Poor transferability of richness models: Predicting plot-scale plant diversity in the Waterberg, South Africa

Tamryn S. Venter*, Nigel P. Barker, Peter C. le Roux

Department of Plant and Soil Sciences, University of Pretoria, Private Bag X20, Hatfield, Pretoria 0028, South Africa

A B S T R A C T

Modelling measures of biodiversity for understudied taxa or regions is one method to address taxonomic and geographic biases in biodiversity data. However, modelling biodiversity metrics, such as species richness, to unsampled areas is only useful if predictions are reliable. As a result, testing the transferability of richness models is necessary for assessing the potential for models to predict to unsampled areas. Here we test the transferability of plant richness models between two reserves to examine if the richness-environment relationship from one reserve can accurately estimate richness patterns in the other reserve, using the vascular plant species richness of the Waterberg region (savanna biome; northern South Africa) as a model system. Six richness response variables (total species, grass species, herb species, woody species, genus, and family richness) and a set of 16 predictor variables were analysed with three modelling approaches and two statistical techniques to build: (1) models comprising all available predictor variables, (2) models using a subset of predictor variables chosen based on model performance, and (3) models using a subset of predictor variables that reduce the difference in the environmental conditions between the two reserves. The models' performance in the training area varied considerably, but soil variables were consistently the most important predictors of plant richness. However, the transferability of all the models was consistently poor across all modelling approaches and both techniques, possibly reflecting the degree to which each reserve contains novel environments absent from the other reserve (despite being separated by only c. 60 km and sharing vegetation types). Due to the poor performance of these richness models, they are currently not useful for predicting richness to other areas in the vicinity of the reserves or in the broader region. However, in areas, like the Waterberg region, that have high plant diversity and are poorly sampled, there may be value in continued development of richness models to address biodiversity gaps, thereby providing better data to inform conservation decisions.

© 2024 The Authors. Published by Elsevier B.V. on behalf of SAAB. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

1. Introduction

The current knowledge of biodiversity, including species distributions and broad-scale patterns of species richness, is vulnerable to taxonomic and geographic biases (Rocchini et al., 2011). These biases can limit our understanding of the factors affecting the distribution of biodiversity and can be problematic for effective conservation planning (Ferrier, 2002; Hortal and Lobo, 2006). Modelling of biodiversity data for understudied taxa or regions is one method to address these biases, particularly since this approach is more time- and cost-efficient than extensive field-based data collection (Ferrier, 2002; Parmentier et al., 2011; Sequeira et al., 2018, 2016). Statistically estimating biodiversity for poorly studied groups or areas is also increasingly becoming more feasible for a larger range of taxa and

portions of the world due to the greater availability of datasets online (Guralnick et al., 2007; Soberón and Peterson, 2004).

However, predicting biodiversity to unsampled areas or times is only useful if the reliability of the results is understood. Transferability is an attribute demonstrated by a model when it can be used to accurately predict the response variable in unsampled areas or times (Phillips, 2008). Explicitly testing the transferability of biodiversity models is thus necessary to determine the reliability of a model's predictions to areas or times beyond the dataset which was used to calibrate the model (Randin et al., 2006). Model transferability, for instance across geographic space, can be assessed by using the model (calibrated in the training area) to predict values in a new area (the testing area) and comparing them to values actually observed in the testing area (Fig. A1 in Appendix A) (Randin et al., 2006; Wenger and Olden, 2012). In addition, testing transferability has considerable practical value as an additional method of model evaluation, even

when predicting to unsampled areas is not the main objective of the study (Randin et al., 2006).

Model transferability varies between studies, with good transferability (Lauria et al., 2015; Peterson et al., 2003; Sequeira et al., 2016; Sundblad et al., 2009), poor transferability (Capinha et al., 2018; Hortal and Lobo, 2011; Randin et al., 2006; Wenger and Olden, 2012), and both good and poor transferability (Duncan et al., 2009; Redfern et al., 2017) reported in the literature. Several factors may affect model transferability, including the number of predictor variables (Low et al., 2020; Randin et al., 2006). Models using many variables may be overfitted (i.e. the model is calibrated very closely to the training data and lacks generality), whereas less complex models with fewer variables can be underfitted and, therefore, also provide inaccurate predictions (Warren et al., 2014). The effects that the number of predictor variables have on model complexity are also dependent on the modelling approach and data used and, therefore, requires more investigation (Sequeira et al., 2018).

In addition, model transferability is usually negatively related to the spatial extent of a study (Parmentier et al., 2011; Rousseau and Betts, 2022), with greater distances between the areas typically resulting in more substantial differences in environmental conditions (Rousseau and Betts, 2022). Novel environments, which refer to environmental conditions that are not present in the training area but occur in the testing area, typically reduce model transferability (Hortal and Lobo, 2011; Low et al., 2020; Parmentier et al., 2011; Randin et al., 2006). Novel environments are problematic as the model has not been calibrated outside of the initial environmental range and while the model may be able to accurately predict the response variable in the training area, it may not correctly predict beyond those initial environmental ranges in the testing area (see Fig. A2 in Appendix A) (Hortal and Lobo, 2011; Randin et al., 2006; Rousseau and Betts, 2022).

When richness models are transferable, missing biodiversity data can be estimated, providing information that can be potentially valuable for conservation planning in poorly sampled areas with little to no information on the occurrence or distribution of species (Hortal and Lobo, 2011). Therefore, in the absence of accurate species occurrences (to model species distributions), species richness is the next best approximation of biodiversity for conservation planning exercises (Hortal and Lobo, 2011). As a result, it is important to accurately model species richness, and one way to potentially increase the robustness of species richness models, especially for plant species, is to build separate richness models for different growth forms, since growth forms can be expected to respond differently to some environmental conditions (Díaz and Cabido, 2001 and references within). Indeed, predictions of the richness of different growth forms can provide habitat-specific information, which is useful for conservation planning (Díaz and Cabido, 2001; McNellie et al., 2021). Furthermore, testing the transferability of species richness models per growth form, assuming the models perform well, will identify the environmental correlates of each growth forms' species richness and can highlight the distribution of different growth forms as opposed to just total species richness for different areas.

This study tested which environmental factors are most strongly correlated with plant richness (total species richness and the richness of different growth forms and at different taxonomic levels) within two game reserves in the Waterberg Biosphere Reserve (WBR) and tested the transferability of these richness-environment relationships between the two game reserves. To identify which variables correlated with the spatial variation in richness within the WBR, richness data and environmental data were collected from plots within the two reserves. Two statistical techniques and three modelling approaches were used to explicitly account for the influence of the number of predictor variables and the potential effects of novel environments. If the richness models developed here display high transferability, then they could be used to predict plant richness in other areas within the WBR and provide insight into areas that should be targeted for conservation planning and/or potential vegetation surveys in the future.

