracted from the EUCLID database (Slee *et al.* 2006).

Environmental data

In this study we only used environmental data available as GIS layers throughout the southeastern Australian regions. This is in contrast with the original modelling (Pollock *et al.* 2012), and was necessary because the field-based environmental measurements were not available across the target regions dataset. Covariates were selected from a large set using a combination of cluster analysis and discrimination power (see Supplement 1). Candidate covariates were obtained from the Soil and Landscape Grid of Australia (Grundy *et al.* 2015) and the NSW and ACT Regional Climate Modelling (NARClIM) project (Evans *et al.* 2014). The final set used were: Moisture index in the lowest quarter; Topographic Wetness Index; Topographic relief within 1000 m; Total Nitrogen.

Model building

The trait-SDMs were built using the same GLMM approach as in (Pollock *et al.* 2012) and (Pollock *et al.* 2018) and detailed in Supplement 3. Broadly similar approaches are

described in (Brown *et al.* 2014; Jamil *et al.* 2013) and an overview of related techniques is in Ovaskainen *et al.* (2017). These models can all be thought of extensions to linear regression, where the taxon response includes interactions between environmental and trait predictors. Briefly, the occurrence of the j^{th} taxon at the i^{th} site, $Y_{ij} = 1$ is assumed to be Bernoulli distributed. The corresponding probability is modelled as the inverse-logit of a linear function of taxon-specific intercepts and coefficients for covariates that had submodels incorporating the three traits and taxon-level random effects.

Statistical power was a key consideration. In the original Gariwerd-Grampians analyses there were roughly 460 sites for 20 taxa with three traits and seven environmental covariates for a linear model with 21 trait-environment interactions to estimate. We felt this was an upper limit to model complexity. We used the same three traits and chose to use four environmental covariates yielding 12 trait-environment interactions.

Measuring and comparing model performance

The trait-SDMs were trained on the Grampians data and then used to make predictions. We predicted using the fitted coefficients within the Grampians in two ways: (a) based only on the traits of the Grampians taxa, without taxon identities; and (b) to the Grampians taxa, using traits and including the taxon random effect. This enables a within-sample evaluation of how well the trait-SDM performs in the reference region. The difference between the performance of these first and second predictions indicated what fraction of environmental responses, within the Grampians, were not associated with the traits we used. We then made predictions to our target regions with our fitted

184 trait-SDM. We predicted occurrences for all taxa based only on their traits as in the
185 first test within the Grampians. These out-of-sample predictions are the main part of
186 our study.

187 **Performance measurement of presence-absence predictions**

188 Predictive performance was measured with two metrics using the confusion matrix of
189 prediction and observations; the area under the receiver operator curve statistic (AUROC)
190 (Fielding & Bell 1997), as used and recommended in studies of transferability (Randin
191 *et al.* 2006; Sequeira *et al.* 2018). This can be interpreted as the probability that for
192 a randomly chosen pair of plots consisting of one presence and one absence, the model
193 would correctly rank their probability of occurrence. We also examined the area under
194 the precision recall curve (AUPRC), which has advantages in situations where objects are
195 rare, and it is proposed to map well on to the problem of directing survey effort (Sofaer
196 *et al.* 2019). We also examined the explained deviance. These are detailed in Supplement
197 4.

198 To evaluate whether predictive performance declined farther from the reference region
199 we used three measures. We used the geographic distance in kilometres between the
200 centroids of the reference and target regions. Community composition dissimilarity was
201 measured with Jaccard's index (P. Legendre & L. F. J. Legendre 2012). Environmental
202 dissimilarity was measured with Kullback-Leibler divergence (Cover & Thomas 2006).

Correlation of regression coefficients from trait-SDM with taxon regressions

We expect better calibrated models to better predict occurrences. So, we examine calibration of environmental responses and ask whether lower predictive performance for taxa in target regions can be explained by miscalibration. Predicting the regression coefficients for environmental covariates maps onto the problem of a practitioner in a region asking: for a particular environmental gradient will a focal species increase or decrease in occurrence? To benchmark performance of our trait-SDM, for each taxon in each region, we fitted separate generalised linear models based on the same environmental variables (see Supplement 4 for details). These taxon- and region-specific models, we call taxon regressions to avoid confusion. Taxon regressions were used to estimate coefficients for comparison with the coefficients from the trait-SDM, and serve to evaluate how the trait-SDM is making potentially inaccurate predictions.

We asked whether miscalibration of the taxon regression coefficients explained variation in the performance measures, reasoning that a model that poorly predicted coefficients for a taxon in a region would result in poor occurrence prediction. We used absolute value of the miscalibration ($|\text{predicted coefficient} - \text{taxon regression coefficient}|$) for each environmental variable as predictors in a model of performance, expecting negative effects. We built GLMMs for the performance metrics with distance measures and miscalibration as predictors, and random effects of taxon and region.

Results

Characterising the target and reference regions

Environmental covariates varied within and between regions. The Grampians is located towards high evaporation and temperature and lower moisture index loading on the first component of the principal components analysis (Fig. S1.6). In the subset of covariates used in the model, the Grampians experienced a lower and more limited range in moisture index compared with the other regions of southeast Australia (Fig. S1.5). Topographic relief, wetness and nitrogen were more equably covered in both the Grampians and target regions. Target regions shared fewer than half their species with the Grampians, and displayed distance-decay in compositional dissimilarity (Fig. S1.3). Regions >600 km from the Grampians shared few, if any, eucalypt taxa with the Grampians, while regions <400 km from the Grampians shared 5–20% of taxa. Environmental distance of regions from the Grampians tended to increase with geographic distance, except for two regions, resulting in a lack of correlation ($r = 0.2$) between the two measures.

Trait-SDM

Modelling results from fitting the trait-SDM to the reference region (the Grampians) including taxon coefficients are in Supplement 3. A variance components analysis of the fitted linear model illustrates that taxa varied most in their response to moisture index (~ 0.8 SD), but that traits explained relatively little (<20%) of that variance (Fig. 1). Taxa varied less in response to topographic wetness (~ 0.6 SD), but traits explained over one-third of that between-taxon variance.

