

University of Wollongong - Research Online

Thesis Collection

Title: The spatial variation of environmental factors on the Illawarra escarpment and their influence on vegetation patterns

Author: Michael Ashcroft

Year: 2009

Repository DOI:

Copyright Warning

You may print or download ONE copy of this document for the purpose of your own research or study. The University does not authorise you to copy, communicate or otherwise make available electronically to any other person any copyright material contained on this site.

You are reminded of the following: This work is copyright. Apart from any use permitted under the Copyright Act 1968, no part of this work may be reproduced by any process, nor may any other exclusive right be exercised, without the permission of the author. Copyright owners are entitled to take legal action against persons who infringe their copyright. A reproduction of material that is protected by copyright may be a copyright infringement. A court may impose penalties and award damages in relation to offences and infringements relating to copyright material.

Higher penalties may apply, and higher damages may be awarded, for offences and infringements involving the conversion of material into digital or electronic form.

Unless otherwise indicated, the views expressed in this thesis are those of the author and do not necessarily represent the views of the University of Wollongong.

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

University of Wollongong Thesis Collections

University of Wollongong Thesis Collection

University of Wollongong

Year 2009

The spatial variation of environmental
factors on the Illawarra escarpment and
their influence on vegetation patterns

Michael B. Ashcroft
University of Wollongong

Ashcroft, Michael B., The spatial variation of environmental factors on the Illawarra escarpment and their influence on vegetation patterns, Doctor of Philosophy thesis, School of Earth & Environmental Sciences - Faculty of Science, University of Wollongong, 2009. <http://ro.uow.edu.au/theses/3042>

This paper is posted at Research Online.

NOTE

This online version of the thesis may have different page formatting and pagination from the paper copy held in the University of Wollongong Library.

UNIVERSITY OF WOLLONGONG

COPYRIGHT WARNING

You may print or download ONE copy of this document for the purpose of your own research or study. The University does not authorise you to copy, communicate or otherwise make available electronically to any other person any copyright material contained on this site. You are reminded of the following:

Copyright owners are entitled to take legal action against persons who infringe their copyright. A reproduction of material that is protected by copyright may be a copyright infringement. A court may impose penalties and award damages in relation to offences and infringements relating to copyright material. Higher penalties may apply, and higher damages may be awarded, for offences and infringements involving the conversion of material into digital or electronic form.

The spatial variation of environmental factors on the Illawarra Escarpment and their influence on vegetation patterns

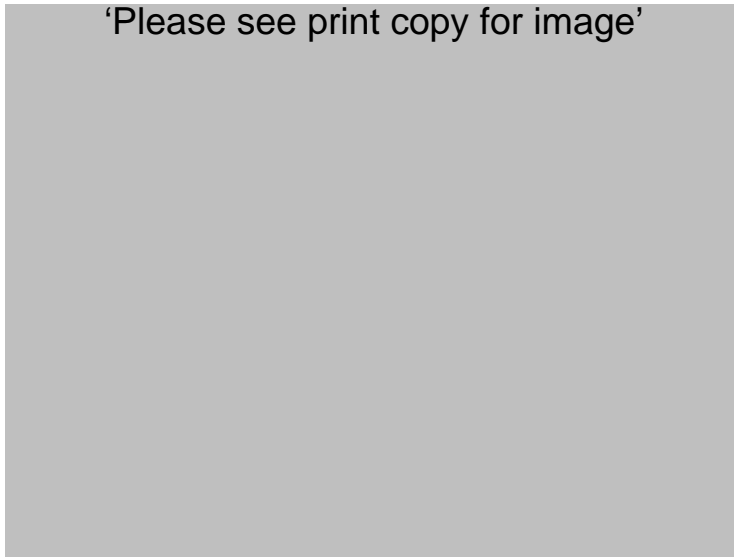
A thesis submitted in fulfilment of the requirements for the award of the degree of

Doctorate of Philosophy

From the School of Earth and Environmental Sciences, University of Wollongong

By Michael B. Ashcroft, BMath/BE (Elec), MEnvSci, October 2009

‘Please see print copy for image’



The escarpment between Mt Keira and the City of Wollongong.
Photo: Imke Büsing © 2005



The relatively undisturbed vegetation on the Woronora Plateau.
Photo: Bernard Ashcroft © 2005

Thesis certification

I, Michael Bernard Ashcroft, declare that this thesis, submitted in fulfilment for the requirements for the award of Doctor of Philosophy in the School of Earth and Environmental Sciences, University of Wollongong, is wholly my own work unless otherwise acknowledged. The document has not been submitted for qualifications at any other academic institution.

The chapters in this thesis that have been published with co-authors are approximately 80% my own work. I developed the initial ideas, aims and methods for each chapter, conducted the analyses, and prepared the first drafts of the articles. These drafts were then subject to change following reviews from my supervisors, journal reviewers and editors, and other people that are included in the acknowledgements. In general, these changes improved the presentation and interpretation of results, but did not change the original aims or methods. For consistency, 'I' is used throughout this thesis, although 'we' is used in the corresponding journal articles that have been prepared with co-authors.

