**Pedology and plant provenance can improve predictions of species distributions of the Australian native flora: a calibrated and validated modelling exercise on 5,033 species**

**Abstract**

Species distribution models (SDMs) are valuable tools for assessing species' responses to environmental factors and identifying areas suitable for their survival. The careful selection of input variables is critical, as their interactions, and correlations with other environmental factors can affect model performance. This study evaluates the influence of climate and soil variables on SDMs’ performance for 5,033 Australian plant species, selected to represent the largest phylogenetic diversity of native terrestrial vascular flora. Using an ensemble of correlative models, we assessed the predictive performance of climate and soil variables, individually and in combination, across four distinct ecoregions: Desert (*n* = 640 species), Mediterranean (*n* = 1,246), Temperate (*n* = 1,936), and Tropical (*n* = 1,211). Our results demonstrate that on a continental scale, climate variables have a greater influence on plant distributions than soil variables. Although incorporating soil and climate variables enhanced model performance in some ecoregions, our results indicate that relying solely on small-scale variables such as soil may increase the likelihood of overfitting. In soil-only models, Clay content (CLY), Nitrogen Total Organic (NTO), and Soil Organic Carbon (SOC) were important across modelled species, with their relevance varying by ecoregion. Our findings have significant implications for understanding the interplay between climate, soil, and plant distribution within diverse ecoregions. By highlighting the crucial role of climate in large-scale models, this study serves as a foundation for developing more accurate predictions of plant distributions, ultimately improving model accuracy for biodiversity assessments.

**Keywords:** Species Distributions; Habitat Suitability; Ecoregion-specific Data; Predictor Variables; Ensemble Models; Precipitation Seasonality

**Introduction**

Modelling the habitat requirements of species has become an increasingly important tool in ecology, biogeography, and conservation (Austin and Meyers 1996, Jarvis and Robertson 1999, Veech 2021, Williams and Hero 2001). Species distribution models (SDMs) are numerical tools that integrate species occurrence and abundance records with environmental data to predict species distributions (Franklin 2010, Guisan and Zimmermann 2000, Kearney, et al. 2010, Stauffer 2002). By considering both abiotic and biotic factors, SDMs offer insights into species' potential occurrences across a landscape (Elith and Leathwick 2009). These models are often extrapolated across space and time to forecast species’ responses to climate change (Brown, et al. 2016, Edwards and Richardson 2004, Hazen, et al. 2013, Ramirez‐Cabral, et al. 2017, Ramos, et al. 2019, Shabani, et al. 2017, Shabani, et al. 2019, Wauchope‐Drumm, et al. 2020). SDMs have become essential for identifying priority areas for conservation and guiding resource allocation, making them invaluable for informed environmental management and biodiversity preservation. By incorporating both climatic and non-climatic factors, SDMs provide insights into habitat suitability and species responses to environmental changes, supporting sustainable ecosystem management strategies.

SDMs can be categorized into two main groups: mechanistic-based and correlative-based models (Dormann, et al. 2012, Kearney and Porter 2009). The performance of both types of models is sensitive to the algorithms (Anderson, et al. 2003, Shabani, et al. 2016), variables (Araújo and Guisan 2006, Warren and Seifert 2011), thresholds (Shabani and Kumar 2014), and the methods used to evaluate their accuracy (Shabani, et al. 2018).

Mechanistic models employ explicit functions to characterize relationships among different environmental variables and ecosystem components which are typically defined a priori based on ecological theory or direct empirical evidence (Connolly, et al. 2017, Dormann, Schymanski, Cabral, Chuine, Graham, Hartig, Kearney, Morin, Römermann, Schröder and Singer 2012). However, mechanistic models require detailed and accurate physiological data for each species, making them less practical for large-scale studies of entire floras (e.g., thousands of species spanning entire phylogenies). Although correlative models are conceptually simpler than mechanistic models, they can perform as robustly (or even better) than mechanistic models in predicting distributions when appropriate input variables are used (Muhling, et al. 2017, Robertson, et al. 2003). Additionally, their independence from explicit assumptions minimizes confirmation bias by relying on observed data to establish relationships between species distributions and environmental variables, rather than depending on predefined ecological or mechanistic hypotheses. This data-driven approach improves the ability to detect emergent patterns and reduces the likelihood of misinterpretation stemming from incorrect or overly simplified assumptions about underlying processes (Connolly, Keith, Colwell and Rahbek 2017). However, it is far from obvious which input variables will have the highest predictive power in correlative models. Often the selection of input variables with low predictive power can result in ecologically unreasonable and inaccurate model output (Shabani, Kumar and Ahmadi 2016). In this study, we focus on correlative modelling because this method is easier to implement for entire floras (i.e., relying on readily available occurrence and environmental data) and is the most commonly used approach in SDM applications (Robinson, et al. 2017). There has been a rapid increase in the availability of correlative tools for modelling species distributions (Elith, et al. 2006). Yet efforts to model entire floras are in their early stages. Consequently, there is a growing need to comprehensively examine the performance and biases of SDMs with particular reference to the selection of input variables.

Significant attention has been devoted to addressing key challenges in SDMs, including the quality and quantity of species data, spatial resolution, and modelling methodologies (Ahmadi, et al. 2023, Datta, et al. 2020, Shabani, Kumar and Ahmadi 2016, Valavi, et al. 2022, Wang, et al. 2016). However, the suitability of environmental data (including which climate and non-climate variables improve the accuracy of SDM predictions) for ecological modelling purposes has not received as much focus, despite its significant influence on modelling outcomes and accuracy (e.g., Fourcade, et al. (2018)). Currently, SDM environmental input data are freely available from various databases. For instance, the WorldClim (Hijmans, et al. 2004) and CHELSA databases (CHELSA – Free climate data at high resolution, available at https://chelsa-climate.org/) provide climatic variables at a variety of spatiotemporal scales, and other environmental variables directly or indirectly related to species distributions including topographic variables (e.g. elevation, slope, aspect, and hillslope position), land use/land cover characteristics (e.g., Searle, et al. (2022) and physicochemical soil properties are also available (Hulshof and Spasojevic 2020)). Given the array of available input data for SDMs, the question of "which input data are best to use?" arises (Ahmadi, Hemami, Kaboli and Shabani 2023, Arenas‐Castro, et al. 2022, Beck, et al. 2014, Kearney and Porter 2009, Velazco, et al. 2017). Additionally, there is growing concern regarding the assumptions and approaches used in SDM studies, especially when applied to broader spatial scales, such as entire biomes and ecoregions (Araújo and Guisan 2006, Luoto, et al. 2006, Svenning, et al. 2008). Recent studies debate on whether climate variables alone provide sufficient predictive power or if covariates representing other key environmental variables of a species’ niche are necessary. While some studies have included non-climatic variables in SDMs, inconsistencies in the number and selection of these variables complicate comparisons (Iverson, et al. 2008, Keith, et al. 2008, Randin, et al. 2009). Pedology and the underlying soil properties are examples of non-climate variables that could improve SDM accuracy in vascular floras, given that soil properties significantly influence the physiology and ecology of terrestrial plants, and hence their distribution (Dubuis, et al. 2013, Velazco, Galvao, Villalobos and De Marco Junior 2017). Soil is vital in the provision of water and nutrients (Aerts and Chapin 1999). It also physically supports root growth (Martre, et al. 2002), and influences how the microbiome operates in the rhizosphere (Berg and Smalla 2009). While the use of soil properties in plant SDMs seems intuitive, little empirical evidence exists on the extent to which the inclusion of pedological factors in SDMs could improve modelling outcomes. Indeed, studies that have incorporated pedological factors have been limited to one or few species (Fitzpatrick, et al. 2008, Hageer, et al. 2017, Martinson, et al. 2011) and no information exists on whether these findings apply to entire terrestrial floras.

In this study, we address the following questions: (1) How do predicted distribution maps change when using only soil datasets, only climate datasets, or a combination of both? (2) do ecoregions differ in the relative importance of variables for predicting the distributions of native species within them?