That variation in range of taxon environmental responses and their explanation by traits are revealed by plotting modelled trait-environment interactions and taxon environmental responses (Fig. 2). First, taxa varied most in response to Moisture Index, but with little relation to traits (Fig. 2, top row). Taxon responses to Topographic Wetness varied less (Fig. 2, second row), but were positively related to SLA, indicating that taxa with thicker, denser leaves were less likely to be found in topographically wet areas. By contrast, taxa with high SLA (i.e., taxa with thinner, flimsier leaves) were likely to respond positively to increased topographic wetness (Fig. 2, second row, left). A positive interaction between seed mass and Topographic Relief indicated that heavier seeded taxa responded positively to increasing ruggedness, while small seeded taxa responded negatively to ruggedness (Fig. 2, third row, center). Predictive performance of the trait-SDM within the reference region was, for some species, comparable to a model that included a taxon random effect (Fig. S4.6).

Calibration of predicted environmental responses

Predicted environmental response coefficients in the reference region were well calibrated for Topographic Wetness and Topographic Relief and less so for Moisture Index (Fig. 3), matching the variance components analysis. Among the target regions, that pattern of calibration was not evident; correlations were weaker, though most positive for Topographic Wetness and Moisture Index. Predicted responses ranged widely in Topographic Relief and Topographic Wetness, less in Moisture Index and Total Nitrogen, which ranged more in the taxon regressions. Relations between responses to Topographic Wetness and traits illustrate that low SLA and high seed mass taxa were consistently found to have negative responses to Topographic Wetness, in the trait-SDM

and taxon regressions (Fig. S3.4).

Predictive performance by target taxa and region

Predictive performance of the trait-SDM varied widely in both AUROC and AUPRC, and more among target taxa within target regions, than between regions (Fig. 5). Median AUROC = 0.65 and 75% of AUROC values > 0.55 (Fig. 5). Many more taxon predictions were excellent for AUROC (AUROC > 0.90) than random or worse (AUROC < 0.5). AUROC within each region ranged roughly over 0.55–0.95, and had performance that declined with prevalence (Figs. S4.1,S4.2). AUPRC was even more tightly (though positively) related to prevalence, so we used AUPRC divided by prevalence, yielding a performance measure relative to that of a random classifier (see Supplement 4). According to AUPRC, most predictions were better than random, with median AUPRC = 1.24 times random, and 25% of predictions > 2.2 times as good as random, but $>25\%$ of predictions were worse than random. Predictive performance was not related to geographic nor environmental distance nor compositional dissimilarity from the reference region (Fig. 5). Neither was performance within the reference region clearly higher than target regions. For the ten taxa that occurred in both the reference region and target regions, AUROC and AUPRC values were similarly, highly variable (Fig. S4.9).

Highly-calibrated environmental response predictions could achieve higher performance measured by AUPRC/prevalence (Fig. 4), and less so for AUROC. GLMMs of the performance statistics confirmed that miscalibrated models performed worse for AUPRC/prevalence, but less clearly so for AUROC (Supplement 4, Table

S4.1). That model of AUPRC/prevalence indicated miscalibration effects on AUPRC/prevalence performance were most important for Topographic Wetness and Topographic Relief (Table S4.1). Miscalibration effects on AUROC were also negative in sign, but were uncertain. The intercept of the model for AUPRC/prevalence (0.43) predicts that for the hypothetical average prediction, with average miscalibration, predictions performed 1.5 times as good as random. In neither performance metric did we find evidence for negative effects of any distance measures (Table S4.1).

Predictions of particular taxa did not perform consistently across regions. Residual variance (measured as standard deviations) was approximately twice that of the taxon level random effect, which was greater than region level (Table S4.2). Taxon level random effects were weakly, negatively correlated with seed mass ($r = -0.30$, 95 % CI $[-0.49, -0.10]$, 80 d.f) S4.10. The effects of miscalibration and the intermediate residual variation among taxa in these models of performance together imply that the variation evident in plots of performance measures would be only partly reducible with better trait knowledge.

Probing predictive performance for some regions and environments

We illustrate predictive performance measured with AUPRC/prevalence for a subset of target regions chosen across the range of median model performance from least to best predicted, as well as the Grampians for reference (Fig. 6). On the right we see that the trait-SDM predicts taxon occurrences with similar performance to the Grampians—taxa vary in their predictive capacity in each of the regions. Most taxa are better predicted in the Victorian Alps than the Snowy Mountains and Jervis, but in each region some taxa are predicted well, with AUPRC > four times as good as random. Notably, the median

AUPRC/prevalence is higher for Victorian Alps than the Grampians, where the model was trained. For AUROC, see Fig. S4.11.

Predicted response of taxa in regions along gradients

We compare the responses for two contrasting environmental covariates : moisture index (which varied widely between taxa but had limited interaction with traits) and; topographic wetness (with less variation between taxa but stronger interaction with traits (cf. Fig. 1). In the Grampians (at top Fig. 6)) we can see that the trait-SDMs produced coefficients for Topographic Wetness similar in sign and magnitude to those from individual taxon regressions. Also taxa with high AUPRC/prevalence values tended to lie farther from the origin and closer to the 1:1 line, indicating that better predictions of occurrence (AUPRC) were associated with well-calibrated predictions of coefficients. Those patterns were not so evident for Moisture Index, where taxon regression responses varied widely but trait-SDM predictions did not capture that and varied little; (Fig. 6).

The correlation between trait-SDM predicted responses to Topographic Wetness in target regions show that some taxon responses were well predicted (lying in top right and lower left quadrants, and close to the 1:1 line). Taxa with high AUPRC/prevalence values were not always close to the 1:1 line, because the plots indicate responses to a single gradient at a time, whereas AUPRC/prevalence measures overall model performance. Some taxon responses were poorly predicted (e.g., in Victorian Alps, the sign was often wrong; positive responses were predicted by the trait-SDM while taxon regressions resulted in negative responses).

Trait-SDM responses to Moisture (Fig. 6, left panels) were less correlated with those from taxon regressions. Still, most responses were in the correct quadrant (i.e., correct sign). High AUPRC/prevalence predictions were generally associated with coefficients in the correct quadrant. In Jervis, it appears that taxa with low AUPRC/prevalence are dispersed widely in the taxon regression coefficients, without corresponding predictive coefficients. That is, taxa in Jervis varied widely in their responses to moisture index, but in a way that was not predicted by the trait-SDM from the Grampians.

Together, the evaluation of environmental response suggests that stronger trait-environment interactions produce well-calibrated response coefficients resulting in better predictions of occurrence. But weak trait-environment interactions resulted in less calibrated coefficients and were not clearly related to predicting occurrence well.

