

2011

## **An evaluation of environmental factors affecting species distributions**

Michael B. Ashcroft

*University of Wollongong*, [ashcroft@uow.edu.au](mailto:ashcroft@uow.edu.au)

Kristine O. French

*University of Wollongong*, [kris@uow.edu.au](mailto:kris@uow.edu.au)

Laurie A. Chisholm

*University of Wollongong*, [lauriec@uow.edu.au](mailto:lauriec@uow.edu.au)

Follow this and additional works at: <https://ro.uow.edu.au/scipapers>



Part of the [Life Sciences Commons](#), [Physical Sciences and Mathematics Commons](#), and the [Social and Behavioral Sciences Commons](#)

---

### **Recommended Citation**

Ashcroft, Michael B.; French, Kristine O.; and Chisholm, Laurie A.: An evaluation of environmental factors affecting species distributions 2011.  
<https://ro.uow.edu.au/scipapers/5229>

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: [research-pubs@uow.edu.au](mailto:research-pubs@uow.edu.au)

---

## An evaluation of environmental factors affecting species distributions

### Abstract

Many different models can be built to explain the distributions of species. Often there is no single model that is clearly better than the alternatives, and this leads to uncertainty over which environmental factors are limiting species' distributions. We investigated the support for different environmental factors by determining the drop in model performance when selected predictors were excluded from the model building process. We used a paired t-test over 37 plant species so that an environmental factor was only deemed significant if it consistently improved the results for multiple species. Geology and winter minimum temperatures were found to be the environmental factors with the most support, with a significant drop in model performance when either of these factors was excluded. However, there was less support for summer maximum temperature, as other environmental factors could combine to produce similar model performance. Our method of evaluating environmental factors using multiple species will not be capable of detecting predictors that are only important for one or two species, but it is difficult to distinguish these from spurious correlations. The strength of the method is that it increases inference for factors that consistently affect the distributions of many species. We discourage the assessment of models against predefined benchmarks, such as an area under the curve (AUC) of more than 0.7, as many alternative models for the same species produce similar results. Therefore, the benchmarks do not provide any indication of how the performance of the selected model compares to alternative models, and they provide weak inference to accept any selected model.

### Disciplines

Life Sciences | Physical Sciences and Mathematics | Social and Behavioral Sciences

### Publication Details

Ashcroft, M. B., French, K. O. & Chisholm, L. A. (2011). An evaluation of environmental factors affecting species distributions. *Ecological Modelling*, 222 (3), 524-531.

Article type: Original research paper

## **An evaluation of environmental factors affecting species distributions**

Michael B Ashcroft<sup>a,c,\*</sup>

Kristine O French<sup>b</sup>

Laurie A Chisholm<sup>b</sup>

<sup>a</sup> School of Earth and Environmental Sciences, University of Wollongong,  
Wollongong, NSW, Australia, 2522.

<sup>b</sup> Institute for Conservation Biology and Environmental Management, University of  
Wollongong, Wollongong, NSW, Australia, 2522.

<sup>c</sup> Australian Museum, 6 College Street, Sydney, NSW, Australia, 2010.

\* Corresponding author. Tel.: +61 2 9320 6475; fax: +61 2 9320 6021; E-mail:

Mick.Ashcroft@austmus.gov.au

**Abstract** Many different models can be built to explain the distributions of species. Often there is no single model that is clearly better than the alternatives, and this leads to uncertainty over which environmental factors are limiting species' distributions. We investigated the support for different environmental factors by determining the drop in model performance when selected predictors were excluded from the model building process. We used a paired t-test over 37 plant species so that an environmental factor was only deemed significant if it consistently improved the results for multiple species. Geology and winter minimum temperatures were found to be the environmental factors with the most support, with a significant drop in model performance when either of these factors was excluded. However, there was less support for summer maximum temperature, as other environmental factors could combine to produce similar model performance. Our method of evaluating environmental factors using multiple species will not be capable of detecting predictors that are only important for one or two species, but it is difficult to distinguish these from spurious correlations. The strength of the method is that it increases inference for factors that consistently affect the distributions of many species. We discourage the assessment of models against predefined benchmarks, such as an Area Under the Curve (AUC) of more than 0.7, as many alternative models for the same species produce similar results. Therefore, the benchmarks do not provide any indication of how the performance of the selected model compares to alternative models, and they provide weak inference to accept any selected model.

*Keywords:* Area under curve; Ecological niche models; Model evaluation; Multivariate statistical analysis; Predictor selection; Species distribution models

## 1. Introduction

Identifying the factors that affect species' distributions is an important unresolved issue in ecology (Araújo and Guisan, 2006). Often there are many combinations of predictors that can explain distributions equally well, especially when environmental factors are correlated, and this introduces uncertainty over the effect of each factor (Freedman, 1983; Graham, 2003; Whittingham et al., 2006; Platts et al., 2008; Murray and Conner, 2009). Further doubts arise because models are typically based only on correlations, and may not reflect causal relationships when both species occurrence records and environmental factors are spatially autocorrelated (Bahn and McGill, 2007; Currie, 2007; Beale et al, 2008). Finally, while it is widely acknowledged that a separate dataset should be used to evaluate model performance, generally it is only the selected model that is evaluated, and therefore this process does not prevent an overfitted or spurious model being selected.

Due to these uncertainties, some form of corroboration assessment (Faith, 2003; Faith et al., 2004) is needed to compare alternative models and examine the weight of evidence supporting the inclusion of different factors in models (Stephens et al., 2007). One such method is hierarchical partitioning (Mac Nally, 2002), which estimates the effect of each environmental factor by averaging its effect in all candidate models in which it occurs. Importantly, this method is not designed to identify one 'true' model, but rather to provide evidence on which factors are most likely to have causal effects. However, while hierarchical partitioning offers a method to address ambiguities between correlated predictors, the method is limited because all models are produced using the same data. If there is a spurious correlation between a species distribution and an environmental factor, due to spatial structure for example, then that variable may appear important in all candidate models that use that same data. Therefore, while

hierarchical partitioning may help resolve ambiguities due to correlated predictors, it cannot help avoid spurious correlations between a species distribution and an environmental factor. These issues also apply to other methods that increase inference by using multiple alternative models but the same data, such as information-theoretic approaches based on as Akaike's information criterion (Anderson et al., 2000).

In this article we introduce a new methodology that is designed to examine the strength of evidence that an environmental factor affects species' distributions. There are two key components of this model building processes. First, all candidate models are assessed using the evaluation dataset to provide better protection against overfitting. If all candidate models are assessed using the evaluation dataset, then overfitted models are penalised during the model building process and we can avoid selecting overfitted models.