To answer these questions, we utilized three input variable datasets: (i) climate data alone (CHELSA – Free climate data at high resolution, available at https://chelsa-climate.org/), (ii) soil data alone (Grundy, et al. 2015), and (iii) both of these datasets combined. With each of these input datasets, we used the same ensemble SDM approach consisting of a set of correlative species distribution models for the ‘present time’ (i.e., covariates datasets spanned 1979-2013). Our modelling protocol aimed to find and utilize the most important environmental variables per species (rather than use a consistent set of variables across all the species). We applied this approach to model most of the Australian native flora across Australian terrestrial ecoregions, for which species data was obtained from various resources. In Australia, diverse ecoregions—from arid deserts to temperate and tropical zones—present large and diverse environmental conditions that influence species distribution patterns. An ecoregion-specific approach may be critical for accurately modelling plant distributions, as species responses to climate, soil, and vegetation characteristics may vary significantly across ecoregions. Ecoregion are distinct geographical units defined by unique combination of species, natural communities, and environmental conditions. These areas are smaller and more localized than broader biomes, and they reflect the complex interactions of climate, soil types, topography, and biological factors (Olson, et al. 2001). The investigated ecoregions included (ⅰ) desert and xeric shrub lands (hereafter Desert ecoregion), which accounted for around 49% of the total study area across Australia; (ⅱ) mediterranean forests, woodlands, and scrub (hereafter Mediterranean), covering around 11%; (ⅲ) temperate grasslands, savannas, and forests (hereafter Temperate), making up 12 %; and (ⅳ) tropical and subtropical grasslands and savannas (hereafter Tropical), representing 28% of the total study area across Australia. By assessing the individual and combined importance of climate and soil properties on SDM predictions of most native plants across different Australian terrestrial ecoregions, we aim to provide a roadmap to input data selection concerning different ecoregion and phylogeny of whole continental vascular floras.

**Methods**

## Species selection & Species data pre-processing

Native plant species’ occurrences were obtained from various resources. including Australian Native Plants Society (Australian Native Plants Society; 2024 ), Global Biodiversity Information Facility (GBIF) (Global Biodiversity Information Facility 2024), and state-specific flora databases (e.g., Electronic Flora of South Australia (Australia; 2024) and among others). These species are distributed across four Australian terrestrial ecoregions. We used the Terrestrial Ecoregions of Australia map version 7 (IBRA 7.0) (<https://data.gov.au/data/dataset/interim-biogeographic-regionalisation-for-australia-ibra-version-7-regions>) that includes seven primary ecoregions, of which we focused on the four largest ones. Further, we downloaded the presence points of all Australian terrestrial plant with more than 100 presence points recorded in Australia from the Atlas of Living Australia (ALA). This initial ALA dataset consisted of 11,322 plant species, which make up over 60% of the estimated flowering plants on the continent, comprising aquatic plants excluded from the study (Broadhurst and Coates 2017). We then matched the scientific names of species with their establishment status—whether native or naturalized—using the Australian Plant Census (APC) database (<https://biodiversity.org.au/>). From this, we focused only on species identified as native, reducing the number of species to 9,670. To achieve our main objective of performing an ecoregion-specific SDM analysis, each species was associated to an ecoregion based on the proportion of its presence points across four Australian ecoregions. In this study, a species was assigned to a particular ecoregion if its frequency within that ecoregion exceeded thresholds of 0.6, 0.7, 0.8, 0.9, and 1 (60%, 70%, 80%, 90%, and 100% frequency). All frequency thresholds were tested and documented; however, only the results for thresholds of 0.6 and 1 are included in the main manuscript, while the remaining results are provided in the Supplementary Information (SI). The spatial resolution of our study was set to 2.5 arcminutes. To ensure that each species had no more than one presence point per grid cell, we applied a geographical filtering approach (Kramer-Schadt, et al. 2013) and removed repeated presence points in each cell. We also excluded records that lacked precise coordinates and eliminated suspected outliers (i.e., records with coordinates falling outside the known range of the species) utilizing the 'CoordinateCleaner ' package in R version 3.4.4 (Zizka, et al. 2019). Ultimately, the number of acceptable species (after removal of aquatic plant species) with at least 60% frequency in any of the four ecoregions was reduced to 5,033 (SI, Table S1). Phylogenetic coverage for the 5,033 angiosperm species analysed was measured using the extended version of the Smith and Brown phylogeny implemented through the R package V.PHYLOMAKER2 (Jin and Qian 2022). This was performed to ensure species selection to cover the largest breadth of phylogenetic diversity in our target flora. Details of the phylogenetic methods and the resulting tree are provided in the SI (SI, Figure S1).

To mitigate the influence of potential biases in species location data used in modelling, in addition to the geographical filtering, we implemented two further strategies: environmental filtering (Varela, et al. 2014), and background weighting (Elith, et al. 2010). Environmental filtering was implemented using the *occfilt\_env* function of the ‘flexsdm’ package in R, following the procedure by Varela, Anderson, García‐Valdés and Fernández‐González (2014). In this process, we created a multidimensional environmental grid based on selected environmental variables, where each dimension represents a variable. This grid divides the range of each variable into discrete intervals, or 'bins,' which are categories representing specific ranges of values. We then filtered the presence points by randomly retaining one point within each unique combination of these bins, thereby reducing the potential clustering of points in similar environmental conditions (Castellanos, et al. 2019, Varela, Anderson, García‐Valdés and Fernández‐González 2014).

To address the issue of potential biases associated with selecting pseudo-absences, we used background weighting. This method involved providing the models with environmental data that matched the spatial biases observed in the occurrence data (Renner and Warton 2013, Sequeira, et al. 2012). To generate background weighting data, a weighting surface was created to emphasizes areas that are geographically less dense in occurrence records. Following the methodology outlined by Elith et al. (Elith, Kearney and Phillips 2010), for each species a 2-dimensional kernel density raster was first created from the presence points using *kde2d* function of the ‘MASS’ package in R. Subsequently, a set of 10,000 background points were allocated based on the probability distribution of the density raster. This approach addresses the bias caused by spatially imbalanced or biased data, favouring pseudo-absences in densely sampled areas over those in sparsely sampled areas (Ahmadi, Hemami, Kaboli and Shabani 2023).

## Explanatory variables

We obtained 19 climate variables from the CHELSA database (Karger, et al. 2017) in raster format at a resolution of 2.5 arc-minutes. Climate variables can be highly correlated, which can cause problems in SDM (Naimi 2015). We then selected nine climate variables for the modelling based on the variance inflation factor (Table 1), i.e.,: Annual Mean Temperature (bio1), Annual Mean Diurnal Temperature Range (bio2), Temperature Seasonality (bio4), Maximum Temperature of Warmest Month (bio5), Minimum Temperature of Coldest Month (bio6), Annual Precipitation (bio12), Precipitation of Wettest Month (bio13), Precipitation of Driest Month (bio14), and Precipitation Seasonality (bio15). In addition to climate variables, we considered eight soil variables: Bulk Density (BDW), Cation Exchange Capacity (CEC), Clay (CLY), Depth of Soil (DES), Nitrogen Total Organic (NTO), Total Phosphorus (PTO), Topographic Wetness Index (TWI), and Soil Organic Carbon (SOC). These variables were incorporated using raster maps produced by the Soil and Landscape Grid of Australia (SLGA; (Grundy, Rossel, Searle, Wilson, Chen and Gregory 2015)) at the same resolution as the climate data (i.e., at 2.5 arc-minutes) (Table 1). The soil variables provide approximate measures of water and nutrient availability for plants (Duursma, et al. 2013), and are likely impacted by soil parent material, climate, topography and soil age (Delgado-Baquerizo, et al. 2020, McBratney, et al. 2003).

Although we considered 17 environmental variables in total (nine climate variables and eight soil variables), we only used a subset of these variables in each species’ SDM. For each species, we selected the most informative and predictive subset of variables out of the candidate variables, choosing five climate and five soil variables using *covsel.embed* function of the ‘covsel’ package in R (Adde, et al. 2023). This function is used to reduce the dimensionality of the predictor-variables set and find highly informative covariates. It achieves this by integrating a collinearity-filtering algorithm with three model-specific embedded regularization techniques: a generalized linear model with elastic net regularization, a generalized additive model with null-space penalization, and a guided regularized random forest. We then used the selected variables in the SDMs of each species with three arrangements: climate-only variables (five variables), soil-only variables (five variables), and climate + soil variables (ten variables).