Discussion

We have demonstrated a method for using trait-SDMs to transfer knowledge from one taxon to another and from one region to another along with ways to measure and visualise the performance of such a transfer. We used that method to demonstrate that taxon environmental responses along gradients (i.e., whether they increase or decrease in probability of occurrence) can be predicted reasonably from their traits, our first general result. But some taxa and some regions were better predicted than others. Performance in target regions displayed no distance-decay from the model reference region, our second general result. Higher performance was related to well-calibrated environmental response predictions that resulted from strong trait-environment associations for AUPRC, but not so clearly for AUROC, our third general result. We discuss these

further below.

The performance we documented is notable, as we work within one genus, which makes the problem harder, because it potentially limits the trait range. Predictive performance using AUROC was comparable to that for spatial transferability within species (Randin *et al.* 2006): 54 tree species models transferred between Swiss and Austrian Alps resulting in median AUROC scores of 0.63 (minimum 0.44, interquartile range 0.55–0.72 and maximum 0.93) from Swiss to Austrian Alps and 0.65 in reverse (minimum 0.45, interquartile range 0.60–0.73 and maximum 0.83). Comparable performance is remarkable, given that our trait-SDM is blind to taxon.

Yet less is known about comparable performance using AUPRC, as it has been less used in SDM (Sofaer *et al.* 2019). Our experience suggests that AUPRC should be more widely used — it better reflected model calibration. Interpretation of AUPRC is confounded by dependence on prevalence, but it matches the problem of directing survey effort, and by expressing relative to the performance of a random classifier (= prevalence), interpretation is aided. Substantive interpretation of the scale and rules of thumb for judging performance under AUPRC and AUPRC/prevalence would be assisted by accumulating published model performance results. One could use any metrics based on the confusion matrix. Undoubtedly these would yield different answers in the detail. Yet our central message is unlikely to change: responses of new species in new regions are variably predicted, some quite well and some quite poorly.

We have presented a difficult case for transferability. The model reference region was peripheral, environmentally and geographically, to the wider region that we wished to predict to. This was partly historical—it is where we first built models—but also where

we had greatest confidence in the locational accuracy of the dataset to fit the model. Another way that our problem is hard is that we fitted the model on a small region and predicted to multiple regions over much greater extent. One expects better performance starting with a bigger (and wider-extent) dataset to predict to a small one. But that is not the problem that we believe presents itself to ecologists and practitioners, who are faced with larger areas where comparatively little is known and some intensively studied areas, from which one may wish to transfer knowledge. This is the problem that we have attempted to address, with some encouraging signs. Analogously, as one attempts to predict suitability under climate change, one is attempting to predict into conditions of great uncertainty from a smaller, well-understood current situation.

Better predictive performance of trait-SDM stemmed from well-calibrated environmental responses that could be predicted through strong trait-environment interactions, like Topographic Wetness in our study. Predicted responses to Moisture Index were not well calibrated, owing to larger variation between species but weak interactions with studied traits. A likely direction for improving performance would be a trait that modulates species performance along the Moisture Index gradient. Perhaps, traits better reflecting the water costs of photosynthetic capacity e.g., Rubisco-dependent carboxylation capacity (V_{cmax}) or Leaf Nitrogen per Area (Prentice *et al.* 2014).

Additionally, Total nitrogen responses were mainly negatively calibrated. Perhaps because the fit was accurate within the reference region, but that trait-environment relationship had no generality. Or the covariate was incorrectly selected, which could be because it was mistakenly chosen ahead of some other environmental covariate with which it was correlated, but which was more meaningful. Poorly calibrated responses

could also emerge when predicting to trait ranges outside those of the reference dataset. This suggests that some explanation for the good predictive performance we found lies in the good coverage of trait space among the taxa in our reference region relative to the target regions (Sequeira *et al.* 2018).

Surprisingly, we found no evidence to suggest declining transferability across geographic, compositional or environmental space in our study, nor little difference from the reference region. The random effects of region in models of performance were substantially lower than for taxa, which implies it is not about choosing the correct region from which to transfer. By contrast, about twice the improvement in model performance could be gained by considering taxa. We found that taxa that had better predictions tended to have lighter seeds. One speculative explanation draws on classical cost-benefit theory about community assembly along gradients of favourability (Orians & Solbrig 1977; Austin & T. M. Smith 1989; T. Smith & Huston 1989; Normand *et al.* 2009). According to that theory, the most productive/resource-acquisitive taxa are restricted to the most favourable sites. More tolerant species could occur in those sites but are competitively excluded from them by the resource-acquisitive taxa, hence they generally occur in less favourable sites. In our case, light seed mass reflects lower tolerance of hazards of seedling establishment, more limited by environment, and heavy seed mass species with tolerant seedlings are less limited by environment, with greater role of competition (Leishman 2001; Muller-Landau 2010). Our models include only environment, no biotic interactions.

Caveats and extensions

Performance of our models was hampered by two types of data limitations: spatial inaccuracy of occurrence data; and spatial models of environment used as covariates. In the first case, the occurrence data across our target regions result from compilations of survey campaigns over >30 years. Over that time, civilian GPS availability, accuracy and precision have improved substantially, meaning that older locations are less reliable. This limitation interacts with our second limitation, the need to use modelled environmental covariates. Our original modelling of the Grampians dataset utilised some field-measured covariates including rockiness and soil texture, which were strongly influential (Pollock *et al.* 2012), as soil texture was in semi-arid areas (Pollock *et al.* 2018). Landscape position can vary dramatically across lateral distances of tens of meters with potent effects on environmental variables related to soil depth, texture, nutrients and water availability as well as irradiance-mediated micro-climate (Austin & Van Niel 2011). And the DEM-based models that exist for such environmental variables do not approach what one can achieve with plot-based measurement. Our soil nitrogen responses would appear to be least reliable. So when combined with spatial inaccuracy of occurrence plots, capacity to predict relationships with environmental variables is diminished (Van Niel & Austin 2007).

These results give confidence in the value of traits to assist in the hard problem of predicting responses to environmental gradients for new species and new environmental conditions and regions. They deserve testing in different systems—other clades and landscapes. Factors likely contributing to the success here are that the reference region was environmentally diverse and the species there were functionally diverse, ranging

445 widely in all three traits.

