The need for reliable prediction of species distributions dependent upon traits has been hindered by a lack of model transferability testing. We tested the predictive capacity of trait-SDMs by fitting hierarchical generalised linear models with three trait and four environmental predictors for 20 eucalypt taxa in a reference region. We used these models to predict occurrence for a much larger set of taxa and target areas (82 taxa across 18 target regions) in south-eastern Australia. Median predictive performance for new species in target regions was 0.65 (area under receiver operating curve) and 1.24 times random (area under precision recall curve). Prediction in target regions did not worsen with increasing geographic, environmental or community compositional distance from the reference region, and was improved with reliable trait–environment relationships. Transfer testing also identified trait–environment relationships that did not transfer. These results give confidence that traits and transfer testing can assist in the hard problem of predicting environmental responses for new species, environmental conditions and regions.

Keywords: community modelling, environmental filtering, Eucalyptus, functional traits, generalisation, prediction

Introduction

Improving our understanding of where species occur and why, are core aims in both fundamental and applied ecology. One approach to addressing this priority area has been through species distribution modelling (SDM), which has been the focus of substantial research over the past two decades (Elith and Leathwick 2009). The vast majority of SDM work is correlative, based on fitting statistical models to datasets of species occurrences and making predictions from these models. This approach is commonly used to predict to new situations (space, time and environments), but the ability of these models to be transferred to new conditions is rarely tested. Correlative models can fail to transfer owing to invalid assumptions that distributions represent niches (e.g. biotic interactions, dispersal limitation and non-equilibrium distributions) and due to modelling failures (e.g. chance correlations and extrapolation) (Colwell and Rangel 2009, Briscoe et al. 2016). Mechanistic models are more explicit about including process (Briscoe et al. 2019), and may include climate dependent phenology (Morin et al.

© 2020 The Authors. Ecography published by John Wiley & Sons Ltd on behalf of Nordic Society Oikos
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
cies vary in their responses. While these hierarchical models can help both establish the functional role of traits and the environment. Incorporating traits into hierarchical models can help both establish the functional role of traits and the environment. 

In addition to transferring models for a particular species to new areas or time periods, there is also value in being able to transfer knowledge across species given that there roughly 300,000 angiosperm plant species worldwide, many lacking accurate distributional data. In both correlative and mechanistic modelling, species are typically modelled one at a time without the ecologically relevant information contained in other species. However, hierarchical (or multi-species) models can be used to model multiple species simultaneously by assuming some commonality of response among species and sharing statistical strength between species (Gelfand et al. 2006, Dorrough et al. 2011, Ovaskainen and Soininen 2011), which is important as many species are rare with limited occurrence records. This is done by defining hyperparameters that specify distributions for model parameters. Multi-level models are not restricted to correlative models, they can also be useful in parameterising process-explicit models, such as multi-species demographic SDM (Pagel et al. 2020). Parameter estimates can then be extracted and statistically explained by species traits in a post hoc analysis, providing a functional explanation of species responses across environmental gradients (Treurnicht et al. 2020). Yet, to be operationally useful, it is not enough for traits to simply indicate response, they need also to predict response transferably (Sequeira et al. 2018). We use the term ‘response’ in the statistical sense, rather than in a mechanistic sense. By adding species submodels into multi-level models, such that some parameter of the model is a function of some species property (e.g. demographic, life-history or functional traits), then the basis of predicting responses for new species from traits is enabled (Vesk 2013).

Plant functional traits can be used as predictors of species distribution along environmental gradients (Dorrough and Scroggie 2008, Laughlin et al. 2012, Pollock et al. 2012, 2018, Jamil et al. 2013, Brown et al. 2014, Ovaskainen et al. 2017, Miller et al. 2019). Trait-SDMs 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 better explain why species vary in their responses. While these hierarchical models are still correlative, with many of the shortcomings mentioned before, they contain additional information on the link between species and their modelled environment that could be highly relevant when predicting to new species or new conditions.

The ability to transfer ecological models from reference to target regions or environmental conditions (Randin et al. 2006) is plainly useful for decisions in conservation and natural resource management when data and understanding are scarce (Sequeira et al. 2018). Transfer between species is at least as important. For example, if the distribution of threatened species is unknown, but we know some of the traits for that species, we could direct survey effort by predicting where that species might occur based on a model for other species in that same area, or based on a model trained elsewhere (e.g. in a more intensively surveyed area). Moreover, if trait–environment models can be demonstrated to have improved predictive capacity, then they may also be preferable for studying species responses to environmental change in situ.

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) (Pollock et al. 2012, 2015, 2018). Drawing on the leaf–height–seed (LHS) scheme of Westoby (1998), in Pollock et al. (2012), we incorporated leaf–height–seed (LHS) traits (specific leaf area (SLA), maximum height and seed mass) into generalised linear mixed models (GLMMs) across gradients in Gariwerd-Grampians Ranges, Victoria, Australia using seven climatic, topographic and edaphic variables. LHS traits are key traits underlying variation in function and performance among the world’s plants (Westoby and Wright 2006, Díaz et al. 2016), capturing leaf economics (Wright et al. 2004), establishment strategies (Leishman 2001, Muller-Landau 2010) and competitive height races under productivity and disturbance (Falster and Westoby 2003). Subsequent work has shown the importance of LHS traits in global analyses (Wright et al. 2004, Moles et al. 2007, 2009, Reich et al. 2007, Brueelheide et al. 2018). In Pollock et al. (2012), we found strong trait–environment relationships with these trait-SDMs, in particular: heavier-seeded species were more likely to occur in sandy (cf. clay) soils; low SLA species were more likely to occur in sites with higher rock cover (and less exploitable soil volume). Species occurrence across gradients of irradiance and rock cover within Gariwerd-Grampians was predicted from the trait-SDM for combinations of SLA and maximum height, but predictions were not evaluated (Pollock et al. 2012).