Mick Ashcroft

Laurie Chisholm

Kristine French

October 2009

Table of Contents

Abstract	xv
Acknowledgments	xix
Chapter 1: Introduction	1
1.1 Modelling the spatial distribution of vegetation	1
1.1.1 Dynamic, mechanistic or process-based models	1
1.1.2 Static, correlative, niche-based models	4
1.1.3 Statistical methods for developing SDMs	4
1.1.4 Model evaluation	6
1.1.5 Issues with species distribution modelling	8
1.2 The vegetation of the Illawarra Escarpment	11
1.3 Research aims and thesis outline	17
Chapter 2: A method for improving landscape scale temperature predictions and the implications for vegetation modelling	20
2.1 Introduction	20
2.2 Materials and methods	24
2.2.1 Environmental predictor variables	24
2.2.2 Predicting maximum and minimum temperatures	26
2.2.3 Using a low pass filter to improve estimates	29
2.2.4 Vegetation modelling	30
2.3 Results	32
2.3.1 Predicting minimum temperature	32
2.3.2 Predicting maximum temperature	34
2.3.3 The effects of low pass filters	34
2.3.4 Intra-day trends	36
2.3.5 Vegetation modelling	38
2.4 Discussion	39
2.4.1 Temperature prediction	39
2.4.2 Vegetation modelling	43
2.5 Conclusions	47
Chapter 3: The effect of exposure on landscape scale soil surface temperatures and species distribution models	48
3.1 Introduction	48
3.2 Methods	51
3.2.1 Study area	51
3.2.2 Data	53
3.2.3 Temperature analysis	57
3.2.4 Species distribution models	58
3.3 Results	59
3.3.1 Maximum temperatures	60
3.3.2 Minimum temperatures	63
3.3.3 Average temperatures	65
3.3.4 Species distribution models	65
3.4 Discussion	69
3.4.1 Factors affecting landscape scale temperatures	69
3.4.2 Which environmental factors influence the distribution of species?	72
3.4.3 Why is elevation a better predictor for some species?	73

3.5	Conclusions	74
3.6	Supplementary material	74
3.6.1	Source and accuracy of predictors	74
3.6.2	The differences between exposure and aspect	77
3.6.3	Results for average temperatures	78
Chapter 4:	Addressing spatial autocorrelation in species distribution models without using survey data as a predictor	85
4.1	Introduction	85
4.1.1	Extending the contextual indices methodology	89
4.2	Methods	90
4.2.1	Single-predictor models	93
4.2.2	Multiple-predictor models	94
4.3	Results	97
4.3.1	Single-predictor models	97
4.3.2	Multiple-predictor models	100
4.4	Discussion	106
4.4.1	What determines the optimal radius?	107
4.4.2	Comparison with other models and predictors	109
4.4.3	How should habitat quality be determined?	110
4.4.4	Further recommendations	111
4.5	Supplementary material	111
4.5.1	Geology and vegetation communities in the study area	111
4.5.2	Survey design	113
Chapter 5:	Climate change at the landscape scale: predicting fine-grained spatial heterogeneity in warming and potential refugia for vegetation	120
5.1	Introduction	120
5.2	Materials and methods	122
5.2.1	Overview of approach	122
5.2.2	Temporal changes in weather patterns	123
5.2.3	Relationships between air and soil temperatures	125
5.2.4	Estimating spatial variations in temperature change from 1972 to 2007	127
5.2.5	Estimating the impact of climate change on plants	128
5.3	Results	131
5.3.1	Temporal changes in weather patterns	131
5.3.2	The relationships between soil and air temperatures	133
5.3.3	Estimating the spatial distribution of warming from 1972 to 2006	135
5.3.4	Effect of warming on vegetation	138
5.4	Discussion	139
5.4.1	The importance of weather patterns for climate change predictions	139
5.4.2	Selecting the correct temperature predictors in models	142
5.4.3	The relationship between soil and air temperatures	143
5.4.4	Assumptions and limitations	144
5.5	Conclusions	144
Chapter 6:	An evaluation of the environmental factors in species distribution models for vegetation on the Illawarra Escarpment	146
6.1	Introduction	146
6.2	Methods	150
6.2.1	Study area	150
6.2.2	Environmental predictors	150
6.2.3	Vegetation data	152

6.2.4	Model production	154
6.3	Results	155
6.4	Discussion	159
6.4.1	Factors affecting the distribution of vegetation on the Illawarra Escarpment 159	
6.4.2	Distinguishing causal from spurious predictors	162
6.4.3	Implications for model validation	163
Chapter 7:	Discussion: The benefits and challenges of Species Distribution Models at the landscape scale	165
7.1	Introduction	165
7.2	The challenges of landscape scale models	166
7.3	Benefit 1: reduction of pseudoreplication	169
7.4	Benefit 2: less predictor colinearity	170
7.5	Benefit 3: identification of refugia from climate change	175
7.6	Benefit 4: matching scale and process	178
7.7	Conclusions	180
References	182