2. Methods

2.1. Study areas

Welgevonden Game Reserve (hereafter 'Welgevonden') and Lapalala Nature Reserve (hereafter 'Lapalala') are both privately owned game reserves in the WBR of northern South Africa, within the Savanna biome (Fig. 1). The WBR was designated in 2001 by UNESCO and aims to ensure sustainable socio-ecological land management of the area (Pool-Stanvliet, 2013). Welgevonden and Lapalala are c. 60 km apart and both have a mountainous topography. Welgevonden, approximately 35,000 ha in size, is located in the south-east of the WBR and, prior to becoming a game reserve, parts of the land were used for farming (both crops and livestock; Slater and Long, 2012). Lapalala, approximately 48,000 ha in size, is located northeast of Welgevonden, and parts of the land were also previously used for farming (Womack, 2022). Both reserves have a wide variety of animals including megaherbivores. In addition to both reserves being located within the Savanna biome, they are also located within the same vegetation type (Waterberg Mountain Bushveld; one of 87 vegetation types within the South Africa's Savanna biome) and their underlying geologies are both mainly sandstone of the Kransberg Subgroup of the Waterberg Group (Fig. B1 in Appendix B) (Mucina and Rutherford, 2006). The resulting soil is nutrient poor with high acidity and plants have evolved and adapted to survive in these soil conditions (Flood, 2015). The reserves, however, differ in their elevational range: Welgevonden's elevation ranges between 1100 and 1700 m above sea level (with an average of c. 1300 m a.s.l.), whereas Lapalala's elevation ranges between 900 and 1200 m a.s.l. (with an average of c. 1100 m a.s.l.) (Womack, 2022). The mean annual daily maximum and minimum temperatures is 28.5 °C and 13 °C for Lapalala (Womack, 2022) and 26.5 °C and 11 °C for Welgevonden (Codron, 2004), respectively. The mean annual rainfall for Lapalala is 546 mm (Womack, 2022), and for Welgevonden is 642 mm (Jonathan Swart, Welgevonden Ecologist, personal communication, 2021), although large inter-annual variation in rainfall occurs in this region.

2.2. Data collection

Two datasets were analysed for this study, both of which were collected using equivalent field methods. Each dataset comprises quantitative plot-based data, composed of a complete vascular plant survey and associated environmental data. The Lapalala dataset was extracted from Womack (2022), and was collected from January to March 2019, while the Welgevonden dataset was collected from February to March 2021. The location of each plot (a demarcated 20 × 20 m square) was chosen prior to fieldwork following a stratified sampling approach (with Lapalala stratified by elevation and soil type, and Welgevonden stratified by vegetation type).

Total species richness, genus richness (number of different genera) and family richness (number of different families) was calculated per plot. In addition, grass species richness (i.e. Poaceae species only), woody species richness (i.e. all woody tree and shrub species), and herb species richness (i.e. all herbaceous, fern, succulent and geophyte species) was determined for each plot, resulting in the six response variables examined in this study: total species, genus, family, grass species, woody species, and herb species richness.

Environmental data were also recorded at each plot. The cover of rock, woody debris, bare soil, and leaf litter were visually estimated for each plot. Soil from the four corners of each plot was collected (roughly 1 kg of soil in total per plot). These samples were air dried
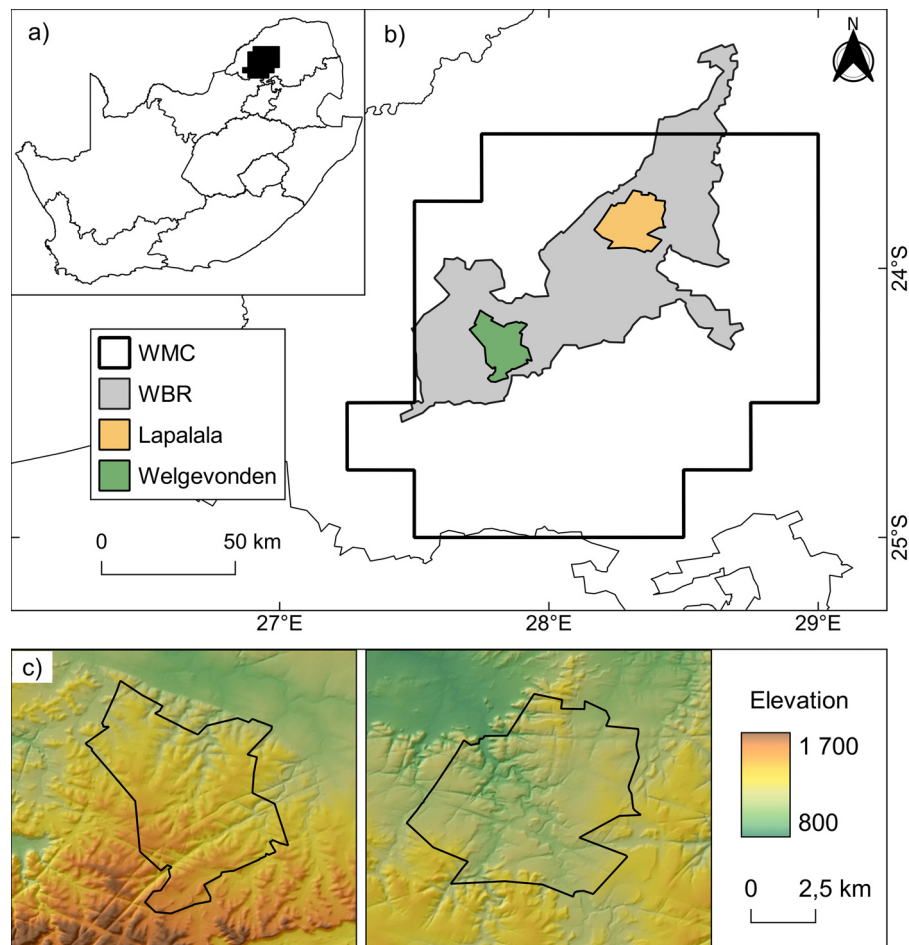
Fig. 1. a) The location of the Waterberg Mountain Complex (WMC) within South Africa. b) The location of the two reserves, Lapalala and Welgevonden, within the WMC and the Waterberg Biosphere Reserve (WBR). c) Elevation maps of Lapalala (right) and Welgevonden (left).

and sieved through a 2 mm sieve. At the Soil Laboratory of the University of Pretoria, particle size distribution was determined using the hydrometer method (following, Kroetsch and Wang, 2008), soil pH was determined using the H_2O method (following, Hendershot et al., 2008), the Mehlich 3 extraction method was used to determine the concentration of potassium, calcium, magnesium, sodium, and phosphorus in each soil sample (following, Ziadi and sen Tran, 2008), and the dichromate redox titration method was used to determine the percentage of organic carbon in the soil sample (following, Skjemstad and Baldock, 2008).