Here we ask whether such trait-SDMs can be transferred between regions and between species (Fig. 1). We have in mind the problem facing an ecologist standing in a forested landscape spanning climatic, edaphic and topographic gradients. Lacking any distributional data for a particular species, but knowing some traits of the species, which way should the ecologist head to be more likely to find the species? Where should survey effort be directed? Trait-SDMs could potentially provide a first-cut model to help our ecologist, by making trait-based predictions using a model trained on
other species in other regions. Such predictive transfer across regions and species is the subject of this paper. Can we predict the distributions of target species (that lack distribution data) in new target regions based only on their traits and the modelled trait–environment interactions from a small outlying reference region? This is a challenging out-of-sample test (Hooten and Hobbs 2015) – previous tests of single-species SDM transferability between regions (Randin et al. 2006) or time periods (Dobrowski et al. 2011, Morán-Ordóñez et al. 2017), report performance ranging from failure to excellent, but with most less than fair. Those studies were not testing trait-SDMs predicting to new species. The only study of trait-SDM that we are aware of employing internal, between-species cross-validation of trait-SDM showed good performance between species, within one single dataset of insects, but still within only one region (Brown et al. 2014). The closest analogues of our approach here can be found in trait-based models of height growth which employed internal, between-species cross-validation (Thomas et al. 2019) and between-region transfer testing (Thomas and Vesk 2017).

In this paper, we evaluated potential (and its variation) to predict species distributions in environmentally complex regions where we lack distribution data for target taxa or target regions but know just three species-level (LHS) traits. We expected that, 1) trait-SDMs will make reasonable predictions, based on model fit within previous studies (Pollock et al. 2012, 2018, Jamil et al. 2013, Miller et al. 2019). 2) Models should make better predictions when environmental responses are well-calibrated and when trait–environment relationships are strong. 3) Model performance will decline when predicting to target regions that are further from the reference region or with more dissimilar environments or species composition. To undertake this evaluation, we fitted a GLMM to 20 taxa in a reference region Gariwerd-Grampians, a small outlier of Australia’s Great Dividing Range, and then used this model to predict environmental responses and distributions for 82 target taxa in 18 target regions across over 117 900 km² along the Great Dividing Range in southeast Australia. We also evaluated metrics of predictive performance that have been proposed as appropriate for testing within-species predictive transfer: the area under the receiver operating curve (Sequeira et al. 2018); and area under the precision recall curve, which is specifically recommended for guiding survey effort for low-moderate prevalence species (Sofaer et al. 2019). In addressing these three expectations, we hope to demonstrate a general method for using trait-SDMs (of any form) to transfer knowledge from one taxon to another and from one region to another, as well as ways to measure and visualise the performance of such transfer.

Methods

Study system and datasets

Our study uses trait-SDM fitted in one reference region to multi-species occurrence data using species traits and environmental covariates (Fig. 1). With new trait data for target taxa in new target regions, we predict their environmental response coefficients. Then occurrences are predicted from environmental responses and environmental covariates and
are evaluated with an independent plot dataset (Fig. 1). The data and steps are described below.

Geography
Our work was conducted in southeastern Australia (Supplementary material Appendix 1 Fig. A1.1), which has a temperate Köppen climate. We used subregions from the the Interim Biogeographic Regionalisation of Australia (IBRA ver. 7) (Dept of the Environment and 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 (2400 km²). 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. The 20 eucalypt taxa found there are listed in Supplementary material Appendix 2 Table A2.1.

Our target (or test or transfer) regions cover a roughly triangular area of 117 900 km² spanning five degrees of latitude and five degrees of longitude. This yielded 18 target regions ranging in area from 400 to 17 300 km², containing 10–72 eucalypt taxa each (Supplementary material Appendix 1 Table A1.1). The target regions cover diverse habitats across the Great Dividing Range, from coastal plains and valleys through escarpments and low plateaus and mountain ranges and inland slopes. Köppen class climates are temperate with mild or warm summers. Vegetation types include woodlands, dry and wet forests.

Occurrence data
From all datasets we extracted binary presence–absence data. For the reference region, Gariwerd-Grampians, we used the plot dataset from Pollock et al. (2012). Briefly, ~460 plots were surveyed for occurrence of eucalypt species 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 occurrence data for testing in the transfer regions were compiled from the Victorian Biodiversity Atlas (The State of Victoria, Dept 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 900–1000 m²) plots with the presence of all woody species recorded. Subspecies were recognised and so, ‘taxon’ is the more correct term, however we occasionally use the term ‘species’ for simplicity.

Trait data
We used the leaf–height–seed (LHS) traits, specific leaf area (SLA), seed mass and maximum attainable height (Westoby 1998), as in our previous trait-SDM analyses in Gariwerd-Grampians (Pollock et al. 2012, 2015). Traits were measured according to standard protocols (Perez-Harguindeguy et al. 2016). Sampling of plants and trait measurement for the Gariwerd-Grampians is described in Pollock et al. (2012) and those data appear in Supplementary material Appendix 2 Table A2.1.