Table 1. Environmental input variables (climate & soil) used in this study.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Name** | **Abbreviation** | **Unit** |
| **Climate** | Annual mean temperature | bio1 | °C\*10 |
| Annual mean diurnal temperature range | bio2 | °C×10 |
| Temperature seasonality | bio4 | % |
| Maximum temperature of warmest month | bio5 | °C×10 |
| Minimum temperature of coldest month | bio6 | °C×10 |
| Annual precipitation | bio12 | mm |
| Precipitation of wettest month | bio13 | mm |
| Precipitation of driest month | bio14 | mm |
| Precipitation seasonality | bio15 | % |
| **Soil** | Bulk Density | BDW | g cm-3 whole earth |
| Cation Exchange Capacity | CEC | mEq per 100g |
| Clay Content | CLY | % |
| Depth of Soil | DES | cm |
| Nitrogen Total Organic | NTO | % |
| Total Phosphorus | PTO | % |
| Topographic Wetness Index | TWI | adimensional |
| Soil Organic Carbon | SOC | % |

## Species distribution modelling

To predict suitable habitats and generate individual species distribution models, we utilized the biomod2 package ensemble platform in the R v. 3.3.4 environment (Thuiller, et al. 2009). This platform employs multiple modelling techniques simultaneously, which can be used to build a consensus or "ensemble" model (Araújo and New 2007, Thuiller, Lafourcade, Engler and Araújo 2009). We specifically selected four commonly used techniques known for their effectiveness in predicting species distribution: generalized linear model (GLM), generalized additive model (GAM), maximum entropy (MaxEnt), and random forest (RF). The choice of these techniques was based on various factors, including their ease of use, ability to handle uncertainty, and capacity to provide reliable error estimates (Merow, et al. 2014). Furthermore, the combination of two regression-based methods with two complex machine learning models allowed us to balance extrapolation (under-fitting) and interpolation (over-fitting) (Ahmadi, Hemami, Kaboli and Shabani 2023, Merow, Smith, Edwards Jr, Guisan, McMahon, Normand, Thuiller, Wüest, Zimmermann and Elith 2014). Each technique was applied with ten replicates for each species. We used the area under the ROC curve (AUC) and true-skill statistic (TSS) to evaluate model performance. The AUC is a robust measure of discrimination ability, but its ecological realism in modelled distributions can be limited, especially when using presence and pseudo-absence data rather than true absences [93-95]. Hence, we also calculated TSS as a classification accuracy measure since it is independent of prevalence (the ratio of presence versus absence records) and provides a more robust assessment of the predictive performance of SDMs when converted into presence/absence data (Hageer, Esperón-Rodríguez, Baumgartner and Beaumont 2017). To evaluate each model, the performance metrics were averaged across the 10-fold validation subsets. We examined the relative importance of each explanatory variable to each model by analyzing the Pearson rank correlation between standard predictions and those based on five random permutations for each variable separately (Thuiller, Lafourcade, Engler and Araújo 2009). For every variable, we then calculated the average variable importance across all implemented models.

## Sensitivity analysis of the species ecoregion assignment

The thresholds used to assign species to ecoregions (i.e., the percentage of observations from an ecoregion required for assignment) can influence comparisons of SDM performance between ecoregions. To account for this, we conducted a sensitivity analysis across a range of thresholds (60%, 70%, 80%, 90%, and 100%) to refine species allocation within ecoregions. We evaluated SDM performance and the relative importance of variables across all thresholds using a two-way ANOVA, followed by a post-hoc Tukey HSD (Honestly Significant Difference) test, to compare results across five thresholds and four ecoregions. Due to space limitations and the consistency of results across thresholds, we present findings from the 60% and 100% thresholds in the main text, with results from other thresholds available in the SI.

## Species response to individual environmental variables

We extracted raster values corresponding to specific geographic points in R, starting by loading the occurrence records for each species as spatial objects and the environmental variables as separate raster layers using the `sf` and `raster` packages, respectively. The `extract()` function from the `raster` package was then used to retrieve raster values at each point location. These results enabled comparison of environmental ranges and suitability across ecoregions, and facilitated the identification of overall species response patterns.

**Results**

## Species richness patterns under different density thresholds

Table 2 summarizes species richness in relation to occurrence thresholds (60%, 70%, 80%, 90%, and 100%) in each ecoregion. As the threshold increased from 60% to 100%, the species richness decreased across all ecoregions. Temperate ecoregion consistently has the highest species richness across all thresholds, indicating that this ecoregion might be more conducive to species survival or more diverse in suitable habitats compared to the others (Figure 1). However, this result should be interpreted with caution, as the higher species richness could also reflect potential sampling bias; the Temperate ecoregion may have been sampled more intensively. The Mediterranean and Tropical ecoregions also have a relatively high species richness, but they decrease more steeply as the threshold increases, particularly in the Mediterranean ecoregion. The Desert ecoregion has the lowest species richness at all thresholds, reflecting its extreme environmental conditions that support fewer species, as well as its relatively limited survey and sampling effort.

Table 2. Species richness in relation to occurrence thresholds of 60%, 70%, 80%, 90%, and 100% within the investigated ecoregions (Desert, Mediterranean, Temperate, Tropical).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Threshold** | **Desert** | **Mediterranean** | **Temperate** | **Tropical** |
| **60 % of total species occurrences** | 640 spp | 1,246 | 1,936 | 1,211 |
| **70 %** | 547 | 1,132 | 1,747 | 1,110 |
| **80 %** | 445 | 1,042 | 1,530 | 997 |
| **90 %** | 289 | 921 | 1,221 | 862 |
| **100 %** | 38 | 476 | 502 | 453 |