List of Tables

Table 3.1: The 37 species from the Illawarra Escarpment that were modelled as part of this study. * marks an introduced species.....	54
Table 3.2: The frequency (Freq) of winds at 9am over 35 years from 16 main compass directions. Data was obtained from the Bureau of Meteorology (they must be contacted to purchase it) and relates to daily observations from the University of Wollongong station (068188), which is located in the foothills east of Mt Keira (Figure 3.2). Recording period was March 1972 to February 2007 inclusive, and there were more than 3000 observations per season. Note that Summer is December to February, and Winter June to August. The Humidity (Humid), wind speed (Speed, km/h) and minimum temperature (Min, °C) were averaged over the days when winds from each direction occurred.....	81
Table 3.3: The frequency (Freq) of winds at 3pm over 35 years from 16 main compass directions. Data was obtained as per Table 3.2. The Humidity (Humid), wind speed (Speed, km/h) and maximum temperature (Max, °C) were averaged over the days when winds from each direction occurred.	82
Table 3.4: Seasonal temperatures were regressed against elevation, radiation, moisture and exposure predictors. The numbers in this table indicate the percentage of regressions in which each predictor (other than exposure) was selected. The numbers in parentheses indicate how often they were significant ($P < 0.05$). Min, Max and Ave are short for minimum, maximum and average temperatures respectively. Blank cells indicate predictor was never selected.	83
Table 3.5: Seasonal temperatures were regressed against elevation, radiation, moisture and exposure predictors. The numbers indicate the percentage of regressions in which each exposure predictor was selected (each column represents the combined total from three directions). The number in parentheses indicates how often they were significant ($P < 0.05$). Min, Max and Ave are short for minimum, maximum and average temperatures respectively. Blank cells indicate predictor was never selected.....	84
Table 5.1: The 37 species from the Illawarra Escarpment that were modelled as part of this study.	130
Table 5.2: The amount of warming in ten seasonal temperature variables was estimated at 600 sites where a vegetation survey was conducted. I calculated the average	

amount of warming for each of 37 species by averaging the warming at the sites at which those species were observed. The middle columns represent the mean warming for those species with the lowest and highest averages respectively. The final column indicates the percentage difference between these two values.	138
Table 6.1: The 37 species from the Illawarra Escarpment that were modelled as part of this study.	153

List of Figures

Figure 1.1: The location of the study area.....	12
Figure 1.2: The elevation of the Illawarra Escarpment (yellow to blue), with the area cleared for the city of Wollongong shown in red.....	13
Figure 1.3: The north south trend in eucalypts that occurs on the Illawarra Escarpment (from NPWS 2002).	14
Figure 1.4: The patchy distribution of the moist (blue, green) and dry (red) rainforest communities that are located on the Illawarra Escarpment (from NPWS 2002). Elevation is depicted in the background as a grey-scale image (Figure 1.1).....	16
Figure 2.1: The topography of the Illawarra Escarpment in the vicinity of Wollongong, Australia. The inset is a Digital Elevation Model showing the rising elevation from the coastal plain to the Woronora Plateau, with Mt Keira and Mt Kembla protruding eastward.....	22
Figure 2.2: The relationship between the predicted average summer minimum (a) and maximum (b) temperatures and the corresponding minimum and maximum temperatures that were recorded by 39 temperature sensors placed at ground level on the Illawarra Escarpment. The predicted average minimum temperature is based on the regression of the actual recorded minimum temperatures against the elevation of each site, whilst the predicted average maximum temperature is based on the regression of the actual recorded maximum temperature against the elevation and canopy cover of each site.....	33
Figure 2.3: The r^2 and P values that resulted from the regression of the average summer maximum temperature against various predictors for 39 temperature sensors placed at ground level on the Illawarra Escarpment. Predictors were the low pass filtered radiation (a), the low pass filtered elevation (b), the low pass filtered radiation and the low pass filtered elevation (c), and the canopy cover, low pass filtered radiation and low pass filtered elevation (d).	35
Figure 2.4: A comparison between the r^2 values for the different regressions predicting the average temperature for each 30-minute period throughout the day (based on 39 temperature sensors at ground level). Elevation refers to the regression using the unfiltered elevation (m) taken directly from the DEM. El1000Rad1000 refers to the regression using the elevation (m) and radiation (W/m^2) as averaged over a radius of 1000 m. CanElRad refers to the regression using the canopy cover, unfiltered	