The second component is that the performance of an environmental factor is assessed using models for multiple species. Environmental factors that are only important for one species are unlikely to be detected using this approach, but it is difficult to distinguish these from spurious correlations. Environmental factors that are strongly correlated with multiple species are less likely to be spurious correlations, as different distributional data are used in each of the models. Hence, the method avoids the problems associated with hierarchical partitioning and information-theoretic approaches, which increase inference using multiple models but are confounded because they use the same data used to test each combination of predictors.

Our method involves using a paired t-test with multiple species to determine whether there is a consistent drop in evaluation AUCs when selected predictors are excluded from models. If AUCs are lower when a predictor is excluded then this would suggest that that the other predictors were not able to explain the species' distributions

as well without this predictor. Alternatively, if AUCs are higher without a predictor, then this would suggest its inclusion led to overfitting to the training data. There would be some statistical noise introduced whenever a different combination of predictors is selected, so the t-test is effectively testing whether each predictor had a consistent overfitting or explanatory effect over all species. While there may be spurious correlations between some environmental factors and species, and these would impact all models using this same data, these would not influence the results of the t-test unless the factor was correlated with multiple species.

Other studies have also looked at drop contributions (e.g. Lehmann et al., 2003), and these have been incorporated into tools such as Maxent (Phillips et al., 2006). However, these previous applications are only with individual species, with no statistical test to confirm whether a drop in performance is significant. Other studies have also made generalisations about which environmental factors are most important over multiple models (e.g. Peterson and Cohoon, 1999; Watson and Peterson, 1999), however these did not include a statistical test of this consistency, and were limited to assessing the predictor that was clearly the most important in each case. Our method has stronger inference due to the statistical test, and can be used to assess all predictors. In addition, while the previously mentioned studies (Peterson and Cohoon, 1999; Watson and Peterson, 1999) implied that one factor was important in all models they assessed, this is not necessary for our method to work. The t-tests can be significant even if some species are unaffected by a variables exclusion. This makes intuitive sense, as we expect different species to be affected by different environmental factors.

We applied the method to provide a more robust examination of three environmental factors that appeared to be key determinants of the distribution of vegetation on the Illawarra Escarpment. Geology, summer maximum temperature, and

winter minimum temperature all have strong correlations with the distribution of multiple species (Ashcroft et al., 2008) but have not been comprehensively evaluated in a multivariate context. Here, the three predictors were examined to determine if they made a unique contribution to the performance of SDMs, or whether other predictors could combine to produce alternative, and potentially equally plausible hypotheses for the distribution of species.

We also evaluated whether species distributions were better correlated with past or present temperatures. The spatial distribution of warming was not uniform between 1972 and 2006 (Ashcroft et al., 2009), and species that had not adjusted their distribution may have been better correlated with past temperatures than current temperatures. Alternatively, if species had adjusted their distribution then we expected current temperatures to be significantly better than past temperatures. SDMs assume that species are in equilibrium with their environment, so we were effectively testing whether or not this was the case. Roubicek et al. (2010) recently suggested that for short-lived species it is important to match the time period of the climate data with the collection of biotic data. However, it is not clear which time period should be used for longer-lived species that may or may not have adjusted their distributions in response to the warming trend over the last 40 years.

## **2. Methods**

### *2.1. Study area*

This study was conducted on approximately 12000 ha of the Illawarra Escarpment and Woronora Plateau, 80 km south of Sydney, Australia (34.4°S, 150.9°E; Fig. 1). The



escarpment runs northeast to southwest through the study area, and separates the Woronora Plateau in the northwest from the city of Wollongong on the coastal plain in the southeast. The escarpment slopes and gullies on the Woronora Plateau contain a complex mosaic of moist and dry rainforests, moist eucalypt (*Eucalyptus* spp.) forests, and tall open eucalypt forests, while the Hawkesbury sandstone ridges and mountaintops support upland swamps and eucalypt woodlands (NPWS, 2002). The foothills and coastal plain are largely disturbed by urban development, but there are many semi-natural areas in parklands and along creek lines.

The geology of the study area consists of approximately horizontal layers, with Hawkesbury sandstone at the highest elevations, and interspersed layers of sandstones, claystones and coal seams on the escarpment slopes. The gullies on the Woronora Plateau are predominately on the uppermost two of these units—Bald Hill claystone and Bulgo sandstone.

## *2.2. Environmental predictors*

A categorical geology layer (Moffit, 1999) was obtained courtesy of the NSW Department of Primary Industries and contained spatial errors of up to 150 m in the locations of boundaries. We developed a range of fine-scale temperature surfaces for the study area, including the minimum and maximum temperatures for spring (September–November), autumn (March–May), summer (December–February), and winter (June–August), as well as the average annual maximums and minimums. The surfaces were developed by recording temperatures using small microclimatic temperature loggers (DS1921G iButtons; Maxim) at 40 sites for a period of 21 months from December 2004 to August 2006 (Ashcroft et al., 2008). Temperature grids were

produced using linear regression and a variety of climate forcing factors, as commonly used elevation sensitive interpolations such as BioClim and WorldClim are not appropriate at finer resolutions (Daly, 2006). The factors considered were elevation, radiation, exposure to winds from various directions, and proximity to streams and coastlines. The data was originally collected and analysed in three-week periods (Ashcroft et al., 2008), however all surfaces for each season were later averaged to produce the seasonal temperature surfaces (referred to hereafter as 2005 temperature surfaces). Many seasonal temperature grids displayed patterns similar to that of elevation, however winter minimum temperatures were determined mostly by distance to coast, and summer maximum temperatures were associated mostly with shelter from hot, dry northwesterly winds. These extreme temperatures were poorly correlated with elevation, and had better ability to explain the distribution of vegetation in univariate models than other seasonal temperatures (Ashcroft et al., 2008). Estimated winter minimum temperatures ranged from 7.2°C to 10.6°C, and summer maximums from 18.1°C to 23.9°C, although these reflect soil surface temperatures in forests rather than air temperatures inside Stevenson screens 1.5m above cleared areas.

The spatial distribution of climate change between 1972 and 2006 was estimated by establishing the relationships between Bureau of Meteorology weather station records and the iButton observations. Results were used to estimate the trend in warming in each grid cell during this period (Ashcroft et al., 2009). In general, inland sites warmed more than coastal sites, and there was more warming at sites that were exposed to hot-dry northwesterly winds, away from streams, or at lower elevations. We subtracted the amount of warming in each grid cell from the 2005 temperature surfaces to create the 1972 temperature surfaces. As the amount of warming varied spatially, the 1972 surfaces displayed slightly different patterns to the 2005 surfaces. In all, there

were 20 temperature surfaces used—minimum and maximum temperatures for spring, summer, autumn, winter and annual periods for both 1972 and 2005.