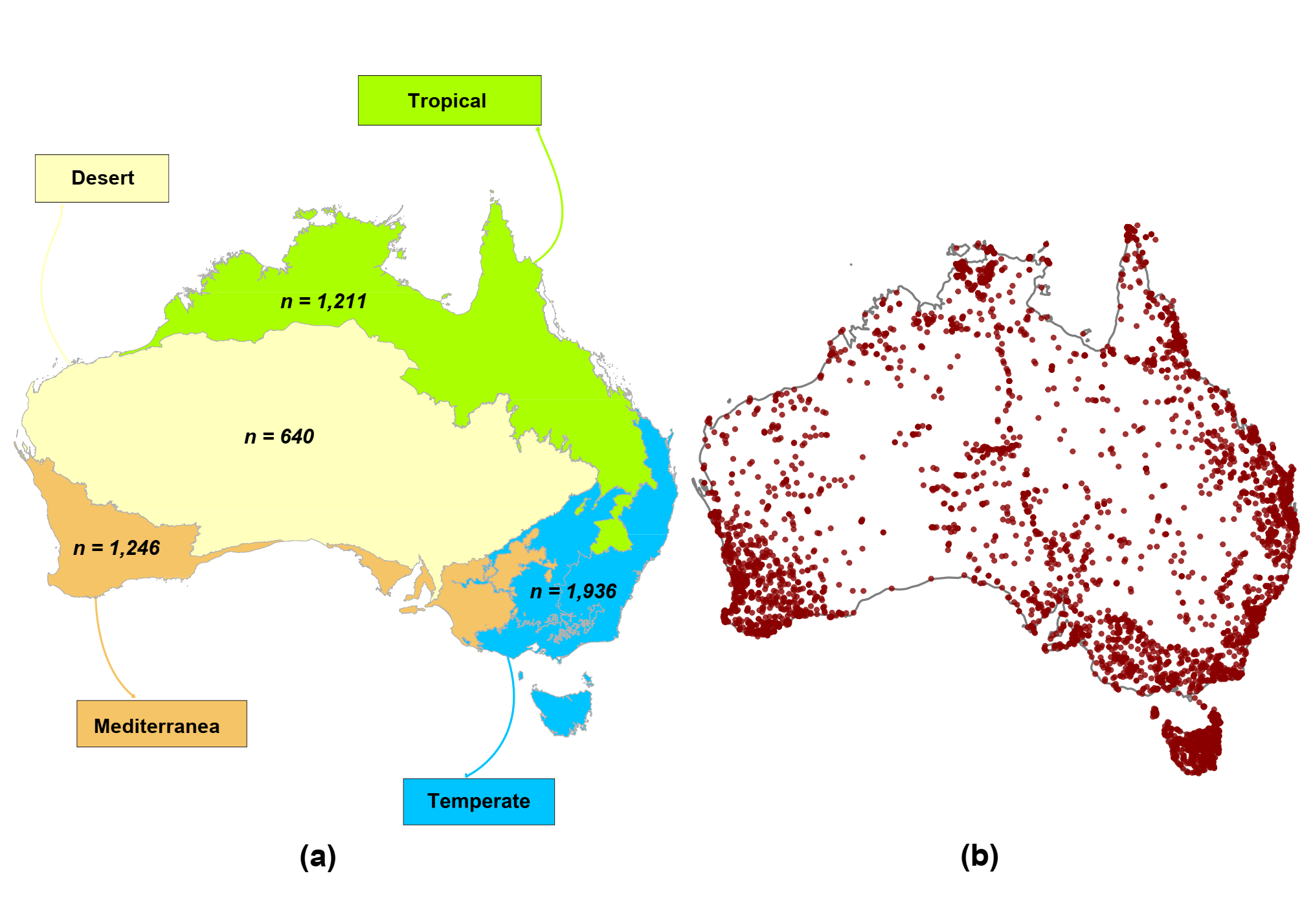


Figure 1. Panel (a) depicting the four major ecoregions assessed in this study—Desert (640 species), Mediterranean (1,246 species), Temperate (1,936 species), and Tropical (1,211 species)—showing the number of species considered in each ecoregion for evaluating the predictive performance of climate and soil variables. Panel (b) shows spatial distribution of plant families (*n =* 166) used in the phylogenetic tree (refer to Figure S1).

## Predictive performance

### Density occurrence threshold across ecoregion

The highest AUC and TSS values were observed at the 100% occurrence density threshold across all four ecoregions, regardless of whether soil, climate, or both types of input variables were used in the SDMs (Figure 2). Conversely, the 60% threshold, which includes species with broader distributions, demonstrated the lowest model performance, particularly when only soil predictor variables were used. This pattern indicates that model accuracy might lower for species with wider distributions.

### Ecoregions

Overall, the predictive performance was highest in the Temperate ecoregion, followed by the Mediterranean ecoregion (Figure 2). The Tropical ecoregion showed moderate performance, while the Desert exhibited the lowest predictive accuracy (Figure 2)

### Model type

The comparison of predictive performance among the model types revealed that the MaxEnt model had the highest accuracy (TSS and AUC), particularly when both climate and soil variables were used together (Figure 2). Conversely, RF models using only soil data showed the lowest accuracy (TSS; Figure 2).

|  |  |
| --- | --- |
| **Threshold 60** |  |
| **Threshold 100** |  |

Figure 2. SDM performance results of generalized linear model (GLM), generalized additive model (GAM), generalized boosting model (GBM), and maximum entropy (MaxEnt) for species across Australian ecoregions under 60% (*n =* 5,033) and 100% (*n =* 1,469) density occurrence ecoregion assignment thresholds. Refer to Figure S2 for details for 70%, 80% and 90% density occurrence threshold.

## Variables importance

When combining climate and soil variables, climate variables had a greater contribution to model predictions than soil variables, regardless of the threshold used. To facilitate comparison, we calculated average relative contribution of variables in climate + soil models across all species in each ecoregion (Table 3). The results showed that, given the 100% and 90% assignment thresholds, i.e., specialist plants, Annual Mean Temperature (bio1) was by far the most important variable in the Mediterranean ecoregion. For Deserts ecoregion, Temperature seasonality (bio4) and Annual precipitation (bio12), for Temperate ecoregion, Precipitation of driest month (bio14), Maximum temperature of warmest month (bio5), and Annual Mean Temperature (bio1), and for Tropical ecoregion, annual precipitation, temperature seasonality, and annual mean temperature, were the most important variables. Assessing relative contribution of soil variables to specialist plants showed that for Deserts ecoregion, Nitrogen Total Organic (NTO) and Cation Exchange Capacity (CEC), for Mediterranean ecoregions, Clay Content (CLY) and Soil Organic Carbon (SOC), for Temperate ecoregion, Soil Organic Carbon (SOC) and Bulk Density (BDW), and for Tropical ecoregion, SOC and NTO were the two most important soil variables (Table 3). Interestingly, the mean relative contribution of climatic variables was highest in the Tropical ecoregion and lowest in the Deserts ecoregion. Conversely, the mean relative contribution of soil variables was greatest in the Deserts ecoregion and lowest in the Tropical ecoregion (Table 3). For both climatic and soil variables, these contrasts consistently became more pronounced as the assignment threshold increased from 60% (common species) to 100% (specialist species).

Table 3. Average relative contribution of variables in climate + soil models across all species in each ecoregion. A complete set variables contribution across all assignment thresholds is provided in the SI Table S2.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | **Threshold 60** | | | | **Threshold 100** | | | |
| Desert | Mediterranean | Temperate | Tropical | Desert | Mediterranean | Temperate | Tropical |
| **Climatic variables** | bio1 | 13.55 | 33.82 | 18.66 | 17.14 | 8.94 | 34.16 | 15.92 | 14.76 |
| bio2 | 3.59 | 1.49 | 3.22 | 2.48 | 5.8 | 1.14 | 3.32 | 1.9 |
| bio4 | 13.05 | 5.11 | 6.09 | 12.15 | 11.18 | 4.11 | 6.93 | 15.75 |
| bio5 | 9.09 | 6.65 | 14.76 | 4.55 | 5.52 | 4.64 | 16.32 | 2.46 |
| bio6 | 7.54 | 10.05 | 4.97 | 6.63 | 10.73 | 12.32 | 5.62 | 8.47 |
| bio12 | 9.88 | 6.27 | 6.88 | 16.17 | 10.28 | 4.95 | 5.43 | 17.15 |
| bio13 | 5.59 | 3.65 | 2.16 | 9.29 | 6.28 | 3.64 | 1.68 | 11.1 |
| bio14 | 0.82 | 2.08 | 16.28 | 1.61 | 1.76 | 1.21 | 17.96 | 1.07 |
| bio15 | 5.34 | 8.11 | 3.59 | 8.78 | 5.24 | 10.72 | 4.1 | 8.92 |
| **Mean** |  | **7.61** | **8.58** | **8.51** | **8.75** | **7.3** | **8.54** | **8.59** | **9.06** |
| **Soil variables** | BDW | 2.47 | 2.58 | 3.82 | 1.99 | 2.75 | 2.7 | 4.84 | 1.7 |
| CEC | 5.66 | 3.49 | 2.13 | 2.38 | 5.89 | 2.7 | 1.74 | 2.49 |
| CLY | 3.75 | 5.9 | 2.28 | 3.41 | 3.73 | 7.22 | 1.6 | 3.06 |
| DES | 3.48 | 0.89 | 0.95 | 0.9 | 2.71 | 0.53 | 0.84 | 0.5 |
| NTO | 7.7 | 3.1 | 3.21 | 3.27 | 10.74 | 2.74 | 4.67 | 3.39 |
| PTO | 2.07 | 1.65 | 0.73 | 1.15 | 2.32 | 1.19 | 0.4 | 0.77 |
| SOC | 3.8 | 4.97 | 8.98 | 6.55 | 4.13 | 5.83 | 7.59 | 5.38 |
| TWI | 2.63 | 0.2 | 1.29 | 1.58 | 2.01 | 0.21 | 1.04 | 1.15 |
| **Mean** |  | **3.94** | **2.85** | **2.92** | **2.65** | **4.29** | **2.89** | **2.84** | **2.31** |

## Species response based on environmental variables

The response of modelled plants to bioclimatic and soil variables vary among the studied ecoregions, highlighting the unique environmental conditions that characterize each ecoregion (Figure 3). Some ecoregions, like the Desert, show more extreme conditions with generally narrower ranges (see yellow line for Soil Organic Carbon (SOC) or Cation Exchange Capacity (CEC) or Temperature seasonality (bio4) or Annual precipitation (bio12) in (Figure 3), while others like the Tropical or Mediterranean ecoregions exhibit broader ranges (see green and orange lines for bio4 and bio5).