446 **Supplements**

447 Supplement 1 Study area, data and covariates.

448 Supplement 2 Eucalypt traits.

449 Supplement 3 Model fitting and results.

450 Supplement 4 Performance metrics.

Acknowledgements

We thank Angela Stock, Martin Stock and Daniel White for assistance in the field. This work was supported by funding through Australian Research Council Centre of Excellence in Environmental Decisions, the Thomas Davies Research Grant from the Australian Academy of Science, and Eucalypt Australia. Field work was performed under permits from NSW National Parks & Wildlife Service SL101959, and the Victorian Department of Environment, Land, Water & Planning 10008557. Data come from lands of the following Indigenous Peoples: Djabwurung, Jardwadjali, Wauthurong, Woiworung, Boonwurrung, Taungurong, Kurnai, Bidwell, Waveroo, Wiradjuri, Jaimatang, Ngarigo, Yuin, Ngunawal, Gudungurra, Tharawal. We acknowledge them.

References

- Araujo, M. B. & Guisan, A. (2006). Five (or so) challenges for species distribution modelling. *Journal of Biogeography* 33, 1677–1688.
- Austin, M. P. (2002). Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological Modelling* 157, 101–118.
- Austin, M. P., Nicholls, A. O. & Margules, C. R. (1990). Measurement of the realized qualitative niche: environmental niches of five *Eucalyptus* species. *Ecological Monographs* 60, 161–177.
- Austin, M. P., Pausas, J. G. & Nicholls, A. O. (1996). Patterns of tree species richness in relation to environment in southeastern New South Wales, Australia. *Australian Journal of Ecology* 21, 154–164.
- Austin, M. P. & Smith, T. M. (1989). A new model for the continuum concept. *Vegetatio* 83, 35–47.
- Austin, M. P. & Van Niel, K. P. (2011). Improving species distribution models for climate change studies: variable selection and scale. *Journal of Biogeography* 38, 1–8.
- Briscoe, N. J., Kearney, M. R., Taylor, C. A. & Wintle, B. A. (2016). Unpacking the mechanisms captured by a correlative species distribution model to improve predictions of climate refugia. *Global Change Biology* 22, 2425–2439.
- Brown, A. M., Warton, D. I., Andrew, N. R., Binns, M., Cassis, G. & Gibb, H. (2014). The fourth-corner solution—using predictive models to understand how species traits interact with the environment. *Methods in Ecology & Evolution* 5, 344–352.
- Colwell, R. K. & Rangel, T. F. (2009). Hutchinson’s duality: the once and future niche. *Proceedings of the National Academy of Sciences* 106, 19651–19658.

- Cover, T. M. & Thomas, J. A. (2006). *Elements of information theory*. 2nd ed. Hoboken, New Jersey: John Wiley & Sons, Inc.
- Department of the Environment & Energy (2012). *Revision of the Interim Biogeographic Regionalisation for Australia (IBRA), Version 7*. Department of the Environment & Energy, Canberra.
- Dobrowski, S. Z., Thorne, J. H., Greenberg, J. A., Safford, H. D., Mynsberge, A. R., Crimmins, S. M. *et al.* (2011). Modeling plant ranges over 75 years of climate change in California, USA: temporal transferability and species traits. *Ecological Monographs* 81, 241–257.
- Dorrough, J., McIntyre, S. & Scroggie, M. (2011). Individual plant species responses to phosphorus and livestock grazing. *Australian Journal of Botany* 59, 670–681.
- Dorrough, J. & Scroggie, M. P. (2008). Plant responses to agricultural intensification. *Journal of Applied Ecology* 45, 1274–1283.
- Elith, J. & Leathwick, J. R. (2009). Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, & Systematics* 40, 677–697.
- Evans, J. P., Ji, F., Lee, C., Smith, P., Argüeso, D. & Fita, L. (2014). Design of a regional climate modelling projection ensemble experiment—NARClIM. *Geoscientific Model Development* 7, 621–629.
- 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.

- Fithian, W., Elith, J., Hastie, T. & Keith, D. A. (2015). Bias correction in species distribution models: pooling survey and collection data for multiple species. *Methods in Ecology & Evolution* 6, 424–438.
- Gelfand, A. E., Silander, J. A., Wu, S., Latimer, A., Lewis, P. O., Rebelo, A. G. *et al.* (2006). Explaining species distribution patterns through hierarchical modeling. *Bayesian Analysis* 1.1, 41–92.
- Grundy, M. J., Rossel, R. A. V., Searle, R. D., Wilson, P. L., Chen, C. & Gregory, L. J. (2015). Soil and landscape grid of Australia. *Soil Research* 53, 835–844.
- Jamil, T., Ozinga, W. A., Kleyer, M. & Braak, C. J. F. ter (2013). Selecting traits that explain species–environment relationships: a generalized linear mixed model approach. *Journal of Vegetation Science* 24, 988–1000.
- Laughlin, D. C., Joshi, C., Bodegom, P. M. van, Bastow, Z. A. & Fulé, P. Z. (2012). A predictive model of community assembly that incorporates intraspecific trait variation. *Ecology Letters* 15, 1291–1299.
- Legendre, P. & Legendre, L. F. J. (2012). *Numerical ecology*. 3rd ed. Developments in Environmental Modelling 24. Amsterdam: Elsevier.
- Leishman, M. R. (2001). Does the seed size/number trade-off model determine plant community structure? An assessment of the model mechanisms and their generality. *Oikos* 93, 294–302.
- Miller, J. E. D., Damschen, E. I. & Ives, A. R. (2019). Functional traits and community composition: A comparison among community-weighted means, weighted correlations, and multilevel models. *Methods in Ecology & Evolution* 10, 415–425.
- Morán-Ordóñez, A., Lahoz-Monfort, J. J., Elith, J. & Wintle, B. A. (2017). Evaluating 318 continental-scale species distribution models over a 60-year prediction horizon:

what factors influence the reliability of predictions? *Global Ecology & Biogeography*
26, 371–384.

Muller-Landau, H. C. (2010). The tolerance–fecundity trade-off and the maintenance of
diversity in seed size. *Proceedings of the National Academy of Sciences* 107, 4242–4247.

Normand, S., Treier, U. A., Randin, C., Vittoz, P., Guisan, A. & Svenning, J.-C. (2009).
Importance of abiotic stress as a range-limit determinant for European plants:
insights from species responses to climatic gradients. *Global Ecology & Biogeography*
18, 437–449.