Environmental data
In this study we only used environmental data available as GIS layers throughout the target 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 region dataset. Covariates were selected from a large set using a combination of cluster analysis and discrimination power, informed by our ecological intuition that aspects of climatic moisture availability, landscape position affecting water and soil shedding or accumulation, soil fertility, texture and rockiness were each important (Pollock et al. 2012, 2018) (see Supplementary material Appendix 1 Section: Selecting environmental covariates, Supplementary material Appendix 1 Fig. A1.2). 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. Supplementary material Appendix 1 Section: Characterising the reference and target regions, presents maps of the covariates across the reference and training regions. Supplementary material Appendix 1 Fig. A1.3, with bivariate plots (Supplementary material Appendix 1 Fig. A1.4), principal components analysis (Supplementary material Appendix 1 Fig. A1.5) and test region divergence from the reference region (Supplementary material Appendix 1 Fig. A1.6).

Model building

The trait-SDMs were built using the same approach as in Pollock et al. (2012, 2018) and detailed in Supplementary material Appendix 3 Section: Reference model fitting. We built generalized linear mixed models (GLMM) with intercepts and slopes varying by taxon, and fixed effects for traits modulating those slopes. These are also known as hierarchical models or multi-level models. Broadly similar approaches are described in Jamil et al. (2013), Brown et al. (2014) 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 (SLA, SM, MH) and taxon-level random effects.

Statistical power was a key consideration for model fitting in the reference region. 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 (Pollock et al. 2012, 2015). We chose to use four environmental covariates yielding 12 trait–environment interactions, which was moderately high complexity for our statistical power. Our aim was not to determine the best possible trait-SDM for the Gariwerd-Grampians, but rather to fit one that was similar to that developed previously (Pollock et al. 2012), because our focus was on how to test predictive transfer. Supplementary material Appendix 3 Section: Expected trait–environment interactions, presents our ecological understanding of the environmental variables used, how they relate to those from Pollock et al. (2012), and our expectations for the relationships to be inferred from the trait-SDM are illustrated in Supplementary material Appendix 3 Fig. A3.1.

Measuring and comparing model performance

The trait-SDMs were trained on the trait and occurrence data for our reference region, Gariwerd-Grampians and then used to make predictions. We predicted using the fitted coefficients within the reference region, Gariwerd-Grampians in two ways: a) based only on the traits of the Gariwerd-Grampians taxa, without taxon identities; and b) to the Gariwerd-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 Gariwerd-Grampians, were not associated with the traits we used. We then made predictions to our target regions with our fitted trait-SDM. We predicted occurrences for all target taxa and regions based only on their traits, as in the first test within the reference region. These out-of-sample predictions are the main part of our study.

Performance measurement of presence–absence predictions

Predictive performance was measured with two metrics using the confusion matrix of prediction and observations. First, the area under the receiver operator curve statistic (AUROC, often presented as the acronym AUC, but spelled out here and throughout this paper for precision) (Fielding and Bell 1997), as used and recommended in studies of transferability (Randin et al. 2006, Sequeira et al. 2018). This can be interpreted as the probability that for a randomly chosen pair of plots consisting of one presence and one absence, the model would correctly rank their probability of occurrence. We also examined the area under the precision recall curve (AUPRC), which has advantages in situations where objects are rare, and it is proposed to map well on to the problem of directing survey effort (Søfaer et al. 2019). These are detailed in Supplementary material Appendix 4 Section: Performance metrics for distributions.

To evaluate whether predictive performance declined farther from the reference region we used three measures. We used the geographic distance in kilometres between the centroids of the reference and target regions. Community composition dissimilarity was measured with Jaccard’s index (Legendre and Legendre 2012). Environmental dissimilarity was measured with Kullback–Leibler divergence (Cover and 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: along 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 (Supplementary material Appendix 3 Section: Taxon models for target taxa and regions). We called these taxon- and region-specific models ‘taxon regressions’ to
avoid confusion. Taxon regressions were used to estimate coefficients for comparison with the coefficients from the trait-SDM, and serve as ‘gold standards’ to evaluate how the trait-SDM is making potentially inaccurate predictions. We asked whether miscalibration of the predicted 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 (|predicted coefficient - 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. These are described in Supplementary material Appendix 4 Section: Models of performance measures.

Results

Trait-SDM fitting in the reference region, Gariwerd-Grampians

We found that some environmental variables were important predictors, and that traits helped explain these responses. For example, taxa varied most in their response to moisture index (-0.8 SD), but traits explained little (15%) of that variance (Fig. 2). By contrast, the between-taxon variance in effect of topographic wetness (-0.6 SD) was relatively well explained (30%) by traits. Taxa with thicker, denser leaves were less likely to be found in topographically wet areas (Fig. 3, second row). Heavier seeded taxa responded positively along gradients of increasing ruggedness, while small seeded taxa responded negatively to ruggedness (Fig. 3, third row, center). Taxa varied less in their responses to topographic relief and total nitrogen, but traits explained about 20% of that variance. Supplementary material Appendix 3 Section: Extended trait-SDM results contains modelling results from fitting the trait-SDM to the reference region (Gariwerd-Grampians) including plots of taxon coefficients (Supplementary material Appendix 3 Fig. A3.2), and as well as the estimated fixed effects (Supplementary material Appendix 3 Table A3.2) and random effects (Supplementary material Appendix 4 Table A4.2).