Finally, an ASTER Global Digital Elevation Model (Version 3; downloaded from 10.5067/ASTER/ASTGTM.003) was obtained for both reserves and was used to extract aspect, slope, curvature, elevation, and topographic wetness index (TWI), for each plot. This was computed using QGIS software (QGIS, 2021) with the 'Slope, aspect, curvature', 'Catchment area', and 'Topographic wetness index' functions. The 'Point sampling tool' function was then used to extract the value for each variable at each plot. Potential direct incident radiation (PDIR) was calculated following McCune (2007).

2.3. Data analysis

Environmental and vascular plant species richness data were collected from a total of 180 plots for Lapalala and 65 plots for Welgevonden (Table 1) and used to model plot-level richness. Collinearity between predictor variables was tested using the Pearson's correlation coefficient, with predictors excluded until all pairwise correlations were $<|0.75|$ (see Fig. C1 in Appendix C for the sequential data

processing and analysis steps). Collinearity can be particularly problematic for model transferability (Dormann et al., 2013; Feng et al., 2019) and, as a result, three highly correlated variables (magnesium, sand fraction and silt fraction) were removed. This resulted in a final dataset of 16 predictor variables.

A diagnostic plot was then constructed for each predictor variable against each response variable to identify outlier and/or bad leverage values in each dataset (following: Rousseeuw, 1991; Rousseeuw and van Zomeren, 1990). Depending on the response variable, between three and 10 plots were removed from the Lapalala dataset, and three to eight plots removed from the Welgevonden dataset representing strong outliers or bad leverage values.

After excluding collinear predictors and removing outliers and bad leverage points, univariate linear and quadratic models were run using generalized linear models (GLMs) for each predictor variable-response variable combination, separately for both the Lapalala and Welgevonden datasets, to determine if any predictor variables displayed a quadratic relationship with response variables. Quadratic models were considered to perform better than linear models if (1) the Akaike Information Criterion (AIC) value of the quadratic model was lower than the AIC value of the linear model, and (2) the quadratic model explained at least 5 % more deviance than the linear model. When a quadratic model performed better than the linear model, a quadratic term was included in the modelling approaches for that predictor variable against the respective response variable in subsequent GLMs. Linear models performed better than quadratic models for the majority of the predictor variables in both datasets (Table D1 in Appendix D).

Table 1

Summary of the response and predictor variables used in this study.

Variable	Description and units
Richness data (6 response variables)	
Total species richness	Total number of species per plot
Grass species richness	Number of grass species per plot
Herb species richness	Number of herb species per plot
Woody species richness	Number of woody species per plot
Family richness	Family richness per plot
Genus richness	Genus richness per plot
Environmental data (19 predictor variables)	
Elevation	In metres above sea level (m a.s.l.)
Slope	Ranging from 0° (horizontal) to 90° (vertical)
TWI	Topographic wetness index; used as a proxy for soil moisture, where the higher the value is the wetter the area is predicted to be (based on topography)
Curvature	A value ranging from −1 (concave) to +1 (convex), with 0 being flat
PDIR	Potential annual direct incident radiation (MJ cm^{-2} yr^{-1})
Soil pH	The pH value of the soil
Carbon	Percentage carbon by weight of the soil
Potassium	Concentration of potassium in the soil sample (mg/kg soil)
Calcium	Concentration of calcium in the soil sample (mg/kg soil)
Magnesium	Concentration of magnesium in the soil sample (mg/kg soil)
Sodium	Concentration of sodium in the soil sample (mg/kg soil)
Phosphorus	Concentration of phosphorus in the soil sample (mg/kg soil)
Sand	Percentage of the soil sample comprised of sand (%weight/weight)
Clay	Percentage of the soil sample comprised of clay (%weight/weight)
Silt	Percentage of the soil sample comprised of silt (%weight/weight)
Rock cover	Percentage rock cover of each plot
Bare soil cover	Percentage bare soil cover of each plot
Woody debris cover	Percentage woody debris cover of each plot
Leaf litter cover	Percentage leaf litter cover of each plot

Table 2

The minimum and maximum values and the number of novel environments for each of the predictor variables for both the Lapalala (L) and Welgevonden (W) datasets. Lapalala had a total of 180 plots sampled and Welgevonden 65 plots.

Predictor variables	Minimum		Maximum		% Novel Environments	
	L	W	L	W	L	W
Elevation	921	1134	1292	1661	37	62
Slope	0.00	0.97	19.69	30.12	4	2
Curvature	−0.03	−0.03	0.01	0.03	1	17
TWI	0.20	0.40	14.37	15.70	1	2
PDIR	0.16	0.81	1.18	1.17	8	0
Rock cover	0	0	90	85	1	0
Woody debris cover	0	0	40	30	1	0
Bare soil cover	0	1	80	40	19	0
Leaf litter cover	0	1	80	50	11	0
Soil pH	4.02	4.41	6.91	6.44	8	0
Clay	6.12	11.00	24.49	33.00	51	2
Carbon	0.01	0.32	4.47	4.94	19	2
Potassium	2.08	22.54	51.15	153.69	96	82
Calcium	12.78	23.70	556.00	1251.02	25	8
Sodium	1.75	6.53	4.97	22.98	100	100
Phosphorus	0.15	4.43	7.19	36.23	97	80

As a final data processing step, corresponding predictor variables in the two datasets were compared to determine the number of plots with novel environments (i.e. how frequently observations in one dataset fell outside of the range of the same predictor variable from the other dataset). For example, if a plot in the Welgevonden dataset has an elevation value either above the maximum or below the minimum elevation value in the Lapalala dataset then the plot would be considered a novel environment for the elevation variable. Novel environments are likely to negatively affect a model's transferability (Low et al., 2020; Randin et al., 2006). Therefore, predictor variables having 50 % or more plots' values as novel environments were excluded from one of the modelling approaches (see below for details) to test if it affected model transferability. The average percentage of novel environments per predictor variable were slightly higher for the Lapalala dataset (30 %) compared to the Welgevonden dataset (22 %) (Table 2; see Table D2 in Appendix D for more detail on the ranges of the predictor variables).

2.4. Model building

Generalised linear models and generalized boosted models (GBMs) were used to model richness values as a function of environmental variables, with models repeated for each reserve, using a Poisson distribution (Elith et al., 2008; McCullagh and Nelder, 1989). GLMs have been shown to exhibit stronger transferability in space and time compared to other models (Hortal and Lobo, 2011; Randin et al., 2006), and their outputs allow simple comparisons of coefficients (i.e. allowing co-efficients from models trained on two different datasets to be compared). GBMs (also known as boosted

regression trees) provide a fundamentally different modelling approach compared to GLMs, using an ensemble approach to achieve the flexibility to model complex non-linear relationships and statistical interactions (Elith et al., 2008). Therefore, GBMs were used as a second modelling method to test for transferability to account for differences between statistical techniques. For all GBMs interaction depth was set to 3 (as recommended by Elith et al., 2008 for datasets of this size), and learning rate was sequentially reduced until > 1000 trees were produced (following Elith et al., 2008).