Orians, G. H. & Solbrig, O. T. (1977). A cost-income model of leaves and roots with
special reference to arid and semiarid areas. *American Naturalist* 111, 677–690.

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

Ovaskainen, O., Tikhonov, G., Norberg, A., Guillaume Blanchet, F., Duan, L., Dunson, D.
et al. (2017). How to make more out of community data? A conceptual framework and
its implementation as models and software. *Ecology Letters* 20.5, 561–576.

Perez-Harguindeguy, N., Diaz, S., Garnier, E., Lavorel, S., Poorter, H., Jaureguiberry, P.
et al. (2016). Corrigendum to: new handbook for standardised measurement of plant
functional traits worldwide. *Australian Journal of Botany* 64, 715–716.

Pollock, L. J., Kelly, L. T., Thomas, F. M., Soe, P., Morris, W. K., White, M. *et al.* (2018).
Combining functional traits, the environment, and multiple surveys to understand
semi-arid tree distributions. *Journal of Vegetation Science* 29, 967–977.

Pollock, L. J., Morris, W. K. & Vesk, P. A. (2012). The role of functional traits in species
distributions revealed through a hierarchical model. *Ecography* 35, 716–725.

- Prentice, I. C., Dong, N., Gleason, S. M., Maire, V. & Wright, I. J. (2014). Balancing the costs of carbon gain and water transport: testing a new theoretical framework for plant functional ecology. *Ecology Letters* 17, 82–91.
- Randin, C. F., Dirnböck, T., Dullinger, S., Zimmermann, N. E., Zappa, M. & Guisan, A. (2006). Are niche-based species distribution models transferable in space? *Journal of Biogeography* 33, 1689–1703.
- Sequeira, A. M. M., Bouchet, P. J., Yates, K. L., Mengersen, K. & Caley, M. J. (2018). Transferring biodiversity models for conservation: opportunities and challenges. *Methods in Ecology & Evolution* 9, 1250–1264.
- Slee, A. V., Brooker, M. I. H., Duffy, S. M. & West, J. G. (2006). *Euclid: eucalypts of Australia*. 3rd ed. CSIRO. Canberra.
- Smith, T. & Huston, M. (1989). A theory of the spatial and temporal dynamics of plant communities. *Vegetatio* 83, 49–69.
- Sofaer, H. R., Hoeting, J. A. & Jarnevich, C. S. (2019). The area under the precision-recall curve as a performance metric for rare binary events. *Methods in Ecology & Evolution* 10, 565–577.
- The State of Victoria, Department of Environment, Land, Water and Planning (2018). *Victorian Biodiversity Atlas*.
- Thomas, F. M. & Vesk, P. A. (2017). Are trait-growth models transferable? Predicting multi-species growth trajectories between ecosystems using plant functional traits. *PloS One* 12, e0176959.
- Van Niel, K. P. & Austin, M. P. (2007). Predictive vegetation modeling for conservation: Impact of error propagation from digital elevation data. *Ecological Applications* 17, 266–280.

577 Warton, D. I., Shipley, B. & Hastie, T. (2015). CATS regression—a model-based
578 approach to studying trait-based community assembly. *Methods in Ecology &*
579 *Evolution* 6, 389–398.

580 Yates, K. L., Bouchet, P. J., Caley, M. J., Mengersen, K., Randin, C. F., Parnell, S. *et*
581 *al.* (2018). Outstanding challenges in the transferability of ecological models. *Trends*
582 *in Ecology & Evolution* 9, 1250–1264.

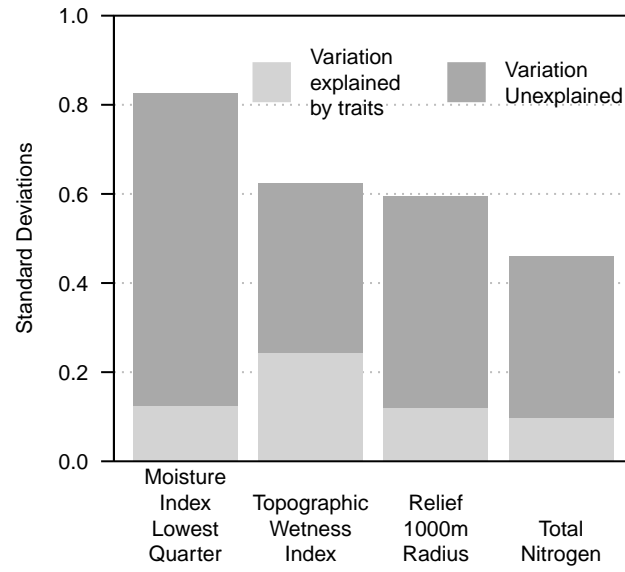


Figure 1: Variance components for environmental response hyperparameters for the Grampians trait-SDM. Bars represent the between-taxa variability in responses to environmental gradients with the component explained by their median trait values in light grey and the unexplained proportion in dark grey. Taller bars represent gradients to which the responses are more variable.