Calibration of predicted environmental responses

Trait-SDM responses to topographic wetness and total nitrogen, topographic relief were also well-calibrated within the reference region (Gariwerd-Grampians) reflecting the variance components analysis (Supplementary material Appendix 4 Fig. A4.1). Environmental responses did not transfer as well to target regions. Response coefficients were less well-calibrated with the ‘gold standard’ taxon regressions; correlations were weaker, most positive for topographic wetness (r = 0.25) and moisture index (r = 0.10) (Supplementary material Appendix 4 Fig. A4.1). Low SLA and high seed mass taxa were consistently found to have negative responses to topographic wetness, in the trait-SDM and taxon regressions for target regions (Supplementary material Appendix 4 Fig. A4.2).

Predictive performance for target taxa and regions

Predictive performance of the trait-SDM within the reference region was, for some species, comparable to a model that included a taxon random effect, the ‘gold standard’ (Supplementary material Appendix 4 Fig. A4.5). However, in most cases the trait-SDM performed less well (Supplementary material Appendix 3 Section: Performance within the reference region, Gariwerd-Grampians).

The models performed very differently for different target taxa and target regions according to both the area under the receiver operating curve (AUROC) and area under the precision recall curve (AUPRC). Predictive performance varied more among target taxa within target regions than between regions (Fig. 4). Median AUROC = 0.65 and interquartile range 0.57–0.77 (Fig. 4). 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 (Supplementary material Appendix 4 Fig. A4.3). 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 (Supplementary material Appendix 4 Section: Area under the precision recall curve). According to AUPRC, most predictions were better than random, with
Median AUPRC = 1.24 times random, and interquartile range 0.89–2.18 times as good as random. Predictive performance was not related to geographic nor environmental distance nor compositional dissimilarity from the reference region (Fig. 4). Neither was performance within the reference region clearly higher than target regions (Fig. 4). For the ten taxa that occurred in both the reference region and target regions, AUROC and AUPRC values were similarly, highly variable (Supplementary material Appendix 4 Fig. A4.6). In other words, taxa that were present in the training dataset did not have consistently higher predictive performance in target regions than target taxa that only occurred in the target regions (and not in the reference region).

Relations between predictive performance and trait-SDM calibration
We found better predictive performance of trait-SDM when the environmental responses were better calibrated and worse
performance when responses were miscalibrated. This effect of miscalibration was stronger for predictive performance measured by AUPRC/prevalence (Fig. 5), and less so for AUROC.

**Explaining variation in predictive performance of target taxa and regions**

In a generalised linear mixed effects model evaluating predictive performance (AUPRC/prevalence) as a function of miscalibration and distance measures, the intercept of 0.43 indicated that, with average miscalibration, predictions to target taxa in target regions performed 1.5 times as good as random. The negative effect of trait-SDM miscalibration on predictive performance (Fig. 5) was most important for topographic wetness and topographic relief (Supplementary material Appendix 4 Table A4.1). Miscalibration effects on AUROC were also negative, but uncertain. This indicates that AUPRC/prevalence is a better measure of predictive performance than AUROC for our case, because it better reflects well-calibrated models.

In contrast to miscalibration, those models explaining variation in predictive performance did not reveal performance decline with distance measures between the reference and target regions (Supplementary material Appendix 4 Table A4.1). In other words, the models built from the reference region transferred similarly well regardless of how different the target regions were to the reference region.

Predictions of particular taxa did not perform consistently across target regions. This is evident because residual variance (measured as standard deviations) was approximately twice that of the taxon-level random effect, which was greater than region-level random effect (Supplementary material Appendix 4 Table A4.2). Taxon-level random effects were weakly, negatively correlated with seed mass ($r = -0.30$, 95% CI ($-0.49$, $-0.10$), 80 df) (Supplementary material Appendix 4 Fig. A4.7). In Supplementary material Appendix 4 Section: Probing predictive performance for some regions and environments, we examine the predictive performance across a subset of regions and environmental variables, illustrating how higher performance was associated with better calibrated models.

**Discussion**

We have demonstrated a predictive framework for using trait-SDMs to transfer knowledge from reference taxa to target taxa and from one reference region to other target regions, along with ways to measure and visualise the performance of such transferability. We show that the distributional responses of target taxa and target regions can be predicted better than random from their traits alone (AUROC median 0.65; AUPRC median 1.24 times random). But variation in predictive
performance was high—some target taxa and regions were better predicted than others. Higher performance at predicting distributions was related to well-calibrated environmental response predictions resulting from strong trait–environment associations. Contrary to expectations, predictive performance in target regions displayed no distance-decay from the reference region where the model was fitted.

**Transferability of predictions**

The predictive performance we documented is notable, because this was a challenging out-of-sample testing framework. Predictions were made to new target taxa, without distributional data or taxa identity, but only using the three traits for the target taxa, multiplied by the coefficients predicted from the trait-SDM. Also, we worked within one genus, which makes the problem harder, because it potentially limits the trait range.

Moreover, the model reference region was peripheral, environmentally and geographically, to the wider domain that we wished to predict to. This was partly historical—the reference region 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 the reference region is small, and we predicted to multiple target regions over a much greater extent. One expects better performance starting with a more extensive reference dataset to predict to a small target one. But that is not the problem that we believe presents itself to ecologists and practitioners, who are faced with large areas where comparatively little is known and some intensively studied reference areas, from which one may wish to transfer knowledge. This is the problem that we have attempted to address, with some encouraging signs. Analogously, predicting suitability under climate change involves predicting into conditions of great uncertainty from a smaller, well-understood reference situation (Porfírio et al. 2014).