Three model building approaches were used in this study: (1) Full Models, (2) Minimum Adequate Models, and (3) Shared Environments Models. Full Models included all the predictor variables (and the relevant quadratic terms for GLMs) for each of the response variables. The Minimum Adequate Models were identified for GLMs using a best subsets model building approach based on minimizing AIC scores, and for GBMs using a backwards stepwise approach run within a 10-fold cross-validation approach (implemented via the gbm.simplify function; Elith et al., 2008). The Shared Environments Model also selected variables to minimize AIC values (GLMs) or via a backwards stepwise approach (GBMs) but had the additional constraint of being limited to predictor variables which had < 50 % of plots as novel environments. Basing model selection on AIC scores penalises complex models and can thus minimise overfitting which can reduce model transferability for GLMs (Randin et al., 2006; Wenger and Olden, 2012), while stepwise model building in GBMs eliminates non-formative variables (Elith et al., 2008) and can, therefore, also be assumed to reduce the potential for overfitting. The Full Models, Minimum Adequate Models and Shared Environments Models were built for each of the six response variables for each dataset, for both modelling techniques, GLMs and GBMs.

2.5. Transferability

All 72 models (36 GLMs and 36 GBMs) were subsequently used to predict richness in the other reserve, whereby the models built with the Lapalala dataset (the training dataset) were used with the Welgevonden dataset (the testing dataset) to predict species richness in Welgevonden, and vice versa. In order to further reduce the effect that novel environments have on the transferability of the models, the datasets were clamped per predictor variable, as clamping has been shown to improve the overall accuracy of models when extrapolating (Capinha et al., 2018). Clamping is when any value from the one dataset falling above the maximum or below the minimum values of the other dataset is 'clamped' to the maximum or minimum

Table 3
A summary of the GLMs transferability results that have been averaged across the three modelling approaches (Full Model, Minimum Adequate Model, and Shared Environments Model). The training dataset (L: Lapalala, W: Welgevonden) was used to build the model with which the testing dataset was then used to predict richness. The adjusted deviance explained (adj. DE) values are for these models and the R^2 values are for the linear models that were used to test transferability. See Table D4 in Appendix D for details on each individual model.

Richness variable	Training dataset	Average adj. DE (%)	Testing dataset	Average unclamped slope estimate	Average unclamped p-value	Average unclamped R^2	Average clamped slope estimate	Average clamped p-value	Average clamped R^2
Total species	L	22.7	W	−0.012	0.701	0.003	0.030	0.505	0.008
	W	43.2	L	−0.069	0.133	0.014	−0.072	0.099	0.016
Grass species	L	11.5	W	−0.122	0.660	0.006	0.086	0.142	0.046
	W	40.1	L	−0.155	0.064	0.029	−0.152	0.039	0.034
Herb species	L	9.0	W	−0.031	0.282	0.021	−0.032	0.288	0.022
	W	39.4	L	−0.007	0.600	0.002	0.010	0.674	0.001
Woody species	L	46.1	W	0.081	0.494	0.118	0.155	0.009	0.202
	W	61.1	L	0.084	0.095	0.031	0.081	0.103	0.028
Family	L	32.3	W	−0.111	0.010	0.124	0.199	0.005	0.151
	W	39.8	L	0.131	0.047	0.093	0.124	0.042	0.084
Genus	L	21.3	W	−0.104	0.213	0.028	−0.036	0.475	0.009
	W	46.7	L	−0.088	0.089	0.021	−0.115	0.022	0.036

value, respectively. Therefore, the models were tested for transferability twice, first with the raw (i.e. unclamped) dataset as the testing dataset and second with the clamped dataset as the testing dataset. A linear model was then used to compare predicted and observed richness values, as a measure of the transferability of the models. The analyses were conducted using R software (R Core Team, 2021), with GBMs built using functions from the dismo (Hijmans et al., 2022; Elith et al., 2008) package.

3. Results

Transferability of richness models was consistently poor, regardless of how well the models performed in the area they were calibrated in and irrespective of the modelling approach used for both the GLMs and GBMs. Due to the strong similarities in the results from the GLMs and the GBMs, just GLM results are reported below, with GBM results reported in detail in Appendix E.

3.1. Model building

The adjusted deviance explained (adj. DE) for the GLMs for the reserves they were trained in ranged from 4.1 % to 64.4 % (Table 3 and Table D4 in Appendix D). The richness variable for which the GLMs performed the best were the models trained with the Welgevonden dataset for woody richness (mean adj. DE = 61.1 % averaged across the three modelling approaches; Table 3). In contrast, the GLMs which performed the worst were for herb and grass richness in the Lapalala dataset (mean adj. DE = 9.0 % and 11.5 %, respectively; Table 3).

For both the Lapalala and Welgevonden datasets, the Minimum Adequate Model approach performed best for predicting species richness in the reserve in which the model was trained for all the models, except one (for GLMs). For this one model (family richness, in the Welgevonden dataset) the Shared Environments Model had the highest adj. DE value, but also a slightly higher (<1 unit different) AIC value compared to the Minimum Adequate Model. Shared Environments Models outperformed Full Models approximately 60 % of the time, to be the second best modelling approach (Table D4 in Appendix D).

The predictor variables included in the greatest number of GLMs overall were soil pH (33 out of 36 models) and soil carbon (29 models) and were, thus, consistently important for both the Lapalala and Welgevonden datasets. In addition, rock cover was important when predicting richness in Lapalala (occurring in 16 out of 18 Lapalala models) and curvature, TWI, woody debris cover, and calcium were important in Welgevonden (occurring in 15, 15, 16, and 14 out of 18 Welgevonden models, respectively). The majority of the predictor variables (c. 60 %) exhibited the same relationship with the response variables for >60 % of the models where it was included in both datasets (Table D3 in Appendix D). Elevation, curvature, PDIR, rock cover, leaf litter cover, soil pH, calcium, and sodium exhibited positive relationships for majority of the models across the two datasets; and TWI and bare soil cover exhibited negative relationships for majority of the GLMs.

3.2. Transferability

All the models displayed poor transferability, using both the unclamped and clamped datasets to predict richness (Fig. 2 and Table 3; see Table D4 in Appendix D for more details, and Appendix E for GBM results). The majority of the GLMs (26 using the unclamped testing dataset and 22 using the clamped testing dataset) did not perform better than chance (i.e. $p > 0.05$). None of the slope estimate values from the linear models were close to 1 (the highest value being 0.3), and 57 % of the model outputs had negative slope values (of which 22 % were significant) (Fig. 2). The majority of the GBMs performed better than random ($p < 0.05$ for 49 out of 72 GBMs for the relationship between predicted and observed richness), all of which had positive relationships. Only four (out of 72) GBMs had negative relationships (all of which did not perform better than random). However, consistent with the GLMs, slope estimate values were not close to one (the highest value being 0.4) and the R^2 values were low (the highest being 0.44; details in Appendix E).