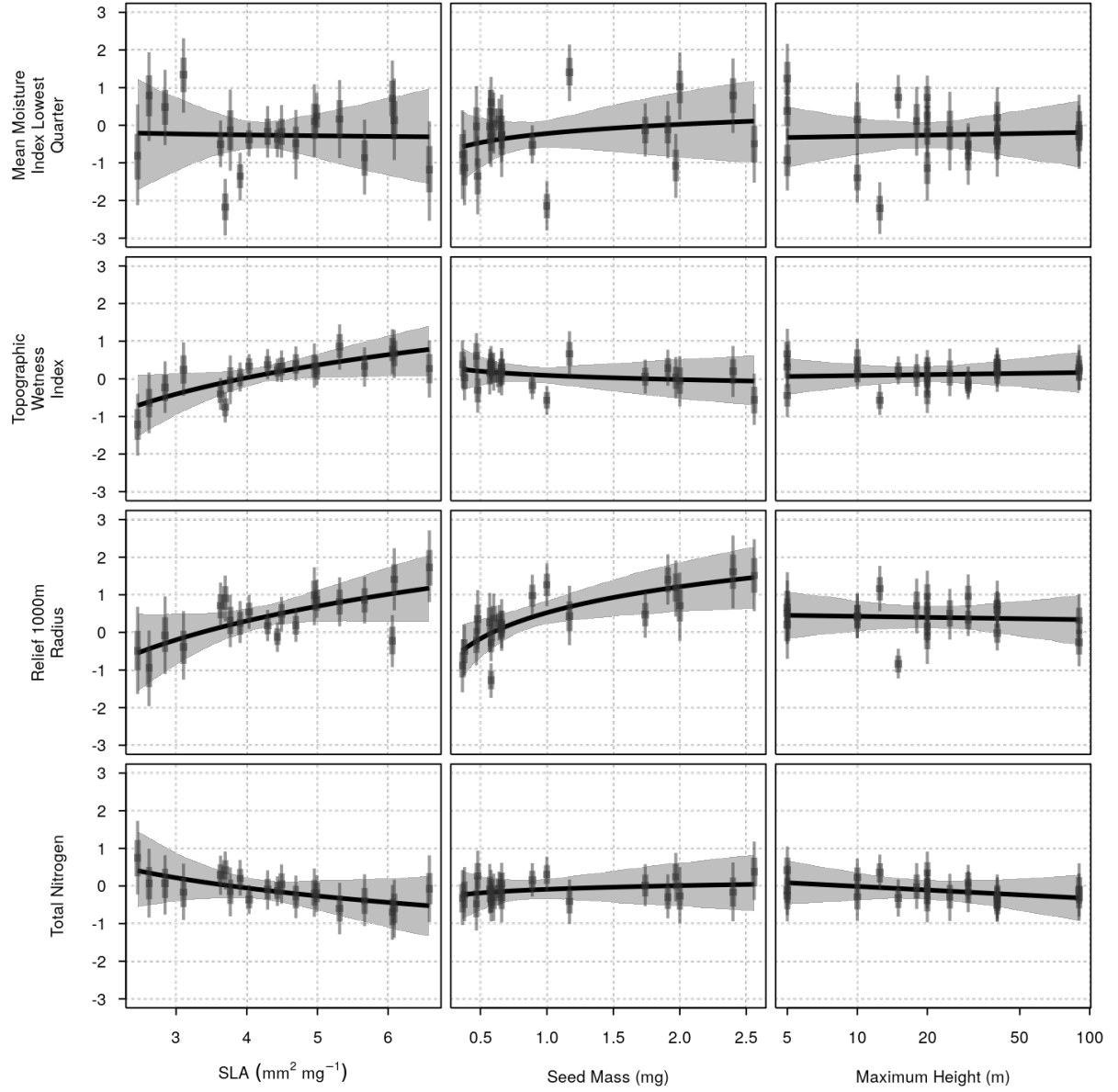


Figure 2: Relationships between median trait values and environmental responses. Black lines show the expected response coefficient as a function of median trait value. Grey envelopes represent the 95% credible bounds of the estimate. Boxplots are the estimated partial responses of each modelled taxon (positioned along the x-axis at their measured trait medians) showing the mean, 50% and 95% credible intervals with segments of decreasing width.

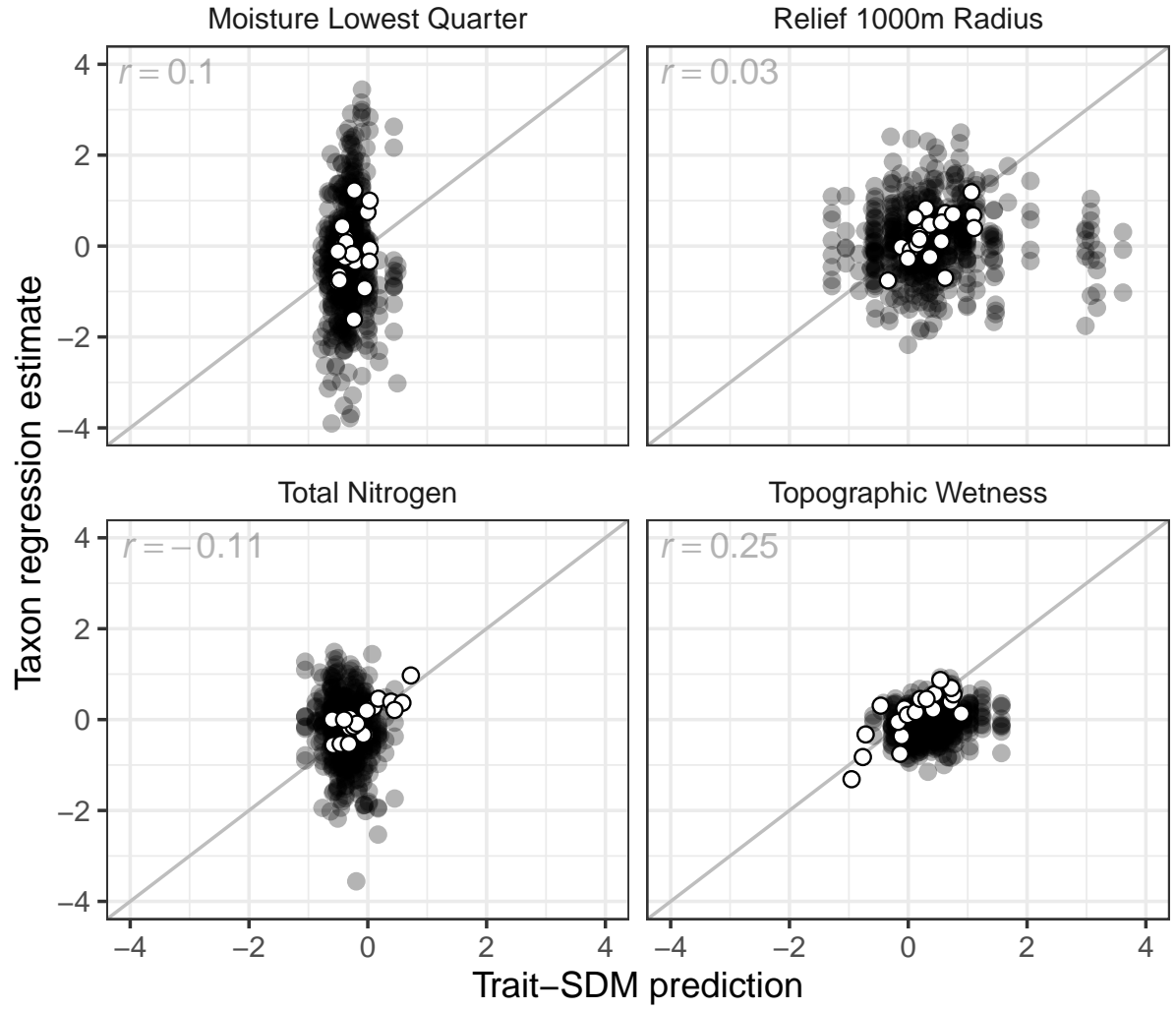


Figure 3: Calibration relationships between environmental response coefficients predicted by the Trait-SDM and fitted by individual taxon regressions. White symbols in foreground are for the reference region (Grampians) and filled transparent grey symbols are for the target regions.