Local and regional analyses have reflected the importance of SLA for species performance along gradients of temperature (Laughlin et al. 2012), moisture (Cornwell and Ackerly 2009) and fertility (Jager et al. 2015). Height is repeatedly shown to increase along gradients of moisture availability (Simpson et al. 2016, Pollock et al. 2018, Treurnicht et al. 2020). Seed mass has been recorded to increase (Treurnicht et al. 2020) and decrease (Pollock et al. 2012) with moisture availability. Those statements of
trait–environment associations are qualitative, spanning different statistical methodologies and different environmental indices. A systematic, predictive framework is needed to determine how general such trends are.

Surprisingly, we found no evidence to suggest declining transferability across geographic, compositional or environmental space in our study. Nor was performance in the reference region much higher than target regions, ruling out an initial steep decline then flattening of performance hiding a relationship between distance and performance. The random effects of target regions in models of performance were roughly half that of taxa. This means that these trait-SDMs can be reasonably transferred in space within the wider domain of our study, but that some taxa are harder to predict and some easier, and that effort to understand why would benefit transferability. We found that taxa that were better predicted tended to have lighter seeds. A similar but uncertain trend was found for high SLA species having better model performance. One speculative explanation draws on classical cost–benefit theory about community assembly along gradients of favourability (Orians and Solbrig 1977, Smith and 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). Likewise, higher SLA species have more resource–acquisitive strategies, and are more likely to be restricted to the most productive habitats and intolerant of ‘harsh’ conditions (Wright et al. 2004, Díaz et al. 2016, Bruelheide et al. 2018).

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. 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). 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, better reflecting the water costs of photosynthetic capacity (e.g. RuBisCO-dependent carboxylation capacity (Vcmax) or leaf nitrogen per area) (Prentice et al. 2014).

Additionally, total nitrogen responses were mainly negatively calibrated, potentially because the fit was accurate within the reference region, but that trait–environment relationship had no generality. Alternatively, the covariate may have been incorrectly selected, potentially being 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 highlights the importance of the transfer testing we demonstrate here. If we accepted the results from the reference region, we might be misled to believe that responses to total nitrogen (positive trend with SLA and negative trend with max height) applied in other regions.

Performance evaluation and metrics

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 species identity.

We used AUROC as a widely-used, convenient metric of model performance, that has been recommended in studies of transferability (Randin et al. 2006, Sequeira et al. 2018). Yet we suspect that AUROC is unrealistically optimistic about performance for the problem of an ecologist standing in a topographically-, climatically- and edaphically-complex landscape, trying to determine where to go to find a target species with known traits. AUROC is sensitive to true negatives, and if a species is relatively rare and a model correctly predicts that it is absent from much of the landscape, then the high values of AUROC can be achieved, regardless of capacity to predict true positives (Sofaer et al. 2019).

Performance measured by AUPRC appeared much more modest than AUROC, though the scales are not comparable. Still, performance that is for the majority of species at least 1.2 times as good as random is no mean feat when all the model knows is the three traits of the target taxa. How our trait-SDM performance compares for AUPRC is unclear, as AUPRC has been less used in SDM (Sofaer et al. 2019), and we know of no studies employing AUPRC for evaluating model transferability. Our experience suggests that AUPRC should be more widely used – it better reflected model calibration in plots and models of the relationship between model miscalibration and predictive performance. Interpretation of AUPRC is confounded by dependence on prevalence—higher prevalence determines higher AUPRC. But it matches the problem of directing survey effort, and its interpretation is aided by expressing relative to the performance of a random classifier (= prevalence). 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.
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 yr. 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 used modelled spatial environmental covariates. Our original modelling of the Gariwerd-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 microclimate (Austin and Van Niel 2011). And the indices based on digital elevation models that exist for such environmental variables do not approach the precision and accuracy that 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 and Austin 2007).

Inevitably the question occurs, are these the right traits? The random effect of species was relatively large, perhaps unmeasured traits can explain some of the residual variation in predictive performance and their incorporation into the trait-SDM might help. Our traits were chosen to reflect allocation trade-offs, and were easily measured (Westoby 1998). Traits like germination-response-to-temperature, would likely improve model performance, but would require such a high level knowledge of the species that predicting distribution and response to environmental gradients would be moot. We have limited information of recruitment and regeneration niche, beyond seed mass, which was expected to reflect a tolerance-fecondity trade-off, which was found, if we interpret rugged areas as harsh environments for seedling recruitment (low moisture, low nutrients, high pressure from seedling enemies) (Leishman 2001, Muller-Landau 2010). Stem sapwood density may be a useful trait to include in models like these, reflecting a growth–mortality trade-off (Wright et al. 2010) and, in particular tolerance to drought mortality risk (Anderegg et al. 2016). Yet, we reiterate that with 20 species in the training set, there is a strong limit to the number of traits which can be fitted. If we consider low and high values of each trait, three traits result in $2^{3} = 8$ independent combinations, with 2.5 species per combination in a perfectly distributed dataset. That is a small sample for reliable statistical inference. Further, our purpose here was not to find the best of all possible models, rather to evaluate whether a model could be transferred.