It is important to note that while we refer to the climate grids as either 1972 or 2005 in this article, the difference between the two is based on the trend in climate over this period rather than the actual climate in the two individual years.

### *2.3. Vegetation data*

A vegetation survey of 600 sites (Fig. 1) was conducted between July 2005 and March 2006. Sites were purposively selected for modelling (see Hirzel and Guisan, 2002) and covered a broad range of environmental conditions. For each of the 21 communities that were common in the study area (NPWS, 2002), sites were randomly selected from a list of potential locations subject to a number of constraints. First, no sites from the same community could be close together, and were distributed among different patches where possible. Secondly, within each community, sites were selected such that they covered a broad elevational range and all geologies on which the community was commonly found. No subjective bias was introduced by searching for pristine or homogeneous sites, and there was no bias towards or away from roads (to eliminate edge effects or gain easy access).

The study area has been subjected to disturbance in the past, mostly from farming, utility easements and mining, but the majority of the escarpment and Woronora Plateau has been protected land for more than 20 years. While some sites showed signs of past disturbance, these were generally in a state of regeneration and contained similar species to more natural areas nearby, albeit with lower abundance and more

interspersed weeds. Sites that were being actively managed without opportunity for natural revegetation were not considered.

All sites were 20 m by 20 m and were surveyed for presence or absence of 37 dominant canopy or sub-canopy species (Table 1). Species were selected based on their expected frequencies in the vegetation communities present in the region (NPWS, 2002), with the sample size of 600 selected so that we would obtain approximately 50 presences (Stockwell and Peterson, 2002) for most of the dominant eucalypt species. This proved effective, and only 4 of the 37 species had less than 40 presences (mean = 126, SD = 94).

Ten of the selected species were eucalypts, which are hard-leaved evergreen trees. Their distributions vary widely, with some species dominant on the Hawkesbury sandstone, some on the escarpment, and some in the plateau gullies. It is not uncommon for three or four eucalypt species to coexist within sites. The eucalypts are typically found in the tallest strata, with moist rainforest species in the understory in sheltered sites, or drier species in the understory in more exposed or fire-prone locations.

The majority of species (21 of 37) were rainforest species, which are typically mesic evergreen species (*Toona ciliata* is the only deciduous species). The rainforest species also had varying distributions, with some species only found in the moistest locations (sheltered and at high elevation), while the lowland dry sub-tropical rainforest species were restricted to the foothills of the escarpment (NPWS, 2002).

There was no a-priori reason to believe all 37 species were all limited by the same environmental factors, and indeed previous research suggested that some species were better explained by winter minimum temperatures, some by summer maximums, and some by factors better correlated with elevation (Ashcroft et al., 2008). However, the aim of the study was to identify the factors affecting the distribution of vegetation in

general, and this does not imply that those same factors affect every species. This is analogous to methods such as Generalised Dissimilarity Modelling (GDM; Ferrier et al., 2007), which combines data for numerous species to estimate compositional turnover using a set of environmental factors. Neither our method nor GDM assume the same environmental factors affect all selected species, yet both aim to identify the factors that are most influential overall. The factors identified will depend on the taxa selected, and so in the present case our results will only reflect the dominant species we selected from the Illawarra Escarpment.

The 37 selected species varied in expected longevity, with some acacias having life spans less than 20 years and eucalypts typically exceeding 100 years. However, the ability for species to adjust their distributions in response to recent warming is not just determined by lifespan, with mortality, reproduction, seed dispersal and seedling establishment also important. Potential mechanisms for change therefore include mortality in areas that became unfavourable, or establishment in newly favourable locations. Some species may be able to survive for some time after conditions become unfavourable for reproduction, however mortality could also be rapid in some circumstances. For example, rainforest species are sensitive to hot, dry winds and bushfires (Fuller, 1995). Similarly, although some rainforest species have mast flowering and eucalypts and acacias may reproduce more abundantly following fire or disturbance, many young plants were observed and species have had some opportunity to reproduce and possibly colonise newly favourable areas. Species' distributions have almost certainly changed to some extent between 1972 and 2005, and we were testing whether it was better to use past or present temperatures to model their current distributions.

#### *2.4. Model production*

All models were produced using Maxent version 3.2.1 (Phillips et al., 2006) with default parameter settings that are suited to a range of presence-only datasets (Phillips and Dudík, 2008). Maxent is a machine-learning modelling method, which has recently gained attention due to its favourable performance in comparison to other modelling methods (Elith et al., 2006). It is more flexible than methods such as Generalised Linear Models (GLMs) and Generalised Additive Models (GAMs) and it can capture complex response curves to environmental gradients.

A benchmark model was produced using geology and all 20 temperature surfaces as candidate predictors. Maxent tends to produce models where many predictors have zero coefficients (Phillips and Dudík, 2008) and therefore not all 21 predictors influenced the models produced. Models were calibrated using a random 70% of the 600 sites, and environmental factors were assessed using the AUC of the remaining 30%.

Note that while the subdivision of data into separate calibration and evaluation datasets is common practice, this does not result in a truly independent dataset. It provides some protection against overfitting to the specific calibration data, but it does not prevent overfitting to the study area or climatic conditions if they are common to both the calibration and evaluation data sets. Methods that perform well when assessed using an ‘independent’ dataset from the same study area (e.g. Elith et al., 2006), can perform poorly when they are applied to other areas (Peterson et al., 2007). While there is currently no solution to this issue, our survey design ensured that spatial autocorrelation between the calibration and evaluation datasets was minimised. That is,

we ensured that adjacent survey sites were from different vegetation communities and geological units where possible.

We used 2-sided paired Student t-tests ( $n = 37$  species) to compare the benchmark models based on all 21 predictors with separate models containing only geology and the ten temperature predictors from 2005. Similarly, we compared the benchmark models with models containing only geology and the ten temperature predictors from 1972. This was effectively testing whether either past or future temperature grids contained unique information. If the 1972 predictors contained unique information then this would provide evidence that species were not in equilibrium with current environmental conditions, while if the 2005 predictors contained unique information it would suggest that species have already shifted their distribution to match current conditions.

We investigated whether the geology, winter minimum or summer maximum predictors were important by looking at their ‘drop’ contributions (Lehmann et al., 2003). If there were a significant drop in performance when predictors were excluded from models, then this would suggest that they contain unique information that cannot be captured by the other predictors. Alternatively, if there were no drop in model performance, then this would reduce confidence in models using that predictor, because there were feasible alternative explanations. For these tests, the models based on geology and the ten 2005 temperature predictors were compared against the models with the geology, winter minimum or summer maximum omitted.