The measures of richness which displayed the best transferability from the GLMs, albeit weak, were woody species richness and family richness, as they showed more positive than negative slope values from their models. For woody species richness 83 % of outputs had a positive slope estimate, of which 80 % were significant, while all the outputs with a negative slope estimate were not significant.

Clamping the testing dataset improved transferability (i.e. displayed a slope estimate closer to 1) for 23 out of 36 GLMs. For 16 of those 23 models, the clamped dataset models also displayed a higher R^2 value. However, of those 16 models, the predictions from only 7 models were significantly better than random (p-value < 0.05).

In contrast to modelling richness in training datasets, of the three modelling approaches, Shared Environments Models displayed the best, albeit weak, transferability for majority of the richness response variables. These models performed best (i.e. had a slope estimate
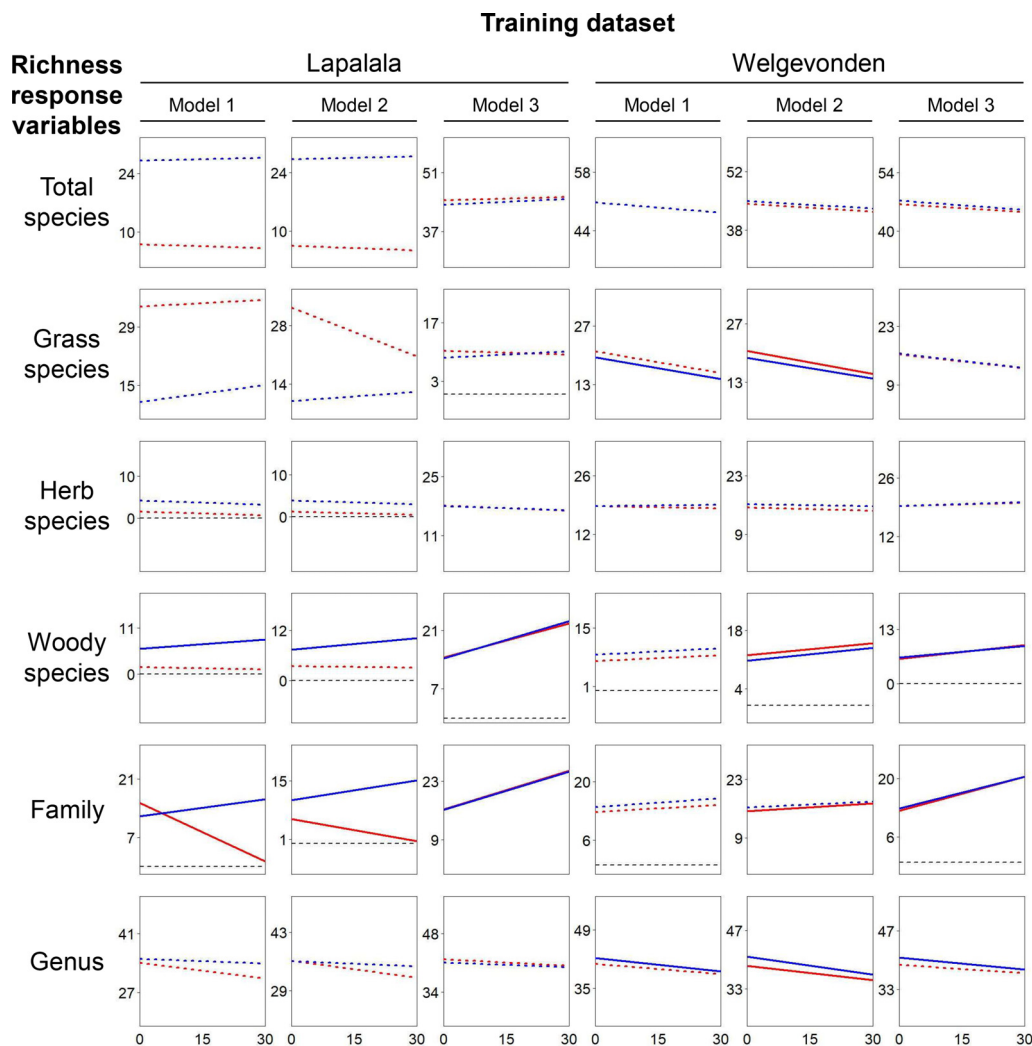
Training dataset

Fig. 2. Comparing the predicted and observed richness values for the three different models (1: Full Model, 2: Minimum Adequate Model, 3: Shared Environments Model). The training dataset is the dataset used to build the models. The red lines are the unclamped dataset's predicted richness against the observed richness, and the blue lines are the clamped dataset's predicted richness against observed richness (solid line: the output was statistically significant [< 0.05], dashed line: the output was not statistically significant). The black dashed horizontal line is at $y = 0$. All the y-axes and x-axes have the same range (30 units).

closer to 1) for nine of the GLMs (four for Lapalala and five for Welgevonden) when using the unclamped testing dataset and for nine GLMs (five for Lapalala and four for Welgevonden) when using the clamped testing datasets (Table 4).

4. Discussion

The transferability of all the richness models were consistently poor, despite the close proximity of Welgevonden and Lapalala and their similarity in vegetation types and geology. The performance of GLMs and GBMs in training sites varied considerably (GLM: maximum adj. DE = 64 %; GBM: maximum Pearson correlation [10-fold cross validation] = 0.70). The Minimum Adequate Model approach performed best, likely as this technique chose models with the lowest AIC, thereby balancing minimizing overfitting and maximizing explanatory power (Lauria et al., 2015; Randin et al., 2006). Shared Environments Models on average outperformed Full Models, highlighting that models which include all available predictor variables may not produce the best results (in agreement with Feng et al., 2019; Warren et al., 2014). Therefore, model building techniques are useful for determining the optimal set of predictor variables to account for variation in the response variable which, in turn, can reduce model complexity and improve transferability.

When comparing growth forms, models for woody species richness on average performed best, while the models for herb and grass species richness on average performed considerably worse. This discrepancy could be due to how species were delimited into these three groups. The woody species richness variable contained all the shrub and tree species, and therefore contained only species with approximately similar growth forms (albeit across a range of genera and families). In contrast, the grass species richness variable contained just the species within Poaceae, but excluded some species with similar growth forms (e.g. Cyperaceae species). Herb species richness contained all other species, such as herbaceous, fern, succulent and geophyte species (i.e. a variety of growth forms). The herb species richness thus, could be especially difficult to predict because two sites which have the same number of total herb species but contain different growth forms (e.g. one site has several geophyte species and no succulent species, whereas the other site has several succulent species but no geophyte species) could experience very different environmental conditions. If plant species are delimited into groups containing similar growth forms (i.e. more refined growth form categories, especially for herbaceous species), this might improve the performance of these models.