Our models include only environment, no biotic interactions. It is unclear how interactions could be introduced into predictions without the species identity being in the model (and thus in the reference or training dataset). This is similar to the problem in community-level models of ‘predict first, assemble later’ or ‘predict and assemble together’ (Ferrier and Guisan 2006). Conceivably, competitive interactions, which have been shown to be associated with leaf-economics and wood-density traits (Lasky et al. 2014, Kunstler et al. 2016) could be useful. Joint species distribution models may offer a route forward, the biotic interactions need to be captured in the linear model, rather than the residuals (Warton et al. 2015). Two problems then remain; the species information is missing when making predictions to new target taxa and regions, and unmeasured trait-mediated interactions are difficult to disentangle from trait–environment associations (Ovaskainen et al. 2017). Equally, other demography-based models also are constrained for making predictions about interactions between unknown species (Pagel et al. 2020, Treurnicht et al. 2020).

Our choice of response (presence–absence) is one that is widely used, though we see no reason to limit to occurrence. Abundance is a useful response, as would be growth. Species traits have been used successfully as predictors of forest tree mortality (Camac et al. 2018), and other vital rates (Adler et al. 2014, Visser et al. 2016) competition effects and responses (Kunstler et al. 2016), and changes through recruitment (Lai et al. 2020) and secondary succession (Lasky et al. 2014). Though in such cases, meso- and macro scale environment is generally ignored. It is worth asking whether those models transfer. Early work on transferring trait–height growth models has not proved highly successful (Thomas and Vesk 2017), yet interestingly a theory-driven trait-parameter subset produced the most generalizable models (Thomas et al. 2019).

Of course, the approach used here to test model transfer can be profitably applied to any of the community models utilising traits as covariate, as these can make predictions to new species or conditions (Ovaskainen et al. 2017). But this predictive framework for transferability testing highlights the need for a model-based approach to community ecology (Warton et al. 2015). A method based on multivariate distances cannot make predictions outside the training dataset. More mechanistic approaches (Higgins et al. 2012, Treurnicht et al. 2020) can also be evaluated in the way we illustrate here, as long as the traits and covariates are known for the target taxa and regions, and there is a hierarchical structure to the model that defines a trait by environment interaction. While our models ignore intraspecific trait variation, in theory it could be incorporated into these multi-species models, as they have been for single species (Benito Garzón et al. 2019), though in practice they would not be useful for the problem we considered, as you would need to know the trait values for all target species along the environmental gradients. If this were so, you would have no need of a model to predict where the species is distributed.

These results give confidence in the value of traits to assist in the hard problem of predicting responses to environmental gradients for new target species and new target environmental
conditions and regions. They deserve testing in different systems – other clades and landscapes. Predictions between similar temperate-climate regions should be tested. Factors likely contributing to the success here are that the reference region was environmentally diverse and the species there were functionally diverse, ranging widely in all three traits.

Data availability statement

Data available from Figshare Digital Repository: <https://doi.org/10.6084/m9.figshare.1026188/13128884> (Vesk et al. 2020).

Acknowledgements – We thank Angela Stock, Martin Stock and Daniel White for assistance in the field. Data come from lands of the following Indigenous Peoples: Djabwurrung, Jardwadjali, Wathurong, Wauworong, Boonwurrung, Taungurung, Kurnai, Bidwell, Waveroo, Wiradjuri, Jaimatung, Ngariyo, Yuin, Ngunawal, Gudungurra, Tharawal. We acknowledge them. The manuscript was improved by comments from two reviewers and the editor.

Funding – 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. Permits – Field work was performed under permits from NSW National Parks and Wildlife Service SL101959, and the Victorian Dept of Environment, Land, Water and Planning 10008557.

Author contributions

| Name                  | Contributions                                      |
|-----------------------|---------------------------------------------------|
| Peter Vesk            | Conceptualization (lead); Data curation (equal);  |
|                       | Formal analysis (equal); Funding acquisition (lead); |
|                       | Investigation (equal); Methodology (supporting);   |
|                       | Project administration (lead); Resources (lead);   |
|                       | Software (supporting); Supervision (lead); Validation (supporting); Visualization (supporting); Writing – review and editing (lead); Writing – original draft (lead); Writing – review and editing (lead). |
| William Morris        | Conceptualization (equal); Data curation (equal); |
|                       | Formal analysis (lead); Investigation (equal);     |
|                       | Methodology (equal); Project administration (supporting); Resources (supporting); Software (lead); Validation (lead); Visualization (lead); Writing – original draft (supporting); Writing – review and editing (supporting). |
| Will Neal             | Data curation (lead); Investigation (supporting);  |
|                       | Project administration (lead); Writing – review and editing (supporting). |
| Karel Mokany          | Data curation (equal); Methodology (supporting);   |
|                       | Resources (equal); Writing – review and editing (supporting). |
| Laura Pollock         | Data curation (supporting); Investigation (supporting); Methodology (supporting); Resources (lead); Visualization (supporting); Writing – review and editing (lead). |