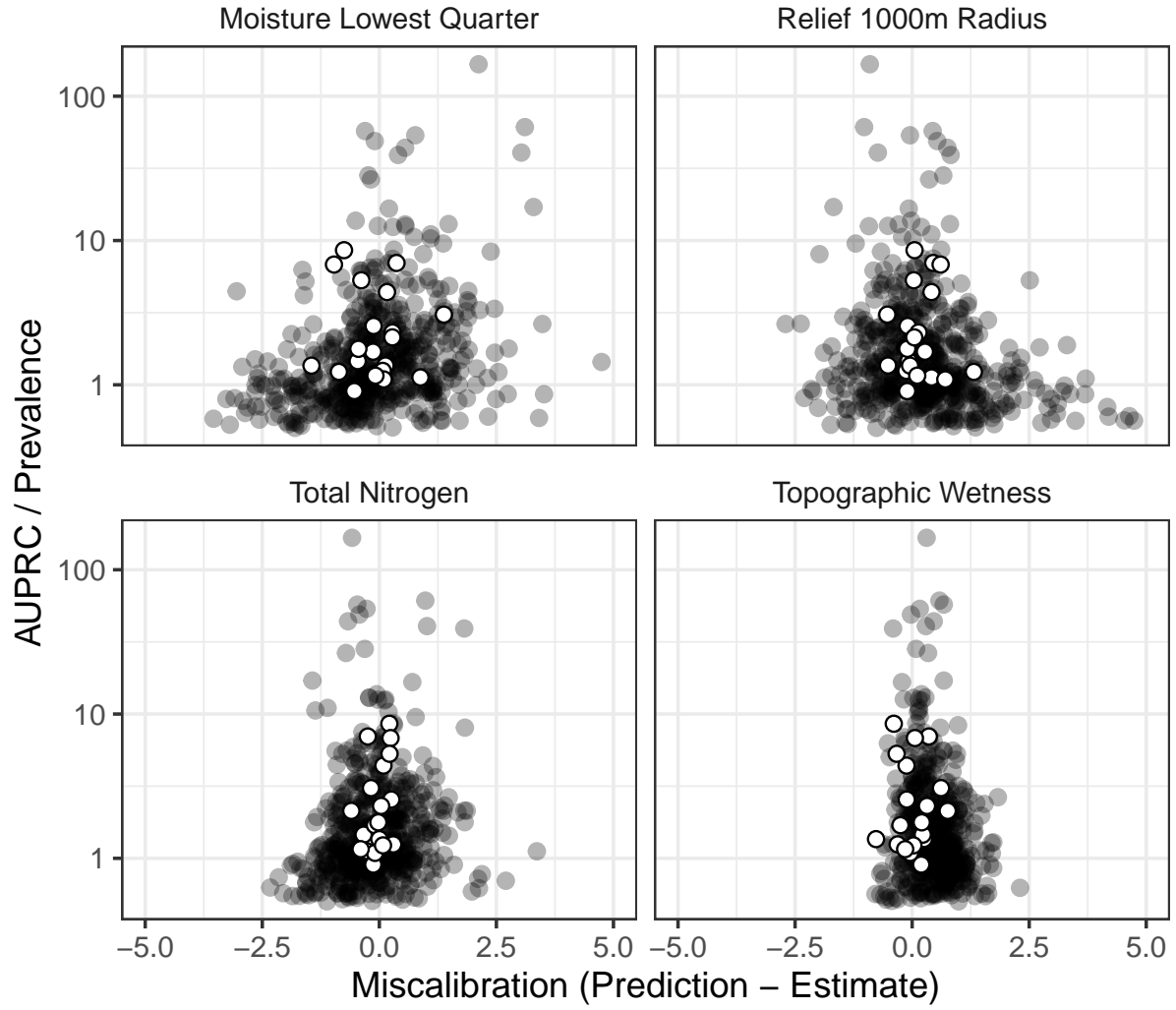


Figure 4: Performance of Trait-SDMs compared to a random classifier (AUPRC / prevalence) plotted against miscalibration of environmental response coefficients. White symbols in foreground are for the reference region (Grampians) and filled grey symbols are for the target regions.