Finally, we determined whether more parsimonious models could perform as well as the models that considered all 21 predictors. We tested two models with only three predictors—geology, winter minimum and either summer maximum or summer minimum (all temperatures predictors from 2005). Summer minimum is well correlated

with elevation, while winter minimum and summer maximum had very different distributional patterns (Ashcroft et al., 2008). All four predictors performed well when evaluated on their ‘alone’ contributions (Lehmann et al., 2003).

### 3. Results

The models produced for the 37 species varied in AUC from 0.599 to 0.976. The differences were mainly due to which species was modelled, with the same species having similar AUC regardless of which predictors were included in models (Fig. 2). Three species were more variable in AUC than others (*Eucalyptus pilularis* (BB), *Syncarpia glomulifera* (TT), and *Acacia mearnsii* (GW)), and therefore t-tests were performed with and without these species to avoid the results being influenced by three ‘outliers’.

There were no significant ( $P < 0.05$ ) differences between the benchmark model and those based only on the 1972 or 2005 predictors (Fig. 2a). The benchmark models had the highest average AUC over the 37 models (0.818) followed by the 1972 models (0.817) and the 2005 models (0.815). The paired t-tests suggested these differences were not significant when the three outlying species were excluded ( $P > 0.4$ ), however the 2005 models were almost significantly poorer than the benchmark models when all species were used ( $P = 0.058$ ).

Excluding summer maximum temperatures from models did not have a significant effect ( $P = 0.235$  with all 37 species,  $P = 0.928$  without 3 outliers), with the average AUC of the 37 models increasing from 0.815 (2005 models) to 0.817 (Fig. 2b). In contrast, excluding winter minimum or geology did have a significant effect ( $P < 0.02$ ) regardless of whether or not the three outlying species were considered (Fig. 2c).



Excluding geology decreased the average AUC of the 37 models from 0.815 to 0.797, while excluding winter minimum decreased it to 0.806.

Models produced using only geology, winter minimum and summer maximum performed significantly worse than the 2005 models ( $P < 0.004$ ), with the average AUC of the 37 models decreasing from 0.815 to 0.806 (Fig. 2d). Models produced using geology, winter minimum and summer minimum were almost significantly worse than the 2005 models when evaluated without the three outlying species ( $P = 0.058$ ), but were not significantly different when evaluated using all species ( $P = 0.859$ ). The average AUC of the 37 models decreased from 0.815 to 0.814.

The large difference in performance between species appeared to be due to the extent and specialisation in their distribution. The species with the highest AUC typically had clustered distributions that were restricted to the Hawkesbury sandstone at higher elevations (SG, SA, RB), the slopes of the escarpment (GST, NC, RC), the gullies on the Woronora Plateau (MGG, GG, SPM), or rainforest patches (SF, CW, BWR, BP, FW, PT). The species that had lower AUC were widespread, and found across a number of these habitats (GW, BGH, VMO, HC, CTP, SP). The nine species with the highest AUC were relatively rare species (fewer than 83 presences), while five of the six worst performing species were common (more than 142 presences).

The species with high AUC had good models with a number of different predictor combinations. For example, the species on the Hawkesbury sandstone had distributions that were predominately restricted to a single geological unit, high elevations, and certain temperature regimes to which these topographic positions are subjected. Any of these factors offered statistically good explanations for their distribution. In contrast, the common and widespread species were found over a range of conditions for all environmental factors we considered, and it was difficult to come up with any

explanation that performed as well as the restricted species (see also Franklin et al., 2009).

As summer maximum temperature performed well in univariate models (Ashcroft et al., 2008), we investigated why excluding summer maximum did not have a detrimental effect on models (see above). We found that although summer maximum temperatures provided a good explanation for the distribution of rainforests, there were other potential explanations such as a simple linear combination of winter minimum and summer minimum (Fig. 3).

## **4. Discussion**

### *4.1. Factors affecting the distribution of vegetation on the Illawarra Escarpment*

The methods we introduced in this article were designed to strengthen our ability to infer the environmental factors that determine the distribution of dominant plant species on the Illawarra Escarpment. We found there was a significant ( $P < 0.05$ ) decrease in model performance when either winter minimum temperatures or geology were excluded from models, thus providing strong support for these predictors. These factors do not necessarily affect all of the 37 species we examined, but they do explain at least some of the species' distributions better than any of the other environmental factors we examined. That is, winter minimum temperatures and geology each contain unique information, at least within the predictors examined, and are therefore likely to be important determinants of local vegetation patterns. Other factors may also be important for other species, at other scales, or in other study areas, and the methods introduced here could be replicated in each of these situations.

We found less support for summer maximum temperature as a predictor. While summer maximum temperature is strongly correlated with the distributions of many rainforest species in this study area (Ashcroft et al., 2008), we found that other predictors could combine to compensate for its exclusion. Summer maximums may still be an important determinant of species distributions, but there are alternative hypotheses that also need further investigation. Further evidence on causal factors may be obtained by further survey work to evaluate and refine models (Engler et al., 2004; Guisan et al., 2006) or by manipulation experiments conducted to test the effect of different environmental factors.

One limitation of our method is that it can only test for unique contributions among the predictors that are considered. There is still a risk that it could identify an important predictor that is simply correlated with an unconsidered environmental factor. For example, our study did not include a moisture predictor, such as soil moisture, rainfall, vapour-pressure deficit or evapotranspiration. Given that there are strong links between surface moisture balance and temperature regimes (Geiger, 1971; Ridolfi et al., 2003; Lookingbill and Urban, 2004; Lu et al., 2009), we cannot exclude the possibility that temperature predictors only appear important because they are correlated with moisture. We have no empirical data to estimate the distribution of moisture on the escarpment or evaluate potential surrogates, but observations from nearby areas using DS1923 hygrochron iButtons (Ashcroft and Gollan, unpubl. data) have confirmed that near-surface temperatures are generally strongly correlated with humidities. However, these correlations are lower for winter minimum temperatures. Therefore, we do not believe that winter minimum temperatures are compensating for missing moisture predictors, although it increases the number of alternative hypotheses regarding the relationship between vegetation distributions and summer maximum temperatures.

#### *4.2 Have species adjusted their distributions in response to recent climate change?*

There was no significant difference in the performance between the benchmark models and those that contained only 1972 or 2005 temperature predictors. There was also no evidence that any species, or group of species, was better modelled by one set of predictors or the other. Therefore, our results do not provide any evidence as to whether the distribution of these plants is determined more by past or current temperatures.