elevation (m) and unfiltered radiation (W/m^2). CanEl1000Rad1000 refers to the regression using canopy cover, the low pass filtered elevation (m) and the low pass filtered radiation (W/m^2) as averaged over a radius of 1000 m. Difference refers to the difference between the r^2 values of the two regressions involving canopy cover, with positive values indicating that the filtered result is higher.	37
Figure 2.5: A comparison between the predicted average summer minimum temperature based on elevation (top left) and the predicted average summer maximum temperature based on the low pass filtered elevation and low pass filtered radiation (bottom right). Whilst the minimum temperature south of Mt Kembla is similar to the minimum temperature north of Mt Keira ($15.5^\circ\text{C} - 16.5^\circ\text{C}$), the maximum temperature appears to be $1^\circ\text{C} - 2^\circ\text{C}$ cooler ($20^\circ\text{C} - 21^\circ\text{C}$ versus $21^\circ\text{C} - 23^\circ\text{C}$).	40
Figure 3.1: The Illawarra Escarpment, 80 km south of Sydney, Australia. The inset shows the topography of the study area as a digital elevation model, with the 600 vegetation survey sites and 40 temperature-monitoring sites shown by black and white circles respectively.	52
Figure 3.2: From top to bottom are the statistics for maximum (a-b), minimum (c-d) and average (e-f) temperatures. The left hand panels (a, c, e) illustrate the spatially averaged temperatures from 40 temperature monitoring sites. The right hand panels (b, d, f) illustrate the correlation (r^2) between the temperatures at each site and elevation, and the correlation between temperatures and a model containing elevation, moisture, wind and radiation. Each point is the mean (s.d.) that was produced by averaging the data from that time of year (2004/5 and 2005/6, surface and subsurface measurements).	61
Figure 3.3: From top to bottom are the statistics for maximum (a-b), minimum (c-d) and average (e-f) temperatures. The left hand panels (a, c, e) illustrate the influence of radiation, elevation, moisture and exposure in the temperature regressions. Each point is the mean (s.d.) that was produced by averaging the data from that time of year (2004/5 and 2005/6, surface and subsurface measurements). The influence is determined by multiplying the coefficient of each predictor by the range of values observed over the 40 temperature monitoring sites. The right hand panels (b, d, f) illustrate the direction of the exposure predictor that was selected to maximise the correlation with elevation.	62

Figure 3.4: 132 SDMs were produced for each species based on different temperature estimates. The performance was assessed in terms of the amount of deviance explained (y-axis), and a model containing elevation and geology was compared against models containing temperature and geology. Each point is the mean (s.d.) that was produced by averaging the data from that time of year (2004/5 and 2005/6, surface and subsurface measurements). Data in each panel is for the first species listed, while the species in parentheses produced similar results.	66
Figure 3.5: The surfaces of elevation, summer minimum, summer maximum and winter minimum temperatures. The temperature surfaces ($^{\circ}\text{C}$) were calculated by averaging all surfaces for each season (2004/5 and 2005/6, surface and subsurface measurements, 4 recording periods). Presence data for two species has been included on the surfaces that explained the most amount of deviance in their models.	68
Figure 3.6: A hypothetical dataset created by assuming temperatures decrease at a rate of $6^{\circ}\text{C}/1000\text{ m}$, but with fluctuations (mean = 0, s.d. = 2°C) introduced to simulate variations in exposure. Two regressions have been performed – one over the full range of 3000 m, and one over only 1000 m.	79
Figure 3.7: A comparison between northwesterly aspect (left) and northwesterly exposure (right). Aspect is more sensitive to DEM errors and does not consider topographic shading.	80
Figure 4.1: A hypothetical example for a study area with two geological units, G1 and G2, and presence/absence data for a selected species. Panels indicate the predicted probability of existence (POE) produced by a Generalised Additive Model without (a) and with (b) contextual indices.	88
Figure 4.2: The topography of the Illawarra Escarpment in the vicinity of Wollongong, Australia (34.4°S , 150.9°E). The inset is a Digital Elevation Model showing the rising elevation (m) from the coastal plain to the Woronora Plateau, with Mt Keira and Mt Kembla protruding eastward. See Supplementary Material (Section 4.5.1) for maps of geology and vegetation.	91
Figure 4.3: A comparison between the deviance explained by single-predictor models with and without contextual indices. The different symbols represent the eight species used in the study: gully gum (GG), mountain grey gum (MGG), coastal white box (CWB), blackbutt (BB), silvertop ash (SA), red bloodwood (RB), red	

cedar (RC) and lilly pilly (LP). There are seven points for each species – one for each of the seven environmental factors.	98
Figure 4.4: The deviance explained by the contextual models for lilly pilly (a), mountain grey gum (b) and red cedar (c) as a function of the radius of the contextual indices. The seven predictors are geology (Geol), elevation (Elev), summer maximum temperature (MaxER), shelter from WNW, S and NE winds (WindWNW, WindS, WindNE), and log of distance to stream (LogDS). Missing values occur due to problems with model convergence.....	99
Figure 4.5: A comparison between the average deviances explained by multiple-predictor models, with and without contextual indices, for eight tree species in the Illawarra region.	101
Figure 4.6: The probability of existence (POE) maps produced by the multiple-predictor models for <i>E. cypellocarpa</i> using predictors of elevation, shelter from southerly winds, geology and shelter from west to northwesterly winds. Circles represent the presence/absence data from the 360 sites that were used to train both models, while the triangles represent the additional sites that were used to test predictions in another area. The model produced without using contextual indices (Non cont., (a)) uses the same environmental variables as the contextual model (b).....	102
Figure 4.7: The estimated average spatial errors of (a) false positives (FP) and (b) false negatives (FN) as determined by averaging the distances (Ave. dist.) to the nearest presences and absences respectively. The two lines illustrate the results for the multiple-predictor contextual and non-contextual models for <i>E. cypellocarpa</i>	105
Figure 4.8 (next page): The geology (left) and vegetation communities (right) present in the study area. Geology map is provided by the NSW Department of Primary Industries. Vegetation map is from NPWS (2002). Black and white spots in the vegetation diagram illustrate the locations where the vegetation survey was conducted. See Section 4.5.1 for definitions and descriptions of each geology and community.	116
Figure 4.9: The optimal radius of the contextual indices for seven predictors, as determined using eight canopy tree species in single predictor models. The diamonds show the mean and standard deviation. The boxes show the median and quartiles. The seven predictors were geology (Geol), elevation (Elev), the average summer maximum temperature (MaxER), the log of distance to streams (LogDS),	