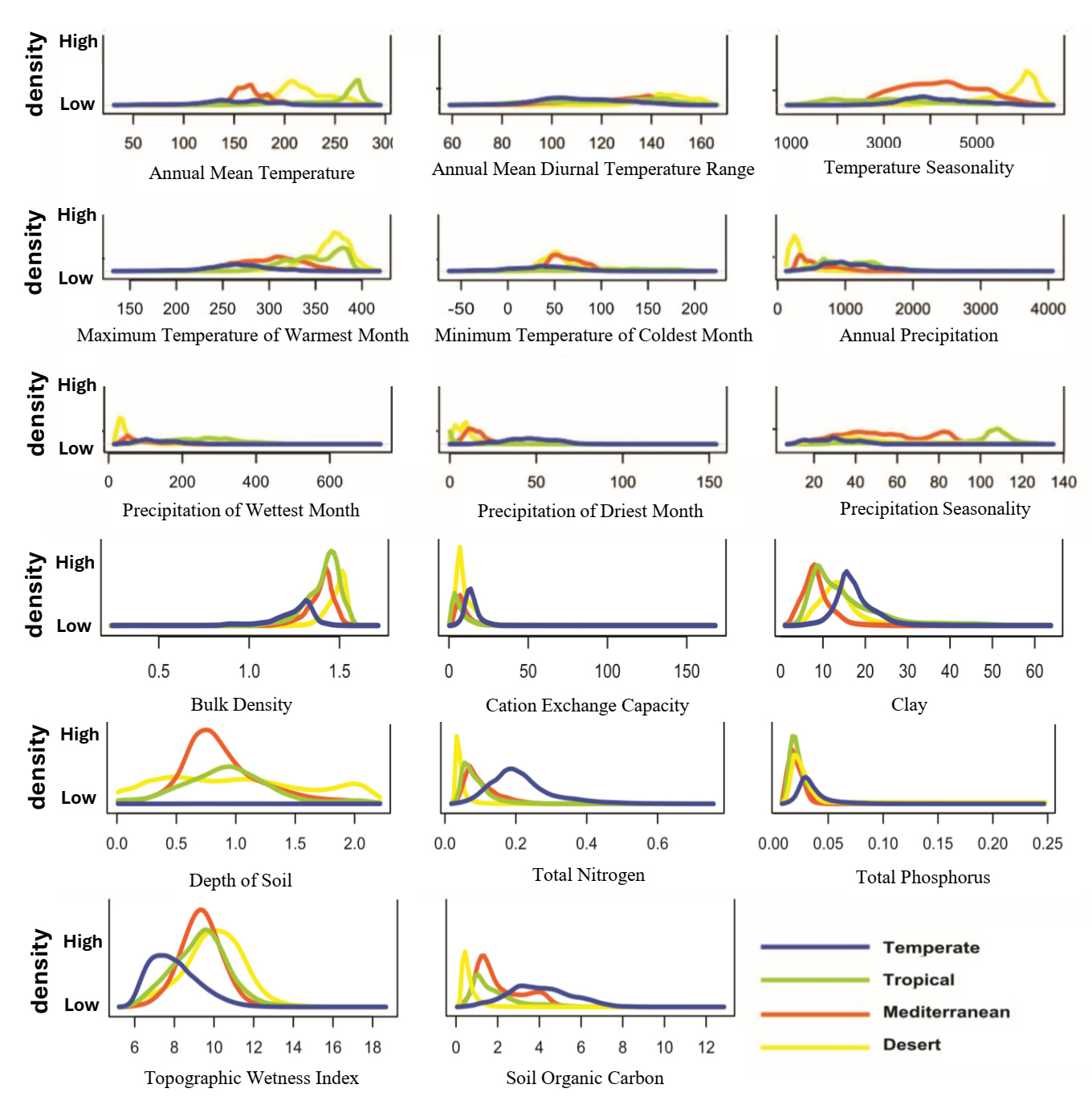


Figure 3. Density plots for different environmental variables across four different Australian ecoregions: Temperate, Tropical, Mediterranean, and Desert. In this study, we modelled 5,033 native species, and the lines are the average of all species. The lines represent the average response of all species modeled within each ecoregion against the extracted raster values for each environmental variable. The x-axis shows the range of each variable (see Table 1 for units, with temperature data presented in °C × 10 and precipitation data in millimeters), and the y-axis represents density, indicating the frequency of values within that range in each ecoregion. For individual density plots of each species per ecoregion, please refer to Figure S3.

## ANOVA analysis of variable importance across ecoregions

The importance of predictor variables differed significantly between ecoregions (Figure 4). Most climate and soil variables showed very strong differences in between-ecoregion comparisons. However, some variables did not show significant differences in certain ecoregion comparisons. Most importantly, the relative importance of Nitrogen Total Organic (NTO) was not significantly different between Mediterranean, Temperate, and Tropical ecoregions. Additionally, the importance of Bulk Density (BDW) and Precipitation of driest month (bio14) does not differ between some pairs, like Desert-Mediterranean and Desert-Tropical (Figure 4). These non-significant differences suggest that these variables have similar levels of importance in these ecoregions, pointing to overlapping environmental features or similar conditions for these factors across those regions. In this regard, Temperate-Tropical with three non-significant variables showed greatest similarity of environmental conditions.

The post-hoc tests also reveal varying significance levels across thresholds for certain environmental variables (Figure 4). For example, the importance of Precipitation seasonality (bio15) shows strong differences among ecoregion assignment thresholds (Figure 4). Similarly, Depth of Soil (DES) and Total Phosphorus (PTO) are highly significant across the 60%-100% comparisons, suggesting that the importance of these variables differ more when the difference between assignment thresholds are larger. In contrast, Soil Organic Carbon (SOC), Annual Mean Temperature (bio1), and Annual precipitation (bio12) show mostly non-significant (ns) results, indicating stable importance across thresholds and consistent influence in species distribution models.

|  |
| --- |
|  |
|  |

Figure 4. Top panel shows post-hoc analysis results; a comparison of environmental variables between every two ecoregions. Bottom panel shows post-hoc analysis results; a comparison of environmental variables between two thresholds. Cation Exchange Capacity (CEC), Depth of Soil (DES), Nitrogen Total Organic (NTO), Bulk Density (BDW), Soil Organic Carbon (SOC), Clay Content (CLY), Total Phosphorus (PTO), Topographic Wetness Index (TWI), and climate variables of Annual Mean Temperature (bio1), Annual mean diurnal temperature range (bio2), Temperature seasonality (bio4), Maximum temperature of warmest month (bio5), Minimum temperature of coldest month (bio6), Annual precipitation (bio12), Precipitation of wettest month (bio13), Precipitation of driest month (bio14), Precipitation seasonality (bio15).

**Discussion**

Ensuring the accuracy of predictive models for plant distributions involves careful selection and integration of both climate and soil variables, as these environmental factors are crucial for capturing the essential ecological requirements of plant species. Predictive accuracy and model realism are largely dependent on the choice of environmental predictor variables used in SDMs (Crimmins, et al. 2013, Mod, et al. 2016). The integration of SDMs into environmental management can enhance decision-making by identifying areas of high conservation value and predicting the impacts of environmental changes on species distributions.

***The effects of pedology on species distribution modelling of floras***

This study aimed to evaluate how environmental variables, both climatic and pedological (soil-related), influence species distribution patterns of terrestrial, vascular plants across Australian ecoregions (Desert, Tropical, Temperate, and Mediterranean), using a range of species distribution models and different occurrence density ecoregion assignment thresholds. Our study found that model performance, assessed using AUC and TSS metrics, was highest when climate and soil variables were integrated. To be more specific, models incorporating both climate and soil data exhibited marginally better performance than those utilizing climate data alone, whereas soil-only models yielded the lowest accuracy. This is consistent with findings suggesting that adding soil factors can improve model accuracy alongside climate data in complex ecoregions (Brenes, et al. 2008, Dubuis, Giovanettina, Pellissier, Pottier, Vittoz and Guisan 2013, Fitzpatrick, Gove, Sanders and Dunn 2008, Hageer, Esperón-Rodríguez, Baumgartner and Beaumont 2017, Martinson, Eddy, Commerford, Blevins, Rolfsmeier and McLauchlan 2011). For example, in another study on Australian shrub species, Hageer, Esperón-Rodríguez, Baumgartner and Beaumont (2017) demonstrated that utilizing soil variable enhance SDM performance. Similarly, Dubuis, Giovanettina, Pellissier, Pottier, Vittoz and Guisan (2013), studying 115 plant species in the Western Swiss Alps, showed that the inclusion of soil variables significantly improved model performance, particularly for species strongly associated with specific soil conditions. However, Zuquim, et al. (2014) and Figueiredo, et al. (2018), in studies with a similar spatial resolution to ours, reported contrasting findings for Amazonian plant species; they found that edaphic variables had the greatest contribution for many of the species they modeled.