Although we catered for spatial variations in warming (Ashcroft et al., 2009), it appears that the magnitude of these spatial variations was insufficient to differentiate between the predictive power of past and current temperature surfaces. The variations in warming varied widely amongst sites (e.g. summer maximum temperatures warmed by 0.13–0.28°C and winter minimums warmed by 0.58–1.11°C) and the distribution of warming was poorly correlated with both past and present temperatures ( $r^2 < 0.19$ ). Nevertheless, these variations were insufficient to prevent a high correlation between past and present temperatures ( $r^2 > 0.985$ ) and hence the models performed similarly. The question of whether past or present temperatures are best for modelling species distributions remains an important goal for future research. For example, both the 1972 and 2005 models we produced performed similarly, yet there is ~0.8°C bias in winter minimum temperatures between the two models. If the two models were projected into future climates then they would make very different predictions. While it is perhaps obvious that you should match the timing of environmental and biotic data when modelling short-lived species (Roubicek et al., 2010), the environmental data that is used for longer lived species may need to be from a prior time period.

#### 4.3. Increasing inference on causality

Methods such as hierarchical partitioning (Mac Nally, 2002) and the ones introduced in this article are designed to increase inference on causal factors rather than identify a single best model for any species. This has been identified as an important issue for SDMs (Guisan and Thuiller, 2005; Araújo and Guisan, 2006). We increased inference on causality by investigating whether a predictor had consistently strong explanatory ability for a range of species, and provided information that could not be captured by other predictors. While this strengthening of inference over multiple species has previously been done qualitatively (Peterson and Cohoon, 1999; Watson and Peterson, 1999), the addition of the statistical test increases the reliability of the method.

Some aspects of the statistical test require further comment. First, it is implicitly assumed that all species are independently distributed. If species share similar distributions then this could introduce pseudoreplication and affect the significance of the t-test. In our study, all eucalypts and acacias had distinctly different distributions that we considered independent. Many rainforest species favoured locations where summer maximum temperatures were low, but some species were only found in the foothills, some only on the escarpment, and some only on the Woronora Plateau. The most similar distributions were between: *Eucalyptus sieberi* and *Corymbia gummifera* (prevalences of 82 and 65 respectively, 48 sites in common); *Cassine australis* and *Croton verreauxii* (prevalences of 129 and 112 respectively, 81 sites in common); and *Polyosma cunninghamii* and *Tasmania insipida* (prevalences of 57 and 49 respectively, 33 sites in common). We considered these sufficiently different for our study, although

care should be taken if replicating the methods with species that have more similar distributions.

Another issue is that if too few species are included in the analysis, it may not be possible to detect significant differences. Similarly, if too many species are used then even small differences may be significant. Focusing only on P-values may obscure the effect size (differences in AUC; Fig. 2), and care should be taken to ensure that these are large enough to have biological relevance (Anderson et al., 2000).

Our method is designed to provide a statistical basis for determining which factors should be considered for inclusion in species distribution models. The most commonly used alternative is to make an *a priori* assumption on which predictors are ecologically important enough to include in candidate models. For example, temperature predictors are often reduced to mean annual temperature, winter minimum, and summer maximum or growing season degree-days (e.g. Araújo and Luoto, 2007). While these variables will be physiologically limiting for many species, this will clearly impact model performance if their distributions are influenced by other seasonal temperatures. Indeed, the results of our study showed that models that were restricted to winter minimum and summer maximum performed significantly ( $P < 0.05$ ) worse than models that also included temperature predictors from other seasons. Distinguishing between multiple correlated and confounded predictors is an important issue to address in SDMs (Dormann, 2007), and our method helps to disentangle these complications. Many statistical methods can only deal with a limited number of environmental variables (Guisan and Zimmermann, 2000), and therefore our method provides a quantitative method to reduce the number of variables before conducting analysis using tools such as GDM or various SDM tools.

While our analysis focused on species distribution models, other methods have also been used to examine which variables are important determinants of species distributions, including K-select analysis (Calenge et al., 2005), Ecological Niche Factor Analysis (ENFA; Hirzel et al., 2002) and General Niche-Environment System Factor Analysis (GNESFA; Calenge and Basille, 2008). However, these methods are also prone to identifying spurious correlations and they have trouble distinguishing between correlated predictors (Hirzel et al., 2002; Basille et al., 2008). Like species distribution models, there is still a need to complement these analysis with an assessment of which environmental factors are important.

Our method is designed to test which environmental factors are significantly affecting species distributions, but it does not provide information on which or even how many species those factors affect. Indeed, it is clear that excluding the significant environmental factors did not negatively affect all species (Fig. 2), and in some cases excluding them even had a positive effect. This could occur where the inclusion of a variable caused overfitting to the training data. It is worth highlighting that assessing candidate models using the evaluation data is not only desirable to prevent overfitting, it is also necessary in our case. The exclusion of variables can only result in a drop in performance when assessed with training data, and this would preclude our method being used exclusively with the training data. However, our method is not limited to situations where the data is split into calibration and evaluation datasets, and could also be used with 10-fold cross validation or other methods.

#### *4.4. Implications for model evaluation*

Many SDMs are justified by proving they exceed predefined performance benchmarks, such as an AUC of more than 0.7 or 0.9 (Swets, 1988; Pearce and Ferrier, 2000). The results of our study suggest that these benchmarks provide little information on the model building process, and do not prove that the selected model is better than the alternatives (see also Lobo et al., 2008; Peterson et al., 2008). All alternative models produced for a given species resulted in a similar AUC, and this was determined by the rarity and extent of its distribution (Franklin et al., 2009). Similarly, Elith et al. (2006) tested 10 different statistical methods and found that predictive success varied more between species than between methods. The use of performance benchmarks needs to be re-examined, and a greater emphasis placed on proving that the selected model is better than the alternatives.

For common species with widespread distributions, it may be difficult to determine any model that exceeds a given benchmark. However, the candidate models may still provide an important contribution to environmental management or suggest valid hypotheses for the factors that limit these species' distributions. Alternatively, there may be many models that exceed the benchmark for rare species with clustered distributions. Some of these models are likely to be based on spurious correlations with spatially structured environmental variables, and exceeding the benchmark may lead to misplaced confidence in the selected model. The alternative models for these species may have similar AUC, but their predictions may be dramatically different, especially when applied in a climate change setting (Araújo and New, 2006; Austin, 2007).

## **Acknowledgements**



This research was conducted as part of a Ph.D. at the University of Wollongong with a University Postgraduate Award scholarship. The research could not have been completed without the data obtained from the Spatial Analysis Laboratory in the School of Earth and Environmental Sciences, much of which has been supplied by AAMHatch, the NSW Department of Environment, Climate Change and Water, and the NSW Department of Primary Industries. Thanks to Bernard Ashcroft, Imke Büsing, Berit Fehlau, Jodie Dunn, Carolyn Blomley, Anders Bofeldt, Roger Ferguson, Elfi Ashcroft, Roger Starkie, Andrea Kubler, and Rike Arnold for help with the fieldwork, to the many landholders who generously granted us permission to access their land. Thanks to Town Peterson, Sara Varela and the anonymous referees for comments that greatly improved this manuscript.