and exposure to northeasterly, southerly and west to northwesterly winds (WindNE, WindS, WindWNW).	118
Figure 4.10: The optimal radius of the contextual indices for eight tree species in the Illawarra region, as determined using seven environmental factors in single predictor models. The diamonds show the mean and standard deviation. The boxes show the median and quartiles. The eight species were gully gum (GG), mountain grey gum (MGG), coastal white box (CWB), blackbutt (BB), silvertop ash (SA), red bloodwood (RB), red cedar (RC) and lilly pilly (LP).	118
Figure 4.11: The response curves generated by a four-predictor GAM for <i>E.</i> <i>cypellocarpa</i> using contextual indices. Each predictor represents the average habitat quality within 500 m, where 100% habitat quality is defined as either: Elevation between 320 m and 420 m (EL500); Exposure to southerly winds of between 0.8 and 1.4 (S500); Geology of either 'TRnbu' or 'TRnz' (G500); or, Exposure to west to northwesterly winds of between 0.5 and 1.0 (W500).	119
Figure 5.1: The topography of the study area (right) as an elevation between 0 and 573 m. Superimposed are 40 locations where temperatures were recorded with iButtons, and 600 sites where a vegetation survey was conducted. Daily Bureau of Meteorology observations were obtained from six nearby locations (left).	126
Figure 5.2: Eight weather patterns were identified for the study area. Each weather pattern is represented as the average northerly (a) and easterly (b) components of the wind across the six BoM weather stations (wind speed \times sine/cosine of wind direction), average minimum and maximum temperatures (c), the average humidity at 9am and 3pm (d). Panel (e) illustrates the seasonal change in weather patterns using the percentage of days in each month that belonged to each of the eight weather patterns (average between 1973 and 2006). Panel (f) shows the inter- annual change in the annual number of days in weather patterns 2 and 8 over that period.	132
Figure 5.3: The relationship between average slope of the air-soil temperature regressions and average humidity, where each point corresponds with one of the eight weather patterns used in this study.	134
Figure 5.4: The mean (and standard deviation) amount of warming (1972-2007) estimated at 40 sites where soil temperatures were recorded.	135
Figure 5.5: The spatial distribution of warming (1972-2007) in the study area (Figure 5.1) as estimated by regressing the amount of warming at the 40 sites where	

temperatures were recorded against environmental factors. All surfaces are in degrees Celsius.....	137
Figure 5.6: The average amount of warming (1972-2006) was estimated for 37 species (Table 5.1) using ten seasonal temperature gradients (Table 5.2). The relative warming for each species on each gradient was calculated on a linear scale from 0% (least warming of all species) to 100% (most warming of all species). This graph illustrates the mean (standard deviation) relative warming for each species over the ten gradients.	140
Figure 6.1: The topography of the Illawarra Escarpment, 80 km south of Sydney, Australia (34.4 °S, 150.9 °E). A vegetation survey was conducted at the 600 sites indicated.	151
Figure 6.2: The results of Maxent models for 37 species (Table 6.1) as evaluated using the Area Under the receiver operating characteristics Curve (AUC) on a random validation 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 ‘all+cont’ model used the predictors from the ‘all’ model, but added a contextual index as an additional predictor. 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.....	156
Figure 6.3: The non-contextual (top) and contextual (bottom) Maxent models for <i>Eucalyptus pilularis</i> (Blackbutt). The non-contextual model is based only on environmental niche factors (temperature and geology), while the contextual model also includes a spatial predictor based on the average model output within 500 m. The presence or absence of <i>E. pilularis</i> illustrates that the species was only observed in the north and east of the area, but the non-contextual model predicts moderate quality habitat in the south and west.	160
Figure 6.4: Panel (a) indicates the location of moist rainforests on the Illawarra Escarpment (Coachwood Warm Temperate Rainforest or Illawarra Escarpment Subtropical Rainforest; NPWS 2002). Panel (b) indicates the locations where summer maximum temperatures are less than 20 °C, and panel (c) indicates where	