In this study, we found that when combining climate and soil variables, climate variables had a more pronounced influence on plant distributions across all Australian ecoregions compared to soil variables (Table 3, Table S2). This finding underscores the pivotal role of geographic scale in determining the relative importance of environmental factors. Indeed, the dominance of climate variables as determinants of plant distribution at a continental scale aligns with the concept of hierarchical environmental filters (Soberón 2007, Williams, et al. 2009). Climate operates as a coarse-scale filter, shaping the broad patterns of species distributions by delineating the physiological limits of plants (Körner 2012). In contrast, soil characteristics, such as nutrient availability, pH, and texture, often act as finer-scale filters that mediate plant distributions within climatically suitable areas (John R, et al. 2007, Laliberté, et al. 2013). More importantly, Australia's diverse biogeography, with its expansive arid interior, tropical north, and temperate south, highlights climate as the primary driver of species distribution, likely overshadowing the localized effects of soil variables across its pronounced climatic gradients (Austin and Van Niel 2011).

While climate dominates at continental and global scales, the inclusion of soil variables enhances model performance by capturing additional nuances in species-environment relationships. For example, in tropical rainforests, fine-scale variations in soil properties are critical for explaining species diversity and distribution patterns (Paoli, et al. 2008, Zuquim, Tuomisto, Jones, Prado, Figueiredo, Moulatlet, Costa, Quesada and Emilio 2014). Additionally, soil variables may become increasingly important in determining plant distributions under climate change scenarios, as species may need to shift their ranges to adapt to changing environmental conditions (Figueiredo, Zuquim, Tuomisto, Moulatlet, Balslev and Costa 2018). Nevertheless, a potential obstacle could arise from the limited availability of environmental variables used as predictors in the models. For instance, soil temperature and soil water capacity, which are crucial factors affecting plant growth (Dunne and Willmott 1996, Reddell, et al. 1985), could be informative for SDMs, but the scarcity of data on these variables at a continental scale may pose a challenge for incorporating them into modelling studies.

***Modelling species distributions of floras across bioregions***

Species distribution patterns varied significantly across ecoregions (Figure S4), with the Temperate ecoregion consistently showing the highest species richness occurrences across all density thresholds, suggesting it supports a more diverse range of native angiosperm species compared to other ecoregions (Table 2). In contrast, the Desert ecoregion had the lowest species richness, reflecting its harsher environmental conditions. In our study, lower ecoregion assignment thresholds (e.g., 60%) allowed for the inclusion of species that span multiple ecoregions and therefore that likely had broader environmental tolerances (i.e., generalist species) (Ahmadi, Hemami, Kaboli and Shabani 2023, Stolar and Nielsen 2015). In contrast, higher thresholds restrict the dataset to species with narrower, more defined distributions, characteristic of ecological specialists (Lay, et al. 2010, McCune 2016).

The highest AUC and TSS values were observed at the 100% density threshold across all four ecoregions, regardless of whether soil, climate, or a combination of both variable types were used in the SDMs (Figure 2). Generalist species, capable of thriving across a range of environmental conditions often display lower predictive accuracy in SDMs due to their broad ecological niches, which could explain the lower performance of our models when lower ecoregion assignment thresholds were used (Connor, et al. 2017, McPherson and Jetz 2007). These species can inhabit a variety of habitats, making it difficult for models to pinpoint precise environmental drivers. In contrast, specialists with narrow ecological niches exhibit stronger associations with a more limited range of environmental factors which may help SDMs predict their distributions more accurately. Our findings highlight that the criteria used to include or exclude species when studying community responses to environmental variables is not merely a data-handling decision but a significant factor that can largely influence a study’s results, interpretation, bias and limitations.

More importantly, in a gradient of common to specialist species, our findings revealed a clear contrast between the Deserts and Tropical ecoregions regarding the significance of climate and soil variables; the importance of climatic variables increased in Tropical ecoregion, while in Deserts ecoregion the contribution of soil variables escalated. In tropical and temperate regions, where rainfall is more abundant and climate conditions are generally milder, the role of soil in shaping plant distribution, while remaining essential in all ecosystems, becomes less pronounced (Brenes, Paredes, Rivas-Torres, Blundo, Coley and Kursar 2008). In these enriched and ecologically diverse regions, other limiting factors such as temperature, sunlight, and competition with other species may play a more significant role in shaping plant distribution (Baltzer, et al. 2008). Moreover, in Tropical regions higher rainfall results in increased leaching or lower mineralization rates, which decreases nutrient availability in the soils (Kurniawan, et al. 2018, Posada and Schuur 2011).

On the contrary, in arid ecosystems, where water availability is limited and the climate is harsh, soil characteristics become crucial determinants of plant survival and distribution (Gamalero, et al. 2020, Martirosyan, et al. 2016). The type and quality of soil directly influence water retention and drainage, which are critical for plant survival (Nielsen and Ball 2015). In line with this, Bui, et al. (2014) discovered that in Australia while climate plays a more significant role in influencing the distribution of Acacia species at a continental scale, the physical and chemical characteristics of soil proved to be more informative in explaining the distribution patterns of shrub species in xeric ecosystems of southern Australia. Altogether, understanding the specific mechanisms driving plant distribution at different spatial scales and various ecoregions are still challenging. In addition to the influence of the environment on species distributions, other factors such as dispersal limitations and neutral processes can also play an important role (Chust, et al. 2006, Franklin 2010).

The post-hoc analysis (Figure 4) shows significant differences in the importance of climate (e.g., Annual Mean Temperature (bio1), Annual precipitation (bio12), Temperature seasonality (bio4)), and soil variables (e.g., Cation Exchange Capacity (CEC), Soil Organic Carbon (SOC), Clay content (CLY)) to species distribution across ecoregions, indicating that species respond differently to environmental variables depending on which ecoregion they come from. These findings suggest that species distribution models should be customized to each ecoregion, as key environmental drivers vary regionally. Additionally, analysis across thresholds reveals that the importance of some variables, such as Precipitation seasonality (bio15), Depth of Soil (DES), and Total Phosphorus (PTO), change depending on the ecoregion assignment thresholds used (e.g., 60-100%), suggesting importance of variables differ between the generalist species included when a low threshold is used and the specialist species modelling is restricted to when a high threshold is used. In contrast, variables like Annual mean diurnal temperature range (bio2) and Nitrogen Total Organic (NTO) remain stable, indicating they are reliable predictors across ecoregion assignment thresholds. This distinction suggests that while certain factors vary with assignment thresholds, others consistently enhance model robustness across ecological contexts.

**Conclusion**

Our findings underscore the potential of SDMs to inform adaptive management practices, particularly in ecoregions where species distributions are strongly influenced by both climate and soil variables. By tailoring environmental management strategies to the specific drivers of species distributions in each ecoregion, SDMs can contribute to the sustainable use of natural resources and the preservation of biodiversity under changing environmental conditions. Our study concludes that climate variables are the most important factors across all ecoregions and thresholds. Further, adding soil variables often improves SDM but only soil is a weak predictor compared to climate only or climate + soil. Our study revealed that, among the various model types, MaxEnt stands out as one of the most effective models employed in this study. Overall, the threshold of 100 yielded the highest SDM performance, while the lowest performance was observed at a threshold of 60. This finding highlights the importance of tailoring ecological models to each ecoregion's environmental dynamics.

**Recommendations**

Future research should investigate the relationship between species' range sizes and the importance of different environmental variables. Species with narrow distributions, such as rare, endemic, or threatened species, may be disproportionately influenced by specific environmental factors, whereas widespread species may respond to a shared suite of variables. Understanding this relationship could yield valuable insights for conservation biology by identifying which environmental factors are most critical for the persistence of species with restricted ranges. While our study illustrates the complex interplay between climate and soil variables in shaping species distributions, other influential factors—such as dispersal limitations and neutral processes, including random dispersal, genetic drift, and stochastic demographic events—also play important roles (Chust, Chave, Condit, Aguilar, Lao and Pérez 2006, Franklin 2010). Additionally, correlated environmental variables complicate interpretations of model output (Brenes, Paredes, Rivas-Torres, Blundo, Coley and Kursar 2008, Davidar, et al. 2007). We recommend that future studies explore the potential benefits of incorporating additional variables, such as topographic and land use factors, to further enhance model accuracy and better capture the mechanisms driving plant distributions across spatial scales and ecoregions.