References

Adler, P. B. et al. 2014. Functional traits explain variation in plant life history strategies. – Proc. Natl Acad. Sci. USA 111: 740–745.
Anderegg, W. R. L. et al. 2016. Meta-analysis reveals that hydraulic traits explain cross-species patterns of drought-induced tree mortality across the globe. – Proc. Natl Acad. Sci. USA 113: 5024–5029.
Austin, M. P. and Van Niel, K. P. 2011. Improving species distribution models for climate change studies: variable selection and scale. – J. Biogeogr. 38: 1–8.
Austin, M. P. et al. 1990. Measurement of the realized qualitative niche: environmental niches of five Eucalyptus species. – Ecol. Monogr. 60: 161–177.
Austin, M. P. et al. 1996. Patterns of tree species richness in relation to environment in southeastern New South Wales, Australia. – Aust. J. Ecol. 21: 154–164.
Benito Garzón, M. et al. 2019. ΔTrait SDMs: species distribution models that account for local adaptation and phenotypic plasticity. – New Phytol. 222: 1757–1765.
Briscoe, N. J. et al. 2016. Unpacking the mechanisms captured by a correlative species distribution model to improve predictions of climate refugia. – Global Change Biol. 22: 2425–2439.
Briscoe, N. J. et al. 2019. Forecasting species range dynamics with process-explicit models: matching methods to applications. – Ecol. Lett. 22: 1940–1956.
Brown, A. M. et al. 2014. The fourth-corner solution – using predictive models to understand how species traits interact with the environment. – Methods Ecol. Evol. 5: 344–352.
Bruijnzeel, H. et al. 2018. Global trait–environment relationships of plant communities. – Nat. Ecol. Evol. 2: 1906–1917.
Camac, J. S. et al. 2018. Partitioning mortality into growth-dependent and growth-independent hazards across 203 tropical tree species. – Proc. Natl Acad. Sci. USA 115: 12459–12464.
Colwell, R. K. and Rangel, T. F. 2009. Hutchinson’s duality: the once and future niche. – Proc. Natl Acad. Sci. USA 106: 19651–19658.
Cornwell, W. K. and Ackerly, D. D. 2009. Community assembly and shifts in plant trait distributions across an environmental gradient in coastal California. – Ecol. Monogr. 79: 109–126.
Cover, T. M. and Thomas, J. A. 2006. Elements of information theory, 2nd ed. – Wiley.
Díaz, S. et al. 2016. The global spectrum of plant form and function. – Nature 529: 167–171.
Dobrowski, S. Z. et al. 2011. Modeling plant ranges over 75 years of climate change in California, USA: temporal transferability and species traits. – Ecol. Monogr. 81: 241–257.
Dorrough, J. and Scroggie, M. P. 2008. Plant responses to agricultural intensification. – J. Appl. Ecol. 45: 1274–1283.
Dorrough, J. et al. 2011. Individual plant species responses to phosphorus and livestock grazing. – Aust. J. Bot. 59: 670–681.
Díaz, S. et al. 2016. The global spectrum of plant form and function. – Nature 529: 167–171.
Elith, J. and Leathwick, J. R. 2009. Species distribution models: ecological explanation and prediction across space and time. – Annu. Rev. Ecol. Evol. Syst. 40: 677–697.
Evans, J. P. et al. 2014. Design of a regional climate modelling projection ensemble experiment – NARClim. – Geosci. Model Dev. 7: 621–629.
Falster, D. S. and Westoby, M. 2003. Plant height and evolutionary games. – Trends Ecol. Evol. 18: 337–343.
Ferrier, S. and Guisan, A. 2006. Spatial modelling of biodiversity at the community level. – J. Appl. Ecol. 43: 393–404.
Fielding, A. H. and Bell, J. F. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. – Environ. Conserv. 24: 38–49.
Gelfand, A. E. et al. 2006. Explaining species distribution patterns through hierarchical modeling. – Bayesian Anal. 1: 41–92.
Grundy, M. J. et al. 2015. Soil and landscape grid of Australia. – Soil Res. 53: 835–844.
Higgins, S. I. et al. 2012. A physiological analogy of the niche for projecting the potential distribution of plants. – J. Biogeogr. 39: 2132–2145.
Hooten, M. B. and Hobbs, N. T. 2015. A guide to Bayesian model selection for ecologists. – Ecol. Monogr. 85: 3–28.
Jager, M. M. et al. 2015. Soil fertility induces coordinated responses of multiple independent functional traits. – J. Ecol. 103: 374–385.
Jamil, T. et al. 2013. Selecting traits that explain species–environment relationships: a generalized linear mixed model approach. – J. Veg. Sci. 24: 988–1000.
Kunstler, G. et al. 2016. Plant functional traits have globally consistent effects on competition. – Nature 529: 204–207.
Lai, H. R. et al. 2020. Functional traits that moderate tropical tree recruitment during post-windstorm secondary succession. – J. Ecol. 108: 1322–1333.
Lasky, J. R. et al. 2014. Trait-mediated assembly processes predict successional changes in community diversity of tropical forests. – Proc. Natl Acad. Sci. USA 111: 5616–5621.
Laughlin, D. C. et al. 2012. A predictive model of community assembly that incorporates intraspecific trait variation. – Ecol. Lett. 15: 1291–1299.
Legendre, P. and Legendre, L. F. J. 2012. Numerical ecology, 3rd ed. Developments in environmental modelling 24. – 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. et al. 2019. Functional traits and community composition: a comparison among community-weighted means, weighted correlations and multilevel models. – Methods Ecol. Evol. 10: 415–425.
Moles, A. T. et al. 2007. Global patterns in seed size. – Global Ecol. Biogeogr. 16: 109–116.
Moles, A. T. et al. 2009. Global patterns in plant height. – J. Ecol. 97: 923–932.
Morán-Ordóñez, A. et al. 2017. Evaluating 318 continental-scale species distribution models over a 60-year prediction horizon: what factors influence the reliability of predictions? – Global Ecol. Biogeogr. 26: 371–384.