summer minimum temperatures – $0.4 * \text{winter minimum temperatures}$ is less than 12.8.....	161
Figure 7.1: A comparison between the distribution of <i>Eucalyptus cypellocarpa</i> (black dots) and summer maximum temperature ($^{\circ}\text{C}$) at the continental (left) and landscape (right) scales. Continental scale data was obtained from the Global Biodiversity Information Facility (http://data.gbif.org) and ANUCLIM (Houlder et al. 2000). Landscape scale data was obtained from my study on the Illawarra Escarpment (Ashcroft et al. 2008; Chapter 3).....	170
Figure 7.2: The graphs illustrate simulated results from standardised (top) and unstandardised (bottom) weather stations. Standardised weather stations have limited noise introduced by factors such as vegetation and topography, and show a strong trend towards factors such as elevation. Unstandardised weather stations are noisier, because temperatures are influenced by a variety of factors. This means that it is more difficult to compare different locations, but it is now possible to work out how multiple factors affect temperatures (see text for more information).	174
Figure 7.3: The locations where summer maximum temperatures are less than 20.5°C are shown in red, while the locations of moist rainforest communities are shown as black polygons. Temperature data is from Ashcroft et al. (2008; Chapter 3), while vegetation data is from NPWS (2002).	176

Abstract

Mapping and explaining the distribution of vegetation helps land managers to make systematic conservation planning decisions. This is typically achieved using models that correlate the distribution of species with environmental factors, and can predict the vegetation at unsurveyed locations. These Species Distribution Models (SDMs) have numerous unresolved issues, but serve as a useful first-pass approximation for planning purposes.

This thesis investigates some of the uncertainties of SDMs, including the impact of data accuracy, the incorporation of spatial processes, the evaluation of alternative models, and the benefits and challenges of producing models at the landscape scale. The research was conducted on the Illawarra Escarpment, 80 km south of Sydney, Australia (34.4 °S, 150.9 °E). The escarpment contains a north-south trend in eucalypts (*Eucalyptus spp.*) that cannot be explained in terms of elevation or geology. It also exhibits a patchy distribution of rainforest communities, some unique to the Illawarra. It is not known which environmental factors determine the distribution of either the eucalypts or rainforest species, or how they may respond to a changing climate.

Species distributions are sensitive to the accuracy of data used, and yet many models only use elevation as a surrogate for temperature, or use simple elevation sensitive interpolations from weather stations. I collected hourly temperature data from 40 sites on the Illawarra Escarpment, and investigated whether elevation was an adequate surrogate for temperatures in this landscape. I then investigated whether temperature surfaces could be improved by considering other topographic and geographic factors, including exposure to wind, distance to coast, radiation, and the average conditions in the surrounding neighbourhood. Elevation was well correlated

with moderate seasonal temperatures (e.g. summer minima and winter maxima), but was poorly correlated with the extreme temperatures (summer maxima, winter minima) that are physiologically limiting for many species. Using neighbourhood influences, exposure to wind and distance to coast improved the accuracy of temperature surfaces, and increased the explanatory performance of vegetation models. I concluded that elevation was not always an adequate surrogate for temperature. Temperatures are also affected by other topographic and geographic factors, and these should be considered when developing models for systematic conservation planning activities.

Species distribution models are typically based solely on niche factors. Where spatial processes are included, it is typically by employing autologistic regression, or other techniques that use survey data as a predictor. This precludes the models being used to make predictions in times or places where survey data is unavailable, and reduces ecological explanation because it is an interpolation technique. I used neighbourhood (contextual) indices based on environmental factors as an alternative method to overcome these problems. I demonstrate that contextual indices improve SDMs over purely niche-based models, and are capable of predicting unknown populations in unsurveyed areas. I conclude that contextual indices have numerous advantages over autologistic regression, and can capture a continuum between niche and dispersal limited species.

Models that predict how species will respond to climate change either use coarse-scale climate surfaces, or idealised predictions of uniform warming. These methods may dramatically over-estimate extinction risk because they neglect fine-scale variations in warming, and refugia where species can persist despite unfavourable regional conditions. I created fine-scale estimates of warming by combining 35 years of Bureau of Meteorology observations with one year of intensive fine-scale temperature

monitoring. I found that warming was greatest at inland locations, at lower elevations, away from streams, and at sites exposed to hot, dry northwesterly winds. As species are biased in the geographic and topographic positions they occupied, some species have experienced more warming than others and are at greater threat from climate change. I concluded that it was important to continue developing methods to downscale coarse-grained climate surfaces, and suggest that the accuracy of this process could be improved by using a range of topographic factors.

There are many methods for selecting predictors in SDMs, and the competing models often make highly variable predictions. I addressed this uncertainty by comparing the performance of models with and without a given environmental factor. I found that there was relatively strong support for the geology and winter minimum temperature predictors, as well as predictors based on contextual indices, as there was a significant drop in model performance when these predictors were excluded. In contrast, there was less support for summer maximum temperature, as other temperature predictors could combine to produce similar model performance. Model performance varied more between models for different species than between different predictor combinations for the same species. I concluded that it was inappropriate to assess models based on subjective benchmarks, such as an AUC of more than 0.7. A comparison between competing models for the same species gives a better indication of the validity of the model building procedure.