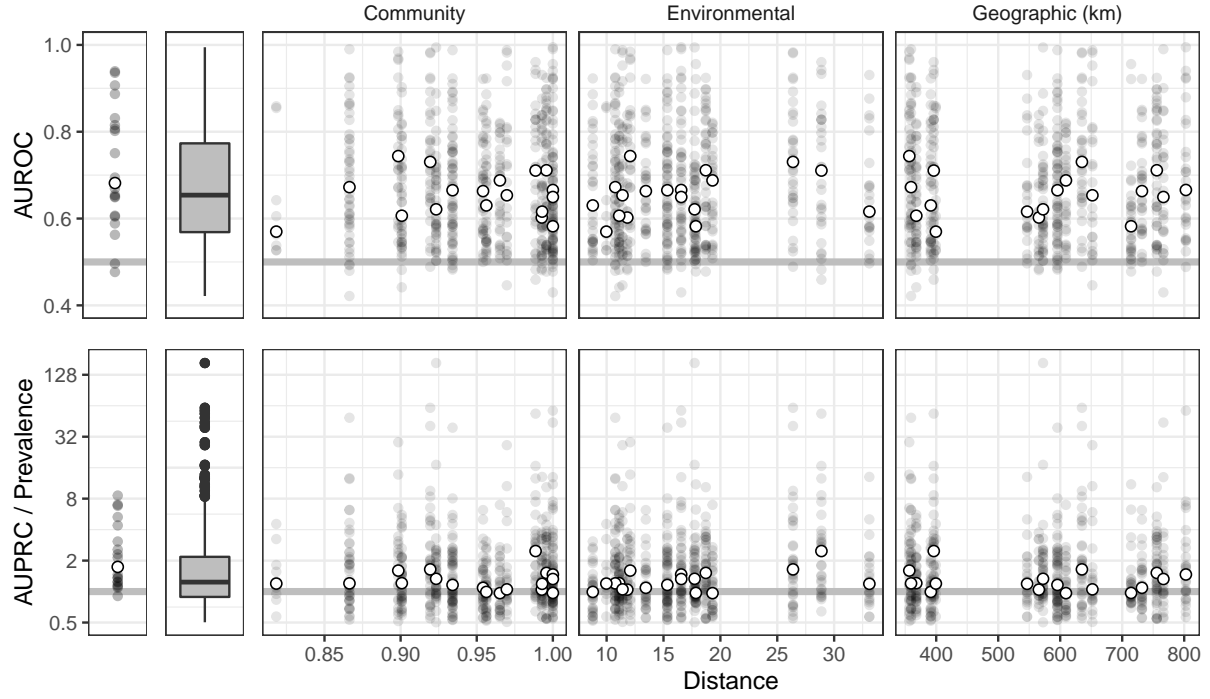


Figure 5: Relationship between within region, taxon-specific performance metrics (AUROC and AUPRC/prevalence) and the distance from the reference to each target region. Distance is measured as: Jaccard dissimilarity of communities, Kullback-Leibler distance of modelled environmental space, and distance in kms between centroids. White circles are the mean performance in each region. Leftmost panels show the performance metrics for the reference region, Gariwerd-Grampians. Boxplots show the distribution of within-region taxon-specific performance across all the target regions.

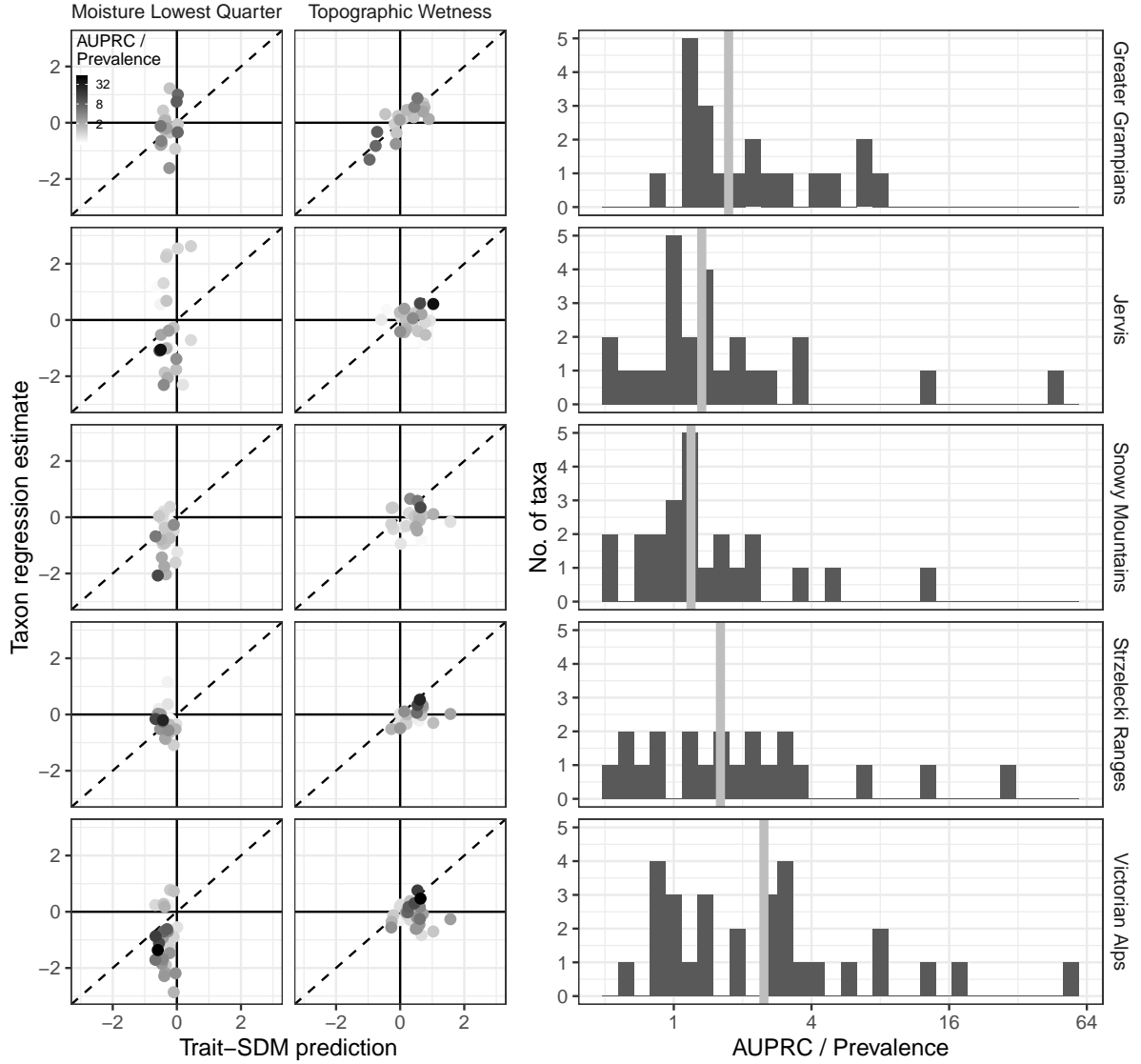


Figure 6: Left panels: Predicted responses from the trait-SDM versus taxon regression estimates. The top row of panels are the reference region, Greater Grampians. The four rows below are other regions in the southeast. Each point represents the response of a taxon within a given region. The position on the y-axis is the expected response predicted trait-SDM conditional on the median trait values. The position on the x-axis is the estimate of the response from taxon regressions of the taxa within the regions. Each point's black level indicates the area under the precision recall curve statistic (AUPRC) divided by the prevalence for the taxon in the region's plots based on the predicted probabilities of occupancy according to the trait-SDM. Right panels: Distribution of taxon-specific AUPRC divided by prevalence for predicted probabilities of occupancy conditional on traits for the regions. Grey line is the median AUPRC divided by prevalence value across the taxa in the region.