The variables that were most consistently important for predicting plant richness were two soil variables, soil pH and carbon. Other

Table 4

The best of the three GLM types (1: Full Model, 2: Minimum Adequate Model, 3: Shared Environments Model) for each richness response variable for the Lapalala (L) and Welgevonden (W) datasets, per the unclamped and clamped testing datasets. The best model was the model with the slope estimate closest to one. Also indicated is if this model had the highest R^2 value of the three models (if not then the model that did have the highest R^2 value is shown in brackets) and if this model was statistically significant.

Richness variable	Training dataset	Unclamped dataset			Clamped dataset		
		Best model	Highest R^2?	Statistically significant?	Best model	Highest R^2?	Statistically significant?
Total species	L	3	Yes	No	3	Yes	No
	W	2	Yes	No	2	No (3)	No
Grass species	L	1	No (2)	No	1	Yes	No
	W	3	No (2)	No	3	No (2)	No
Herb species	L	2	Yes	No	3	No (2)	No
	W	3	No (2)	No	3	Yes	No
Woody species	L	3	Yes	Yes	3	Yes	Yes
	W	3	Yes	Yes	2	No (3)	Yes
Family	L	3	Yes	Yes	3	No (2)	Yes
	W	3	Yes	Yes	3	Yes	Yes
Genus	L	3	No (2)	No	3	No (1)	No
	W	3	No (2)	No	3	No (2)	Yes

studies have also shown soil conditions to be important when modelling vascular plant species richness or individual plant species distributions (Bertrand et al., 2012; Chauvier et al., 2021; Coudun et al., 2006; Dubuis et al., 2013; Lobo et al., 2001; Scherrer and Guisan, 2019). Specifically, both Gould and Walker (1997) and Costanza et al. (2011) report soil pH as being the best predictor of plant species richness in their studies. Soil pH had a positive linear relationship with the species richness variables for 52 % of the models in this study, in agreement with Gould and Walker (1997), and a negative quadratic relationship (i.e. an inverted 'U' shaped graph) in 36 % of the models (in line with Costanza et al., 2011). Therefore, soil conditions are an important factor that should be used in plant richness models where possible.

4.1. Transferability

The transferability of all the GLMs and GBMs, across six different richness response variables and six different combinations of modelling techniques, was consistently poor, regardless of their performance in explaining richness in the reserve in which they were calibrated. Similarly, poor transferability has been reported in the literature for alpine plant species (Randin et al., 2006), for fish species across western United States of America (Wenger and Olden, 2012), for dung beetle species richness across the Iberian Peninsula (Hortal and Lobo, 2011) and for other species, including amphibians, reptiles, mammals and fungi, across a range of scales, including global, continental, and country scales and across oceanic islands (Capinha et al., 2018). However, poor transferability is not universally observed, and good model transferability has been observed, including for plants (e.g. Peterson et al., 2003).

One potential reason for poor model transferability is where models are overfitted, which can occur from the choice of statistical technique and the number of predictor variables (Capinha et al., 2018; Liu et al., 2020; Randin et al., 2006; Sequeira et al., 2016; Wenger and Olden, 2012). However, GLMs are considered robust to overfitting (Capinha et al., 2018; Lauria et al., 2015; Randin et al., 2006; Wenger and Olden, 2012) and the Minimum Adequate Models approach (applied here for both GLMs and GBMs) avoids overfitting by penalising complex models (Lauria et al., 2015; Randin et al., 2006; Wenger and Olden, 2012). Indeed, in this study Minimum Adequate Models did show better transferability than the more complex Full Models, in agreement with Randin et al. (2006) and Wenger and Olden (2012).

A second potential reason for poor transferability in this study is that some predictor variables used here are proxies for, and not direct measurement of, ecophysiologically-relevant environmental variables. As the relationship between proxy variables and the variables that they are assumed to represent may vary through space (e.g. elevation is a proxy for temperature, but thermal lapse rate can vary with humidity; Bonan, 2002), the utility of proxy variables for predicting biological patterns will be affected by how strongly and consistently they are linked to ecophysiologically-relevant variables under different conditions (Austin, 2002; Randin et al., 2006). For example, in this study, TWI was used as a proxy for soil moisture, but it is not a direct measure of water available in the soil (Kopecký et al., 2021; Riihimäki et al., 2021) and the TWI-soil moisture relationship may differ between soil types and seasons (Sørensen et al., 2006). This use of proxy variables could therefore have contributed to limiting the transferability of these models and using variables which are more accurate measures of in-field conditions could potentially improve model transferability.

A third possible reason for poor transferability may be the extent of novel environments in the testing dataset (Hortal and Lobo, 2011; Low et al., 2020; Parmentier et al., 2011; Randin et al., 2006). Even after removing variables with more than 50 % of their sites as novel environments, the remaining variables had an average of 11 % (ranging from 1 % to 37 %; Lapalala dataset) and 3 % (ranging from 0 % to 17 %; Welgevonden dataset) of their sites as novel environments. The two approaches that were implemented in this study to reduce the influence of novel environments (the use of the Shared Environments Model and clamping the testing datasets when predicting richness) both only slightly improved model transferability. This suggests that even the small proportion of novel environments present possibly still limits the transferability of the richness models. This result is similar to Hortal and Lobo (2011) who found poor transferability of models to testing areas with only slight differences (c. 12 % difference in training and testing datasets) in the environmental conditions compared to the training areas. Clamping the testing data is also useful to avoid extreme predictions when novel environments are present but will likely still be less accurate than predicting to testing areas with no novel environments (Capinha et al., 2018). Therefore, when the aim of the study is to predict to unsampled areas, it is advisable to select the training area carefully to incorporate the full range of environmental conditions present in the testing area, which, in turn, would avoid having to remove variables due to novel environments. Indeed, sampling areas of pronounced landscape-scale

heterogeneity that encompass large abiotic gradients should potentially be a priority for studies aiming to develop broadly transferable models.

A different type of mechanism which could account for the relationships between environmental predictor and richness response variables not being conserved between training and testing areas, and thus also potentially contributes to poor transferability, was proposed by Hortal and Lobo (2011) and focuses on the aggregative nature of species richness. The authors highlight that species richness is the sum of the occurrence patterns of different species, with each species potentially responding differently to multiple components in the environment. Therefore, even if a relationship between the species richness and the predictor variables can be identified (i.e. the model performs well in the area in which it was calibrated), these relationships could be different in the testing area due to differences in species composition, with, for example, potentially different growth forms and different life histories and, in consequence, different responses to environmental conditions exhibited by the species in the two areas (Hortal and Lobo, 2011). This may contribute to poor transferability in this study as there is a relatively low proportion of shared species between the reserves (35 % for Welgevonden and 36 % for Lapalala; Jaccard similarity index of 0.21). Moreover, in this study system the proportion of shared taxa increases with taxonomic level, potentially explaining why the transferability of family richness models was better than the transferability of genus and species richness models.