The results of this research provide important insights into the benefits and challenges of creating SDMs at the landscape scale (extent of 10–200 km). It is a major challenge to obtain spatially and thematically accurate environmental predictors and biotic data at this scale, and studies should include the collection of data to ensure models are adequate. Landscapes will not have as much environmental variation as

coarse-scale models, and this will limit the ability to transfer the models to new study areas. However, there are a number of benefits that justify these studies. Producing accurate temperature surfaces at the landscape scale will result in less pseudoreplication and less predictor colinearity. This will improve the robustness of models. Landscape scale studies also allow modellers to capture fine-scale refugia, and this will improve the accuracy of climate change predictions. Finally, many ecological processes operate at a scale that is too fine to be detected with coarse-scale models. Landscape scale models may be the only alternative to detect these processes. There is no optimal scale for SDMs, however, and a future challenge is to better integrate coarse and fine-scale models to make more ecologically robust predictions.

Acknowledgments

The journey towards this thesis began in late 2001, when I decided to move away from my established career in software design. Firstly, I would like to thank the University of Wollongong for supporting this career change, especially John Morrison for encouraging me to do a Masters of Environmental Science in 2002/3, and later to my PhD supervisors, Laurie Chisholm and Kris French, who always had confidence in my ability to complete this thesis. Even when times were tough, they were always positive and never doubted that I was on the right track. I am especially grateful for all the time and effort Laurie and Kris spent reviewing my journal articles, especially when I was employing statistical techniques that were outside their core area of expertise. I also thank my supervisors for introducing me to the iButton temperature loggers that were so crucial for my research.

The fieldwork component could not have been completed without Kris French, Bernard Ashcroft, Belinda Pellow, Anders Bofeldt and the Janet Cosh Herbarium, who helped me learn how to recognise the species I was targeting. I also thank Bernard Ashcroft, Imke Büsing, Berit Fehlau, Jodie Dunn, Carolyn Blomley, Anders Bofeldt, Kris French, Laurie Chisholm, Roger Ferguson, Elfi Ashcroft, Roger Starkie, Andrea Kubler and Rike Arnold for assistance with the fieldwork, and putting up with the leaches, ticks, snakes, dense vegetation and rugged topography. The research was critically dependant on access to the sites that I wanted to survey, and I am grateful to the Sydney Catchment Authority, National Parks and Wildlife Service, Sydney Water, BHP Billiton, Wollongong City Council, and the numerous private landowners who granted me permission to access their land.

Thanks to everyone who has reviewed and provided comments on the journal articles resulting from this research, including my supervisors, Jane Elith, Brian Lees, John Gollan, John Leathwick, Miguel Araújo, Philip Dixon, Bernard Ashcroft and many anonymous referees. John Gollan also provided useful discussion and references, which helped me to write the discussion chapter.

My research was supported by a University Postgraduate Award scholarship between 2004 and 2006. The Spatial Analysis Laboratory at the University of Wollongong was crucial for all the analysis, as was the data that was supplied to them by AAMHatch, the NSW Department of Environment and Climate Change and the NSW Department of Primary Industries. Estimates of radiation were developed using the freeware Solar Analyst extension for ArcView that is made available by the USDA Forest Service, and Maxent is available from the Princeton University webpage. From 2007 to 2009, I was working full-time at the Australian Museum, and I am grateful to them for allocating some of my time towards the completion of my PhD. Both the University of Wollongong and Australian Museum assisted me to present my results at conferences.

Last, but not least, I thank my wife and family for putting up with my heavy workload and stress— especially over the final few years whilst I was working full-time, commuting to Sydney, and still trying to find time for them and my PhD. I gained two beautiful daughters, Lauren and Kira, during the course of this research, and they have provided me with wonderful memories and inspiration to complete the thesis.

Chapter 1: Introduction

1.1 Modelling the spatial distribution of vegetation

Modelling the spatial distribution of vegetation assists environmental managers to design reserve systems and pursue other aspects of systematic conservation planning (Margules and Pressey 2000). These spatial distributions are especially important at the regional and landscape scales, as this is the scale at which most environmental management decisions are made (Ferrier et al. 2002; Lookingbill and Urban 2003).

The distribution of vegetation can be modelled using an increasingly large number of statistical methods (e.g. Elith et al. 2006). These methods can be broadly separated into either dynamic, mechanistic, process-based models; or static, correlative, niche-based models (Guisan and Zimmermann 2000).

Dynamic, mechanistic or process-based models have been presented as being a more robust method for modelling species distributions, as they are based on species physiology and consider transient responses and dispersal limitations (Loehle and LeBlanc 1996; Hampe 2004). However, mechanistic models are not feasible for most species, because there is insufficient knowledge about these traits (Guisan and Zimmermann 2000). Indeed, correlative niche-based models may be less reliably linked with species physiology and stochastic processes, but serve as a useful first pass approximation when there is insufficient knowledge for mechanistic models (Pearson and Dawson 2004).