## References

- Anderson, D.R., Burnham, K.P., Thompson, W.L., 2000. Null hypothesis testing: problems, prevalence, and an alternative. *J. Wildlife Manag.* 64, 912–923.
- Araújo, M.B., Guisan, A., 2006. Five (or so) challenges for species distribution modelling. *J. Biogeogr.* 33, 1677–1688.
- Araújo, M.B., New, M., 2006. Ensemble forecasting of species distributions. *Trends Ecol. Evol.* 22, 42–47.
- Araújo, M.B., Luoto, M., 2007. The importance of biotic interactions for modelling species distributions under climate change. *Global Ecol. Biogeogr.* 16, 743–753.
- Ashcroft, M.B., Chisholm, L.A., French, K.O., 2008. The effect of exposure on landscape scale soil surface temperatures and species distribution models. *Landscape Ecol.* 23, 211–225.

- Ashcroft, M.B., Chisholm, L.A., French, K.O., 2009. Climate change at the landscape scale: predicting fine-grained spatial heterogeneity in warming and potential refugia for vegetation. *Global Change Biol.* 15, 656–667.
- Austin, M., 2007. Species distribution models and ecological theory: a critical assessment and some possible new approaches. *Ecol. Model.* 200, 1–19.
- Bahn, V., McGill, B.J., 2007. Can niche-based distribution models outperform spatial interpolation? *Global Ecol. Biogeogr.* 16, 733–742.
- Basille, M., Calenge, C., Marboutin, E., Andersen, R., Gaillard, J-M., 2008. Assessing habitat selection using multivariate statistics: some refinements of the ecological-niche factor analysis. *Ecol. Model.* 211, 233–240.
- Beale, C.M., Lennon, J.J., Gimona, A., 2008. Opening the climate envelope reveals no macroscale associations with climate in European birds. *P. Natl. Acad. Sci. USA* 105, 14908–14912.
- Calenge, C., Dufour, A.B., Maillard, D., 2005. K-select analysis: a new method to analyse habitat selection in radio-tracking studies. *Ecol. Model.* 186, 143–153.
- Calenge, C., Basille, M., 2008. A general framework for the statistical exploration of the ecological niche. *J. Theor. Biol.* 252, 674–685.
- Currie, D.J., 2007. Disentangling the roles of environment and space in ecology. *J. Biogeogr.* 34, 2009–2011.
- Daly, C., 2006. Guidelines for assessing the suitability of spatial climate data sets. *Int. J. Climatol.* 26, 707–721.
- Dormann, C.F., 2007. Promising the future? Global change projections of species distributions. *Basic Appl. Ecol.* 8, 387–397.
- Elith, J., Graham, C.H., Anderson, R.P., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G.,

- Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.McC., Peterson, A.T., Phillips, S.J., Richardson, K., Scachetti-Pereira, R., Schapire, R.E., Soberón, J., Williams, S., Wisz, M.S., Zimmermann, N.E., 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29, 129–151.
- Engler, R., Guisan, A., Rechsteiner, L., 2004. An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. *J. Appl. Ecol.* 41, 263–274.
- Faith, D.P., 2003. Environmental diversity (ED) as surrogate information for species-level biodiversity. *Ecography* 26, 374–379.
- Faith, D.P., Ferrier, S., Walker, P.A., 2004. The ED strategy: how species-level surrogates indicate general biodiversity patterns through an 'environmental diversity' perspective. *J. Biogeogr.* 31, 1207–1217.
- Ferrier, S., Manion, G., Elith, J., Richardson, K., 2007. Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. *Divers. Distrib.* 13, 252–264.
- Franklin, J., Wejnert, K.E., Hathaway, S.A., Rochester, C.J., Fisher, R.N., 2009. Effect of species rarity on the accuracy of species distribution models for reptiles and amphibians in southern California. *Divers. Distrib.* 15, 167–177.
- Freedman, D.A., 1983. A note on screening regression equations. *Am. Stat.* 37, 152–155.
- Fuller, L., 1995. Wollongong's Native Trees. Kingsclear books, Alexandria.
- Geiger, R., 1971. *The Climate Near the Ground*. Revised edn. Harvard Univ. Press, Cambridge.

Graham, M.H., 2003. Confronting multicollinearity in ecological multiple regression. *Ecology* 84, 2809–2815.

Guisan, A., Zimmermann, N.E., 2000. Predictive habitat distribution models in ecology. *Ecol. Model.* 135, 147–186.

Guisan, A., Thuiller, W., 2005. Predicting species distribution: offering more than simple habitat models. *Ecol. Lett.* 8, 993–1009.

Guisan, A., Broennimann, O., Engler, R., Vust, M., Yoccoz, N.G., Lehmann, A., Zimmermann, N.E., 2006. Using niche-based models to improve the sampling of rare species. *Conserv. Biol.* 20, 501–511.

Hirzel, A., Guisan, A., 2002. Which is the optimal strategy for habitat suitability modelling. *Ecol. Model.* 157, 331–341.

Hirzel, A.H., Hausser, J., Chessel, D., Perrin, N., 2002. Ecological-niche factor analysis: how to compute habitat-suitability maps without absence data? *Ecology* 83, 2027–2036.

Lehmann, A., Overton, J.M., Leathwick, J.R., 2003. GRASP: generalized regression analysis and spatial prediction. *Ecol. Model.* 160, 165–183.

Lobo, J.M., Jiménez-Valverde, A., Real, R., 2008. AUC: a misleading measure of the performance of predictive distribution models. *Global Ecol. Biogeogr.* 17, 145–151.

Lookingbill, T., Urban, D., 2004. An empirical approach towards improved spatial estimates of soil moisture for vegetation analysis. *Landsc. Ecol.* 19, 417–433.

Lu, S., Ju, Z., Ren, T., Horton, R., 2009. A general approach to estimate soil water content from thermal inertia. *Agric. For. Meteorol.* 114, 141–151.