The difficulty in predicting plant species richness to other areas, due to the differences in species assemblages, could potentially be overcome by predicting species richness per growth form. Indeed, woody richness models displayed some of the best transferability, and the poorer performance of grass and herb species richness models might improve if the plant growth forms are better delimited (e.g. based on a finer division of growth forms, including geophytes, succulents and ferns as separate growth forms, since they potentially represent unique biotic and abiotic conditions). Plant growth forms are a simple way to distinguish plant functional groups and predicting their distribution potentially provides a clearer indication of the spatial patterns of the different growth forms in habitats and the relative composition of species within communities instead of just a single value of species richness (McNellie et al., 2021).

4.2. Steps to improve the transferability of models

The success of the transferability of models relies on the ability of the models to find causal relationships between the response and predictor variables, and on this relationship being conserved through space, including in novel environments (Capinha et al., 2018; Hortal and Lobo, 2011). Therefore, when developing models with the aim of predicting to unsampled areas, if novel environments in the testing area can be minimized, or avoided completely, that would be the best option to improve the transferability of biodiversity models. Sequeira et al. (2016) found good model transferability when predicting reef fish species richness from the larger Great Barrier Reef to a smaller reef which had a range of environmental conditions that fell within the Great Barrier Reef's environmental conditions. Similar results of model transferability being successful from one area to another but not the reverse, due to different predictor variable ranges, were also reported for fish species in the Baltic Sea (Sundblad et al., 2009). Therefore, it is suggested that when data is needed to model to unsampled areas, data collection should, where possible, include the full range of the environmental variables to minimize the effects of novel environments on model transferability. This does not necessarily require the sampling of a very large area but rather could be achieved by sampling two areas with different environmental conditions and then compiling the datasets together to create a calibration dataset that spans a wide range of conditions.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.sajb.2024.05.052.

References

Austin, M.P., 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. Ecol. Modell. 157, 101–118.

Bertrand, R., Perez, V., Gégout, J.C., 2012. Disregarding the edaphic dimension in species distribution models leads to the omission of crucial spatial information under climate change: the case of *Quercus pubescens* in France. Glob. Change Biol. 18, 2648–2660.

Bonan, G., 2002. Surface Climates. Ecological Climatology: Concepts and Applications. Cambridge University Press, Cambridge, pp. 248–291.

Capinha, C., Essl, F., Seebens, H., Pereira, H.M., Kühn, I., 2018. Models of alien species richness show moderate predictive accuracy and poor transferability. NeoBiota 38, 77–96.

Chauvier, Y., Thuiller, W., Brun, P., Lavergne, S., Descombes, P., Karger, D.N., Renaud, J., Zimmermann, N.E., 2021. Influence of climate, soil, and land cover on plant species distribution in the European Alps. Ecol. Monogr. 91, 1–14.

Codron, J., 2004. An Isotopic Comparison of Elephant (Loxodonta africana) Diets in the Kruger National Park and Welgevonden Game Reserve. University of Cape Town, Cape Town. (MSc Dissertation).

Costanza, J.K., Moody, A., Peet, R.K., 2011. Multi-scale environmental heterogeneity as a predictor of plant species richness. Landsc. Ecol. 26, 851–864.

Coudun, C., Gégout, J.C., Piedallu, C., Rameau, J.C., 2006. Soil nutritional factors improve models of plant species distribution: an illustration with *Acer campestre* (L.) in France. J. Biogeogr. 33, 1750–1763.

Díaz, S., Cabido, M., 2001. Vive la différence: plant functional diversity matters to ecosystem processes. Trends Ecol. Evol. 16, 646–655.

Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J.R.G., Gruber, B., Lafourcade, B., Leitão, P.J., Münkemüller, T., Mcclean, C., Osborne, P.E., Reineking, B., Schröder, B., Skidmore, A.K., Zurell, D., Lautenbach, S., 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography 36, 27–46.

Dubuis, A., Giovanettina, S., Pellissier, L., Pottier, J., Vittoz, P., Guisan, A., 2013. Improving the prediction of plant species distribution and community composition by adding edaphic to topo-climatic variables. J. Veg. Sci. 24, 593–606.

Duncan, R.P., Cassey, P., Blackburn, T.M., 2009. Do climate envelope models transfer? A manipulative test using dung beetle introductions. Proc. R. Soc. B: Biol. Sci. 276, 1449–1457.

Elith, J., Leathwick, J.R., Hastie, T., 2008. A working guide to boosted regression trees. J. Anim. Ecol. 77, 802–813.

Feng, X., Park, D.S., Liang, Y., Pandey, R., Papeş, M., 2019. Collinearity in ecological niche modeling: confusions and challenges. Ecol. Evol. 9, 10365–10376.

Ferrier, S., 2002. Mapping spatial pattern in biodiversity for regional conservation planning: where to from here? Syst. Biol. 51, 331–363.

Flood, B., 2015. Capturing Baseline Vegetation Data, Including an Assessment of Plant Sensitivity to Increased Acidity. University of the Witwatersrand, Johannesburg. in the Waterberg (MSc Dissertation).

Gould, W.A., Walker, M.D., 1997. Landscape-scale patterns in plant species richness along an arctic river. Can. J. Bot. 75, 1748–1765.

Guralnick, R.P., Hill, A.W., Lane, M., 2007. Towards a collaborative, global infrastructure for biodiversity assessment. Ecol. Lett. 10, 663–672.

Hendershot, W.H., Lalande, H., Duquette, M., 2008. Soil reaction and exchangeable acidity. In: Carter, M.R., Gregorich, E.G. (Eds.), Soil Sampling and Methods of Analysis. CRC Press, Boca Raton, USA, pp. 173–178.

Hijmans R.J., Phillips S., Leathwick J., Elith J. (2022). _dismo: species Distribution Modeling_. R package version 1.3-9, <https://CRAN.R-project.org/package=dismo>.

Hortal, J., Lobo, J.M., 2006. Towards a synecological framework for systematic conservation planning. Biodiversity Inform. 3, 16–45.

Hortal, J., Lobo, J.M., 2011. Can species richness patterns be interpolated from a limited number of well-known areas? Mapping diversity using GLM and Kriging. Braz. J. Nat. Conserv. 9, 200–207.

Kopecký, M., Macek, M., Wild, J., 2021. Topographic Wetness Index calculation guidelines based on measured soil moisture and plant species composition. Sci. Total Environ. 757, 1–10.

Kroetsch, D., Wang, C., 2008. Particle size distribution. In: Carter, M.R., Gregorich, E.G. (Eds.), Soil Sampling and Methods of Analysis. CRC Press, Boca Raton, USA, pp. 713–725.

Lauria, V., Power, A.M., Lordan, C., Weetman, A., Johnson, M.P., 2015. Spatial transferability of habitat suitability models of Nephrops norvegicus among fished areas in the northeast Atlantic: sufficiently stable for marine resource conservation? PLoS One 10, 1–19.

Liu, C., Wolter, C., Xian, W., Jeschke, J.M., 2020. Species distribution models have limited spatial transferability for invasive species. Ecol. Lett. 23, 1682–1692.