Morin, X. et al. 2008. Tree species range shifts at a continental scale: new predictive insights from a process-based model. – J. Ecol. 96: 784–794.
Muller-Landau, H. C. 2010. The tolerance-fecundity trade-off and the maintenance of diversity in seed size. – Proc. Natl Acad. Sci. USA 107: 4242–4247.
Normand, S. et al. 2009. Importance of abiotic stress as a range-limit determinant for European plants: insights from species responses to climatic gradients. – Global Ecol. Biogeogr. 18: 437–449.
Orians, G. H. and Solbrig, O. T. 1977. A cost-income model of leaves and roots with special reference to arid and semiarid areas. – Am. Nat. 111: 677–690.
Ovaskainen, O. and Soininen, J. 2011. Making more out of sparse data: hierarchical modeling of species communities. – Ecology 92: 289–295.
Ovaskainen, O. et al. 2017. How to make more out of community data? A conceptual framework and its implementation as models and software. – Ecol. Lett. 20: 561–576.
Pagel, J. et al. 2020. Mismatches between demographic niches and geographic distributions are strongest in poorly dispersed and highly persistent plant species. – Proc. Natl Acad. Sci. USA 117: 3663–3669.
Perez-Harguindeguy, N et al. 2016. Corrigendum to: a handbook for standardised measurement of plant functional traits worldwide. – Aust. J. Bot. 64: 715–716.
Pollock, L. J. et al. 2012. The role of functional traits in species distributions revealed through a hierarchical model. – Ecol. Lett. 35: 716–725.
Pollock, L. J. et al. 2015. The roles of ecological and evolutionary processes in plant community assembly: the environment, hybridization and introgression influence co-occurrence of Eucalyptus. – Am. Nat. 185: 784–796.
Pollock, L. J. et al. 2018. Combining functional traits, the environment and multiple surveys to understand semi-arid tree distributions. – J. Veg. Sci. 29: 967–977.
Porfirio, L. L. et al. 2014. Improving the use of species distribution models in conservation planning and management under climate change. – PLoS One 9: e113749.
Prentice, I. C. et al. 2014. Balancing the costs of carbon gain and water transport: testing a new theoretical framework for plant functional ecology. – Ecol. Lett. 17: 82–91.
Randin, C. F. et al. 2006. Are niche-based species distribution models transferable in space? – J. Biogeogr. 33: 1689–1703.
Reich, P. B. et al. 2007. Predicting leaf physiology from simple plant and climate attributes: a global GLOPNET analysis. – Ecol. Appl. 17: 1982–1988.
Sequeira, A. M. M. et al. 2018. Transferring biodiversity models for conservation: opportunities and challenges. – Methods Ecol. Evol. 9: 1250–1264.
Simpson, A. H. et al. 2016. Soil–climate interactions explain variation in foliar, stem, root and reproductive traits across temperate forests. – Global Ecol. Biogeogr. 25: 964–978.
Slee, A. V. et al. 2006. Euclivid: eucalypts of Australia, 3rd ed. – CSIRO.
Smith, T. and Huston, M. 1989. A theory of the spatial and temporal dynamics of plant communities. – Vegetatio 83: 49–69.
Sofaer, H. R. et al. 2019. The area under the precision-recall curve as a performance metric for rare binary events. – Methods Ecol. Evol. 10: 565–577.
The State of Victoria, Dept of Environment, Land, Water and Planning 2018. Victorian biodiversity atlas. – The State of Victoria, Dept of Environment, Land, Water and Planning.
Thomas, F. M. and Vesk, P. A. 2017. Are trait–growth models transferable? Predicting multi-species growth trajectories between ecosystems using plant functional traits. – PLoS One 12: e0176959.
Thomas, F. M. et al. 2019. Using functional traits to predict species growth trajectories, and cross-validation to evaluate these models for ecological prediction. – Ecol. Evol. 9: 1554–1566.
Teuwenicht, M. et al. 2016. Environmental drivers of demographic variation across the global geographical range of 26 plant species. – J. Ecol. 104: 331–342.
Teuwenicht, M. et al. 2020. Functional traits explain the Hutchinsonian niches of plant species. – Global Ecol. Biogeogr. 29: 534–545.
Van Niel, K. P. and Austin, M. P. 2007. Predictive vegetation modeling for conservation: impact of error propagation from digital elevation data. – Ecol. Appl. 17: 266–280.
Vesk, P. A. 2013. How traits determine species responses to environmental gradients. – J. Veg. Sci. 24: 977–978.
Vesk, P. et al. 2020. Data from: Transferability of trait-based species distribution models. – Figshare Digital Repository, <https://dx.doi.org/10.6084/m9.figshare.10.26188/13128884>.

Visser, M. D. et al. 2016. Functional traits as predictors of vital rates across the life cycle of tropical trees. – Funct. Ecol. 30: 168–180.

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

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

Westoby, M. and Wright, I. J. 2006. Land–plant ecology on the basis of functional traits. – Trends Ecol. Evol. 21: 261–268.

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

Wright, S. J. et al. 2010. Functional traits and the growth-mortality trade-off in tropical trees. – Ecology 91: 3664–3674.
Author/s:
Vesk, PA; Morris, WK; Neal, WC; Mokany, K; Pollock, LJ

Title:
Transferability of trait-based species distribution models

Date:
2020-11-10

Citation:
Vesk, P. A., Morris, W. K., Neal, W. C., Mokany, K. & Pollock, L. J. (2020). Transferability of trait-based species distribution models. ECOGRAPHY, 44 (1), pp.134-147. https://doi.org/10.1111/ecog.05179.

Persistent Link:
http://hdl.handle.net/11343/274116

File Description:
Published version

License:
CC BY