- Mac Nally, R., 2002. Multiple regression and inference in ecology and conservation biology: further comments on identifying important predictor variables. *Biodivers. Conserv.* 11, 1397–1401.
- Moffit, R.S., 1999. Southern Coalfield Regional Geology 1:100 000, 1<sup>st</sup> edition. Geological Survey of New South Wales, Sydney.
- Murray, K., Conner, M.M., 2009. Methods to quantify variable importance: implications for the analysis of noisy ecological data. *Ecology* 90, 348–355.
- NPWS, 2002. Native Vegetation of the Illawarra Escarpment and Coastal Plain. NSW National Parks and Wildlife Service, Sydney.
- Pearce, J., Ferrier, S., 2000. Evaluating the predictive performance of habitat models developed using logistic regression. *Ecol. Model.* 133, 225–245.
- Peterson, A.T., Cohoon, K.P., 1999. Sensitivity of distributional prediction algorithms to geographic data completeness. *Ecol. Model.* 117, 159–164.
- Peterson, A.T., Papeş, M., Eaton, M., 2007. Transferability and model evaluation in ecological niche modeling: a comparison of GARP and Maxent. *Ecography* 30, 550–560.
- Peterson, A.T., Papeş, M., Soberón, J., 2008. Rethinking receiver operating characteristic analysis applications in ecological niche modeling. *Ecol. Model.* 213, 63–72.
- Phillips, S.J., Anderson, R.P., Schapire, R.E., 2006. Maximum entropy modeling of species geographic distributions. *Ecol. Model.* 190, 231–259.
- Phillips, S.J., Dudík, M., 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* 31, 161–175.

- Platts, P.J., McClean, C.J., Lovett, J.C., Marchant, R., 2008. Predicting tree distributions in an East African biodiversity hotspot: model selection, data bias and envelope uncertainty. *Ecol. Model.* 218, 121–134.
- Ridolfi, L., D’Odorico, P., Porporato, A., Rodriguez-Iturbe, I., 2003. Stochastic soil moisture dynamics along a hillslope. *J. Hydrol.* 272, 264–275.
- Roubicek, A.J., VanDerWal, J., Beaumont, L.J., Pitman, A.J., Wilson, P., Hughes, L., 2010. Does the choice of climate baseline matter in ecological niche modelling? *Ecol. Model.* 221, 2280–2286.
- Stephens, P.A., Buskirk, S.W., del Rio, C.M., 2007. Inference in ecology and evolution. *Trends Ecol. Evol.* 22, 192–197.
- Stockwell, D.R.B., Peterson, A.T., 2002. Effects of sample size on accuracy of species distribution models. *Ecol. Model.* 148, 1–13.
- Swets, J.A., 1988. Measuring the accuracy of diagnostic systems. *Science* 240, 1285–1293.
- Watson, D.M., Peterson, A.T., 1999. Determinants of diversity in a naturally fragmented landscape: humid montane forest avifaunas of Mesoamerica. *Ecography* 22, 582–589.
- Whittingham, M.J., Stephens, P.A., Bradbury, R.B., Freckleton, R.P., (2006) Why do we still use stepwise modelling in ecology and behaviour? *J. Anim. Ecol.* 75, 1182–1189.

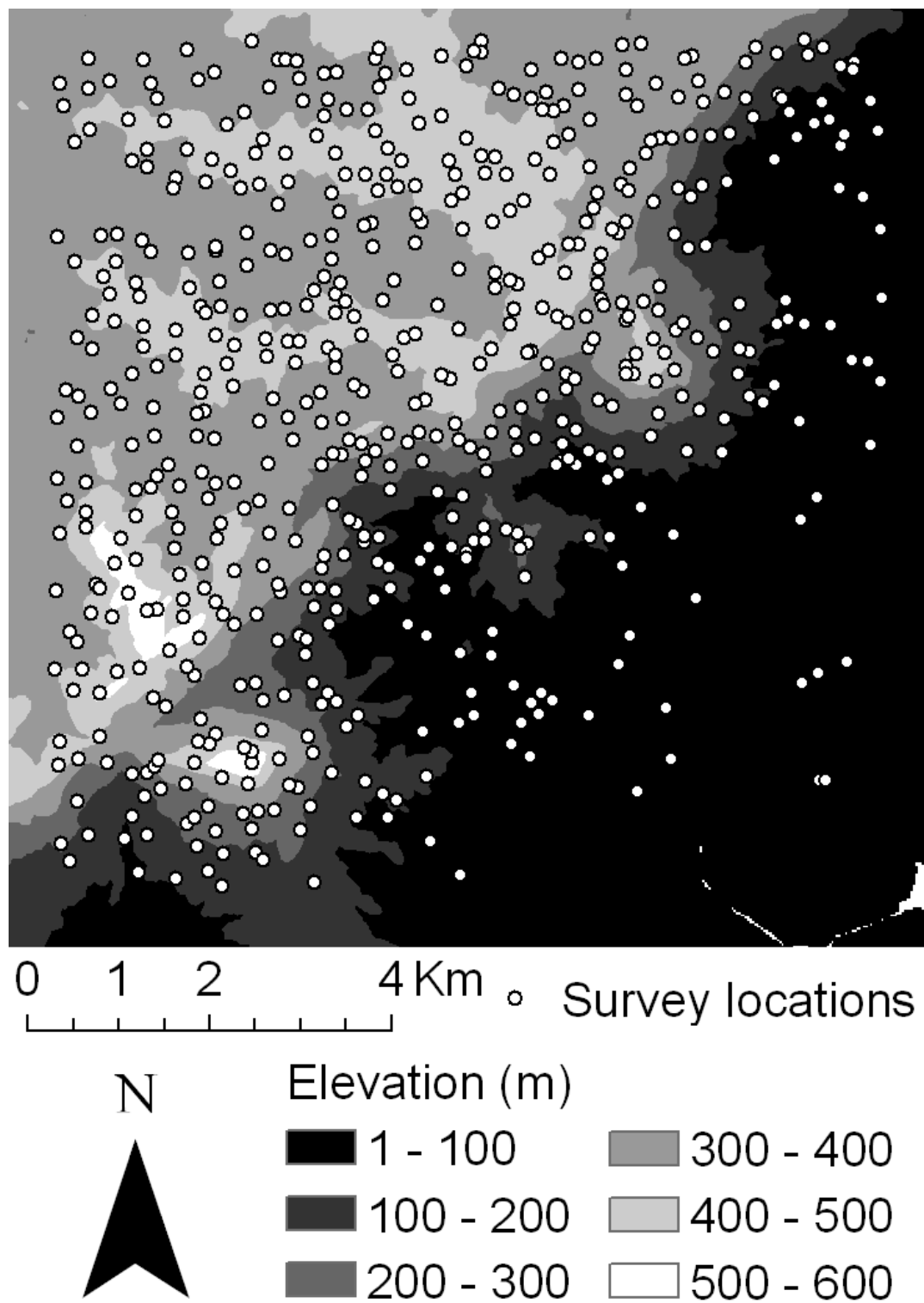


Fig. 1: The topography of the Illawarra Escarpment, 80 km south of Sydney, Australia (34.4°S, 150.9°E). A vegetation survey was conducted at each of the 600 sites indicated.