Lobo, J.M., Castro, I., Moreno, J.C., 2001. Spatial and environmental determinants of vascular plant species richness distribution in the Iberian Peninsula and Balearic Islands. Biol. J. Linnean Soc. 73, 233–253.

Low, B.W., Zeng, Y., Tan, H.H., Yeo, D.C.J., 2020. Predictor complexity and feature selection affect Maxent model transferability: evidence from global freshwater invasive species. Divers. Distrib. 27, 497–511.

McCullagh, P., Nelder, J.A., 1989. Generalized Linear Models, 2nd ed. Chapman and Hall, London.

McCune, B., 2007. Improved estimates of incident radiation and heat load using nonparametric regression against topographic variables. J. Veg. Sci. 18, 751–754.

McNellie, M.J., Oliver, I., Ferrier, S., Newell, G., Manion, G., Griffioen, P., White, M., Koen, T., Somerville, M., Gibbons, P., 2021. Extending vegetation site data and ensemble models to predict patterns of foliage cover and species richness for plant functional groups. Landsc. Ecol. 36, 1391–1407.

Mucina, L., Rutherford, M.C., 2006. The Vegetation of South Africa, Lesotho and Swaziland. Strelitzia 19. South African National Biodiversity Institute, Pretoria.

Parmentier, I., Harrigan, R.J., Buermann, W., Mitchard, E.T.A., Saatchi, S., Malhi, Y., Bongers, F., Hawthorne, W.D., Leal, M.E., Lewis, S.L., Nusbaumer, L., Sheil, D., Sosef, M.S.M., Affum-Baffoe, K., Bakayoko, A., Chuyong, G.B., Chatelain, C., Comiskey, J.A., Dauby, G., Doucet, J.L., Fauset, S., Gautier, L., Gillet, J.F., Kenfack, D., Kouamé, F.N., Kouassi, E.K., Kouka, L.A., Parren, M.P.E., Peh, K.S.H., Reitsma, J.M., Senterre, B., Sonké, B., Sunderland, T.C.H., Swaine, M.D., Tchouto, M.G.P., Thomas, D., van Valkenburg, J.L.C.H., Hardy, O.J., 2011. Predicting alpha diversity of African rain forests: models based on climate and satellite-derived data do not perform better than a purely spatial model. J. Biogeogr. 38, 1164–1176.

Peterson, A.T., Papes, M., Kluza, D.A., 2003. Predicting the potential invasive distributions of four alien plant species in North America. Weed Sci. 51, 863–868.

Phillips, S.J., 2008. Transferability, sample selection bias and background data in presence-only modelling: a response to Peterson et al. (2007). Ecography 31, 272–278.

Pool-Stanvliet, R., 2013. A history of the UNESCO Man and the Biosphere Programme in South Africa. S. Afr. J. Sci. 109, 1–6.

QGIS, 2021. QGIS Geographic Information System. Open Source Geospatial Foundation Project.

R Core Team, 2021. R: A language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Randin, C.F., Dirnböck, T., Dullinger, S., Zimmermann, N.E., Zappa, M., Guisan, A., 2006. Are niche-based species distribution models transferable in space? J. Biogeogr. 33, 1689–1703.

Redfern, J.V., Moore, T.J., Fiedler, P.C., de Vos, A., Brownell, R.L., Forney, K.A., Becker, E.A., Ballance, L.T., 2017. Predicting cetacean distributions in data-poor marine ecosystems. Diversity Distrib. 23, 394–408.

Riihimäki, H., Kemppinen, J., Kopecký, M., Luoto, M., 2021. Topographic wetness index as a proxy for soil moisture: the importance of flow-routing algorithm and grid resolution. Water Resour. Res. 57, 1–21.

Rocchini, D., Hortal, J., Lengyel, S., Lobo, J.M., Jiménez-Valverde, A., Ricotta, C., Bacaro, G., Chiarucci, A., 2011. Accounting for uncertainty when mapping species distributions: the need for maps of ignorance. Prog. Phys. Geogr. 35, 211–226.

Rousseau, J.S., Betts, M.G., 2022. Factors influencing transferability in species distribution models. Ecography e06060, 1–13.

Rousseeuw, P.J., 1991. A diagnostic plot for regression outliers and leverage points. Comput. Stat. Data Anal. 11, 127–129.

Rousseeuw, P.J., van Zomeren, B.C., 1990. Unmasking multivariate outliers and leverage points. J. Am. Stat. Assoc. 85, 633–639.

Scherrer, D., Guisan, A., 2019. Ecological indicator values reveal missing predictors of species distributions. Sci. Rep. 9, 1–8.

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 Ecol. Evol. 9, 1250–1264.

Sequeira, A.M.M., Mellin, C., Lozano-Montes, H.M., Vanderklift, M.A., Babcock, R.C., Haywood, M.D.E., Meeuwig, J.J., Caley, M.J., 2016. Transferability of predictive models of coral reef fish species richness. J. Appl. Ecol. 53, 64–72.

Skjemstad, J.O., Baldock, J.A., 2008. Total and organic carbon. In: Carter, M.R., Gregorich, E.G. (Eds.), Soil Sampling and Methods of Analysis. CRC Press, Boca Raton, USA, pp. 225–237.

Slater, K., Long, P., 2012. Operation Wallacea and WEI Field Research Report: Welgevonden Game Reserve. Lincoln: Operation Wallacea, South Africa.

Soberón, J., Peterson, A.T., 2004. Biodiversity informatics: managing and applying primary biodiversity data. Philos. Trans. R. Soc. Lond. Ser. B: Biol. Sci. 359, 689–698.

Sørensen, R., Zinko, U., Seibert, J., 2006. On the calculation of the Topographic Wetness Index: evaluation of different methods based on field observations. Hydrol. Earth Syst. Sci. 10, 101–112.

Sundblad, G., Härmä, M., Lappalainen, A., Urho, L., Bergström, U., 2009. Transferability of predictive fish distribution models in two coastal systems. Estuar. Coast Shelf Sci. 83, 90–96.

Warren, D.L., Wright, A.N., Seifert, S.N., Shaffer, H.B., 2014. Incorporating model complexity and spatial sampling bias into ecological niche models of climate change risks faced by 90 California vertebrate species of concern. Diversity Distrib. 20, 334–343.

Wenger, S.J., Olden, J.D., 2012. Assessing transferability of ecological models: an underappreciated aspect of statistical validation. Methods Ecol. Evol. 3, 260–267.

Womack, C.P., 2022. The Classification and Mapping of Lapalala Wilderness Reserve, Limpopo. University of Pretoria, Pretoria.

Ziadi, N., sen Tran, T., 2008. Mehlich 3-extractable elements. In: Carter, M.R., Gregorich, E.G. (Eds.), Soil Sampling and Methods of Analysis. CRC Press, Boca Raton, USA, pp. 81–88.