**Data Availability**: All codes will be freely available.

**References**

Adde, A., et al. 2023. Too many candidates: Embedded covariate selection procedure for species distribution modelling with the covsel R package. - Ecological Informatics 75: 102080.

Aerts, R. and Chapin, F. S. I. 1999. The mineral nutrition of wild plants revisited: a re-evaluation of processes and patterns. - In: Advances in ecological research. Elsevier, pp. 1-67.

Ahmadi, M., et al. 2023. MaxEnt brings comparable results when the input data are being completed; Model parameterization of four species distribution models. - Ecology and Evolution 13: e9827.

Anderson, R. P., et al. 2003. Evaluating predictive models of species’ distributions: criteria for selecting optimal models. - Ecological Modelling 162: 211-232.

Araújo, M. B. and Guisan, A. 2006. Five (or so) challenges for species distribution modelling. - Journal of Biogeography 33: 1677-1688.

Araújo, M. B. and New, M. 2007. Ensemble forecasting of species distributions. - Trends in Ecology & Evolution 22: 42-47.

Arenas‐Castro, S., et al. 2022. Effects of input data sources on species distribution model predictions across species with different distributional ranges. - Journal of Biogeography 49: 1299-1312.

Austin, M. P. and Meyers, J. A. 1996. Current approaches to modelling the environmental niche of eucalypts: implication for management of forest biodiversity. - Forest Ecology and Management 85: 95-106.

Austin, M. P. and Van Niel, K. P. 2011. Improving species distribution models for climate change studies: variable selection and scale. - Journal of Biogeography 38: 1-8.

Australia;, E. F. o. S. 2024. Electronic Flora of South Australia. - In.

Australian Native Plants Society; 2024 Australian Native Plants Society, Available at <https://anpsa.org.au/>. - In, p. April

Baltzer, J. L., et al. 2008. The role of desiccation tolerance in determining tree species distributions along the Malay–Thai Peninsula. - Functional Ecology 22: 221-231.

Beck, J., et al. 2014. Spatial bias in the GBIF database and its effect on modelling species' geographic distributions. - Ecological Informatics 19: 10-15.

Berg, G. and Smalla, K. 2009. Plant species and soil type cooperatively shape the structure and function of microbial communities in the rhizosphere. - FEMS Microbiology Ecology 68: 1-13.

Brenes, T., et al. 2008. The effect of soil on the growth performance of tropical species with contrasting distributions. - Oikos 117: 1453-1460.

Broadhurst, L. and Coates, D. 2017. Plant conservation in Australia: Current directions and future challenges. - Plant Diversity 39: 348-356.

Brown, C. J., et al. 2016. Ecological and methodological drivers of species’ distribution and phenology responses to climate change. - Global Change Biology 22: 1548-1560.

Bui, E. N., et al. 2014. Acacia, climate, and geochemistry in Australia. - Plant and Soil 381: 161-175.

Castellanos, A. A., et al. 2019. Environmental filtering improves ecological niche models across multiple scales. - Methods in Ecology and Evolution 10: 481-492.

CHELSA – Free climate data at high resolution, available at <https://chelsa-climate.org/>. - In.

Chust, G., et al. 2006. Determinants and spatial modelling of tree β‐diversity in a tropical forest landscape in Panama. - Journal of Vegetation Science 17: 83-92.

Connolly, S. R., et al. 2017. Process, Mechanism, and Modelling in Macroecology. - Trends in Ecology & Evolution 32: 835-844.

Connor, T., et al. 2017. Effects of grain size and niche breadth on species distribution modelling. - Ecography 41: 1270–1282.

Crimmins, S. M., et al. 2013. Evaluating ensemble forecasts of plant species distributions under climate change. - Ecological Modelling 266: 126-130.

Datta, A., et al. 2020. Origin of climatic data can determine the transferability of species distribution models. - NeoBiota 59: 61-76.

Davidar, P., et al. 2007. The effect of climatic gradients, topographic variation and species traits on the beta diversity of rain forest trees. - Global Ecology and Biogeography 16: 510-518.

Delgado-Baquerizo, M., et al. 2020. The influence of soil age on ecosystem structure and function across biomes. - Nature Communications 11: 4721.

Dormann, C. F., et al. 2012. Correlation and process in species distribution models: bridging a dichotomy. - Journal of Biogeography 39: 2119-2131.

Dubuis, A., et al. 2013. Improving the prediction of plant species distribution and community composition by adding edaphic to topo‐climatic variables. - Journal of Vegetation Science 24: 593-606.

Dunne, K. A. and Willmott, C. J. 1996. Global distribution of plant-extractable water capacity of soil. - International Journal of Climatology 16: 841-859.

Duursma, D. E., et al. 2013. Next-generation invaders? Hotspots for naturalised sleeper weeds in Australia under future climates. - PLoS One 8: e84222.

Edwards, M. and Richardson, A. J. 2004. Impact of climate change on marine pelagic phenology and trophic mismatch. - Nature 430: 881-884.

Elith, J., et al. 2006. Novel methods improve prediction of species’ distributions from occurrence data. - Ecography 29: 129-151.

Elith, J. and Leathwick, J. 2009. Species distribution models: ecological explanation and prediction across space and time. - Annual review of Ecology, Evolution, and Systematics 40:

Elith, J., et al. 2010. The art of modelling range‐shifting species. - Methods in ecology and evolution 1: 330-342.

Figueiredo, F. O. G., et al. 2018. Beyond climate control on species range: The importance of soil data to predict distribution of Amazonian plant species. - Journal of Biogeography 45: 190-200.

Fitzpatrick, M. C., et al. 2008. Climate change, plant migration, and range collapse in a global biodiversity hotspot: the Banksia (Proteaceae) of Western Australia. - Global Change Biology 14: 1337-1352.

Fourcade, Y., et al. 2018. Paintings predict the distribution of species, or the challenge of selecting environmental predictors and evaluation statistics. - Global Ecology and Biogeography 27: 245-256.

Franklin, J. 2010. Mapping Species Distributions: Spatial Inference and Prediction. - Cambridge University Press.

Gamalero, E., et al. 2020. Saline and Arid Soils: Impact on Bacteria, Plants, and their Interaction. - Biology 9: 116.

Global Biodiversity Information Facility 2024. Global Biodiversity Information Facility (GBIF), Available at: <http://www.gbif.org>. - In.

Grundy, M. J., et al. 2015. Soil and landscape grid of Australia. - Soil Research 53: 835-844.

Guisan, A. and Zimmermann, N. E. 2000. Predictive habitat distribution models in ecology. - Ecological Modelling 135: 147-186.

Hageer, Y., et al. 2017. Climate, soil or both? Which variables are better predictors of the distributions of Australian shrub species? - PeerJ 5: e3446.

Hazen, E. L., et al. 2013. Predicted habitat shifts of Pacific top predators in a changing climate. - Nature Climate Change 3: 234-238.

Hijmans, R., et al. 2004. The WorldClim interpolated global terrestrial climate surfaces. - In.

Hulshof, C. M. and Spasojevic, M. J. 2020. The edaphic control of plant diversity. - Global Ecology and Biogeography 29: 1634-1650.

Iverson, L. R., et al. 2008. Estimating potential habitat for 134 eastern US tree species under six climate scenarios. - Forest Ecology and Management 254: 390-406.

Jarvis, A. M. and Robertson, A. 1999. Predicting population sizes and priority conservation areas for 10 endemic Namibian bird species. - Biological Conservation 88: 121-131.

Jin, Y. and Qian, H. 2022. V.PhyloMaker2: An updated and enlarged R package that can generate very large phylogenies for vascular plants. - Plant Divers 44: 335-339.

John R, et al. 2007. Soil nutrients influence spatial distributions of tropical tree species. - Proceedings of the National Academy of Sciences of the United States of America 104: 864-9.

Karger, D. N., et al. 2017. Climatologies at high resolution for the earth’s land surface areas. - Scientific Data 4: 170122.