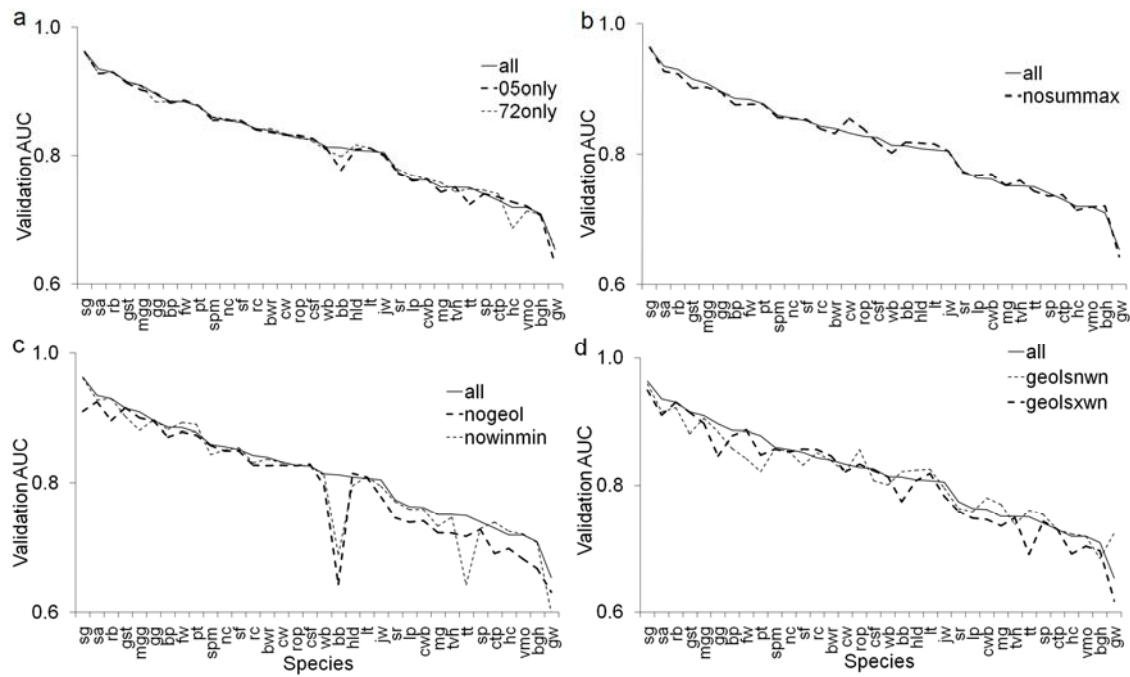


Fig. 2: The results of Maxent models for 37 species (Table 1) as evaluated using the Area Under the receiver operating characteristics Curve (AUC) on a random evaluation dataset consisting of 30% of the surveyed sites. The ‘all’ model used geology and 20 temperature surfaces as predictors. The ‘05only’ and ‘72only’ models used geology and the 10 temperature surfaces from either 2005 or 1972 respectively. The ‘nosummax’, ‘nogeol’ and ‘nowinmin’ models were the same as the ‘05only’ models, but excluded either summer maximum temperature, geology, or winter minimum respectively. The ‘geolsnwn’ and ‘geolsxwn’ models only used geology, winter minimum, and either summer minimum or summer maximum respectively.



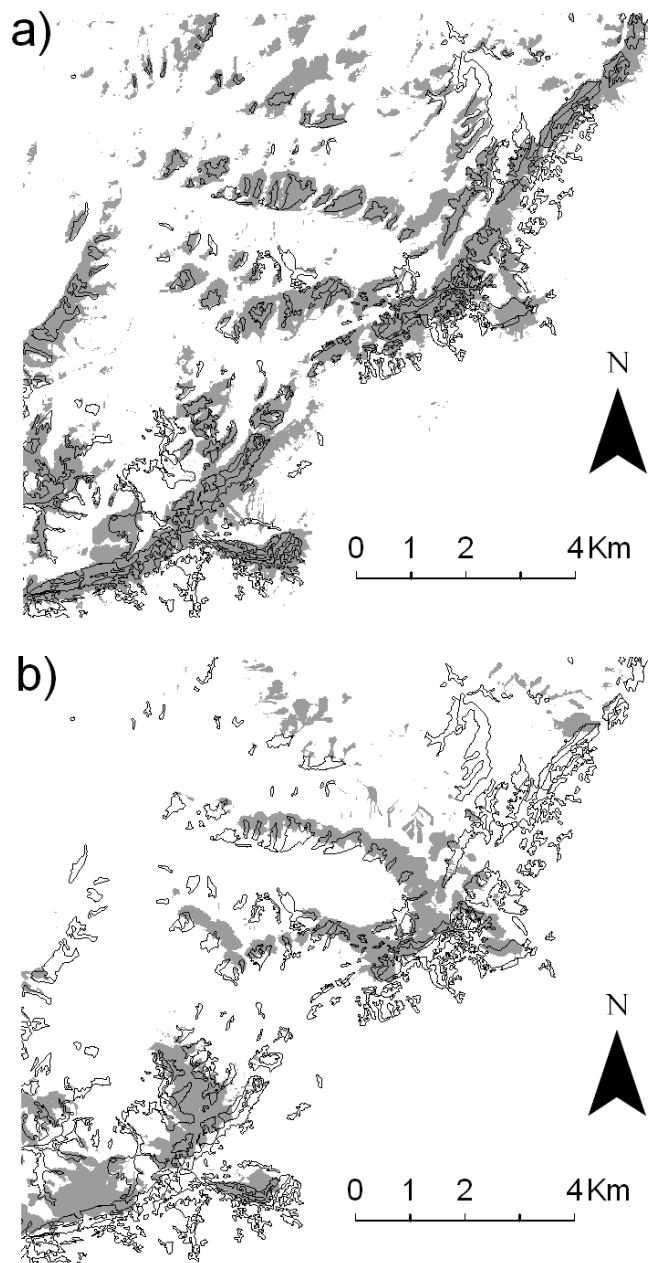


Fig. 3: The location of moist rainforests on the Illawarra Escarpment (Coachwood Warm Temperate Rainforest or Illawarra Escarpment Subtropical Rainforest; NPWS 2002) is shown in black polygons on both panels. Panel (a) indicates the locations where summer maximum temperatures are less than 20.5 °C, and panel (c) indicates where summer minimum temperatures – 0.4 \* winter minimum temperatures is less than 12.8.