Kearney, M. and Porter, W. 2009. Mechanistic niche modelling: combining physiological and spatial data to predict species’ ranges. - Ecology letters 12: 334-350.

Kearney, M., et al. 2010. Correlative and mechanistic models of species distribution provide congruent forecasts under climate change. - Conservation Letters 3: 203-213.

Keith, D. A., et al. 2008. Predicting extinction risks under climate change: coupling stochastic population models with dynamic bioclimatic habitat models. - Biology Letters 4: 560-563.

Körner, C. 2012. Alpine treelines: functional ecology of the global high elevation tree limits. - Springer.

Kramer-Schadt, S., et al. 2013. The importance of correcting for sampling bias in MaxEnt species distribution models. - Diversity and Distributions 19: 1366-1379.

Kurniawan, S., et al. 2018. Conversion of tropical forests to smallholder rubber and oil palm plantations impacts nutrient leaching losses and nutrient retention efficiency in highly weathered soils. - Biogeosciences 15: 5131-5154.

Laliberté, E., et al. 2013. Contrasting effects of productivity and disturbance on plant functional diversity at local and metacommunity scale. - Journal of Vegetation Science 24: 834-842.

Lay, G. L., et al. 2010. Prospective sampling based on model ensembles improves the detection of rare species. - Ecography 33: 1015-1027.

Luoto, M., et al. 2006. Determinants of the biogeographical distribution of butterflies in boreal regions. - Journal of Biogeography 33: 1764-1778.

Martinson, E., et al. 2011. Biogeographic distributions of selected North American grassland plant species. - Physical Geography 32: 583-602.

Martirosyan, V., et al. 2016. Desert Perennial Shrubs Shape the Microbial-Community Miscellany in Laimosphere and Phyllosphere Space. - Microbial Ecology 72: .

Martre, P., et al. 2002. Root deployment and shoot growth for two desert species in response to soil rockiness. - American Journal of Botany 89: 1933-1939.

McBratney, A. B., et al. 2003. On digital soil mapping. - Geoderma 117: 3-52.

McCune, J. L. 2016. Species distribution models predict rare species occurrences despite significant effects of landscape context. - Journal of Applied Ecology 53: 1871-1879.

McPherson, J. M. and Jetz, W. 2007. Effects of species’ ecology on the accuracy of distribution models. - Ecography 30: 135-151.

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

Mod, H. K., et al. 2016. What we use is not what we know: environmental predictors in plant distribution models. - Journal of Vegetation Science 27: 1308-1322.

Muhling, B. A., et al. 2017. Projections of future habitat use by Atlantic bluefin tuna: mechanistic vs. correlative distribution models. - ICES Journal of Marine Science 74: 698-716.

Naimi, B. 2015. Usdm: Uncertainty analysis for species distribution models. 1: 1-12.

Nielsen, U. N. and Ball, B. A. 2015. Impacts of altered precipitation regimes on soil communities and biogeochemistry in arid and semi-arid ecosystems. - Global Change Biology 21: 1407-21.

Olson, D. M., et al. 2001. Terrestrial Ecoregions of the World: A New Map of Life on Earth: A new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. - BioScience 51: 933-938.

Paoli, G. D., et al. 2008. Soil nutrients affect spatial patterns of aboveground biomass and emergent tree density in southwestern Borneo. - Oecologia 155: 287-299.

Posada, J. M. and Schuur, E. A. 2011. Relationships among precipitation regime, nutrient availability, and carbon turnover in tropical rain forests. - Oecologia 165: 783-95.

Ramirez‐Cabral, N. Y. Z., et al. 2017. Global risk levels for corn rusts (Puccinia sorghi and Puccinia polysora) under climate change projections. - Journal of Phytopathology 165: 563-574.

Ramos, R. S., et al. 2019. Risk of spread of tomato yellow leaf curl virus (TYLCV) in tomato crops under various climate change scenarios. - Agricultural Systems 173: 524-535.

Randin, C. F., et al. 2009. Climate change and plant distribution: local models predict high‐elevation persistence. - Global Change Biology 15: 1557-1569.

Reddell, P., et al. 1985. The effects of soil temperature on plant growth, nodulation and nitrogen fixation in casuarina cunninghamiana miq -New Phytologist 101: 441-450.

Renner, I. W. and Warton, D. I. 2013. Equivalence of MAXENT and Poisson point process models for species distribution modelling in ecology. - Biometrics 69: 274-281.

Robertson, M. P., et al. 2003. Comparing models for predicting species’ potential distributions: a case study using correlative and mechanistic predictive modelling techniques. - Ecological Modelling 164: 153-167.

Robinson, N. M., et al. 2017. A systematic review of marine-based species distribution models (SDMs) with recommendations for best practice. - Frontiers in Marine Science 4: 421.

Searle, R., et al. 2022. Tern digital soil mapping raster covariate stacks -In: CSIRO.

Sequeira, A., et al. 2012. Ocean‐scale prediction of whale shark distribution. - Diversity and Distributions 18: 504-518.

Shabani, F. and Kumar, L. 2014. Sensitivity Analysis of CLIMEX Parameters in Modelling Potential Distribution of *Phoenix dactylifera* L. - PLoS ONE 9: e94867.

Shabani, F., et al. 2016. A comparison of absolute performance of different correlative and mechanistic species distribution models in an independent area. - Ecology and Evolution 6: 5973-5986.

Shabani, F., et al. 2017. Climate modelling shows increased risk to *Eucalyptus sideroxylon* on the eastern coast of Australia compared to *Eucalyptus albens*. - Plants 6: 58.

Shabani, F., et al. 2018. Assessing accuracy methods of species distribution models: auc, specificity, sensitivity and the true skill statistic. - Global Journal of Human-Social Science Research 18: 7\_18.

Shabani, F., et al. 2019. Climate‐driven shifts in the distribution of koala browse species from the Last Interglacial to the near future. - Ecography 42: 1587-1599.

Soberón, J. 2007. Grinnellian and Eltonian niches and geographic distributions of species. - Ecology Letters 10: 1115-1123.

Stauffer, D. 2002. Linking populations and habitats: Where have we been? Where are we going? - Predicting Species Occurrences : Issues of Accuracy and Scale

Stolar, J. and Nielsen, S. E. 2015. Accounting for spatially biased sampling effort in presence-only species distribution modelling. - Diversity and Distributions 21: 595-608.

Svenning, J. C., et al. 2008. Glacial refugia of temperate trees in Europe: insights from species distribution modelling. - Journal of Ecology 96: 1117-1127.

Thuiller, W., et al. 2009. BIOMOD–a platform for ensemble forecasting of species distributions. - Ecography 32: 369-373.

Valavi, R., et al. 2022. Predictive performance of presence‐only species distribution models: a benchmark study with reproducible code. - Ecological Monographs 92: e01486.

Varela, S., et al. 2014. Environmental filters reduce the effects of sampling bias and improve predictions of ecological niche models. - Ecography 37: 1084-1091.

Veech, J. A. 2021. Habitat ecology and analysis. - Oxford University Press, USA.

Velazco, S. J. E., et al. 2017. Using worldwide edaphic data to model plant species niches: An assessment at a continental extent. - PLoS One 12: e0186025.

Wang, H., et al. 2016. Study on selecting sensitive environmental variables in modelling species spatial distribution. - Annals of GIS 22: 57-69.

Warren, D. L. and Seifert, S. N. 2011. Ecological niche modelling in Maxent: the importance of model complexity and the performance of model selection criteria. - Ecological applications 21: 335-342.

Wauchope‐Drumm, M., et al. 2020. Using a species distribution model to guide NSW surveys of the long‐footed potoroo (Potorous longipes). - Austral Ecology 45: 15-26.

Williams, J. N., et al. 2009. Using species distribution models to predict new occurrences for rare plants. - Diversity and Distributions 15: 565-576.

Williams, S. and Hero, J. 2001. Multiple determinants of Australian tropical frog biodiversity. - Biological Conservation 98: 1-10.

Zizka, A., et al. 2019. CoordinateCleaner: Standardized cleaning of occurrence records from biological collection databases. - Methods in Ecology and Evolution 10: 744-751.

Zuquim, G., et al. 2014. Predicting environmental gradients with fern species composition in Brazilian Amazonia. - Journal of Vegetation Science 25: 1195-1207.