1.1.1 Dynamic, mechanistic or process-based models

There are many types of dynamic, mechanistic or process-based models (hereafter simply mechanistic models) that can be used to model the distribution of species.

Arguably the simplest of these is a Markov chain model (Balzter 2000). Markov chain models consist of a set of spatially discrete locations, each of which is in one of a discrete number of states at any point in time. A matrix is determined from a time series of observations that defines the transition probabilities for each combination of states. That is, the probability that a location in state x at time t will be in state y at time $t + 1$, for all combinations of x and y . Markov chain models are usually based on the assumption that transition probabilities are constant in time, independent of the state of the surrounding locations, and independent of previous states, although this need not be the case (Balzter 2000).

Markov chain models are perhaps best suited to modelling systems where there is a fixed and deterministic successional process (Childress et al. 1998). The biggest drawbacks are that transition probabilities are not necessarily constant from year to year (Childress et al. 1998), and a long time-series of data is necessary to calibrate models (Dale et al. 2002).

Another type of mechanistic model is called a cellular automata model (Balzter et al. 1998). While Markov chains ignore spatial context, cellular automata determine the state at time $t + 1$ using the state of adjoining locations at time t . These approaches are somewhat complementary, and cellular automata and Markov chain methods have been combined to produce spatio-temporal Markov chains (Balzter et al. 1998). Cellular automata are useful when neighbouring locations interact strongly, such as when there are invading species, or when dispersal or competition has a strong effect (Arii and Parrott 2006). They can also be used when land-use changes are spatially autocorrelated, such as the clearing, farming, abandonment and succession of Amazonian rainforests (Soares-Filho et al. 2002). Like Markov chains, cellular automata models require a large amount of data to calibrate models. Both models are

perhaps best viewed as simple models of the conceptual dynamics, and not a precise description of how the ecosystem behaves (Ruxton and Saravia 1998).

The simplicity of Markov chain and cellular automata models is more evident when they are compared with process-based forest gap models or landscape models. These models typically model at least some aspects of the lifecycle of individual trees (or small cohorts), including growth (Schumacher et al. 2004); mortality (Bigler and Bugmann 2004; Schumacher et al. 2004; Wunder et al. 2006); seed dispersal and germination (Liu and Ashton 1998; He et al. 1999; Bleher et al. 2002); herbivory (Weisberg and Bugmann 2003); and disturbances such as fire (Roberts 1996; Schumacher et al. 2006), windthrow and forestry (He et al. 2002). There are usually numerous assumptions, estimations and simplifications that are needed for the parameterisation and definition of these models, and therefore it is necessary to quantify the uncertainty in model output that is due to potential errors in the model inputs (Crosetto et al. 2000; Lexer and Hönninger 2004).

Another type of mechanistic model is non-dynamic, but relates suitable habitat to the physiology of the species (Kearney and Porter 2004). These models determine the fundamental niche by experimentally determining the range of environmental conditions that are suitable for a species' reproduction, movement, or other critical life phases. These models are different from other mechanistic models in that they are non-dynamic, but they are further differentiated from correlative models in that they are based on mechanisms rather than correlations between the environment and species current distributions.

None of the mechanistic models discussed in this section are used in this thesis, however it was important to consider them as alternative methods. They were not used primarily because there was insufficient data available to estimate how growth,

germination, mortality, and seed dispersal processes varied with environmental conditions for each of the species I considered, and it is difficult to experimentally determine these factors for long-lived tree species. Not all species reproduce every year (e.g. mast flowering in rainforest species), and reproduction for some species may be higher after periodic disturbances such as fire.

1.1.2 Static, correlative, niche-based models

A number of statistical methods have been used to correlate the current distribution of species with the corresponding environmental conditions (e.g. Elith et al. 2006). These have been collectively referred to by a number of terms including species distribution models, bioclimatic envelope models, ecological niche-models, or static-equilibrium models. They have been thoroughly reviewed and discussed in the ecological literature (e.g. Guisan and Zimmermann 2000; Austin 2002; Guisan et al. 2002; Rushton et al. 2004; Guisan and Thuiller 2005; Araújo and Guisan 2006; Guisan et al. 2006a; Austin 2007), and are not reviewed further here. Instead, the following sections establish the context for the chapters in this thesis by discussing how the techniques evolved over the lifetime of this research. Species distribution models, or SDMs as they will now be referred to, are evolving rapidly, and techniques which were considered best-practice at the start of this research are now under greater scrutiny and some considered out-dated. I briefly outline some of these developments to illustrate the context in which the various chapters were prepared, and to better demonstrate the contributions that the chapters made to advance the field of research.

1.1.3 Statistical methods for developing SDMs

All statistical methods used to produce species distribution models are based on correlations between species' distributions and environmental conditions. It is usually

assumed these relationships are causal, and can be used to predict the invasion potential of species (e.g. Peterson 2003) or how species will respond to climate change (e.g. Thomas et al. 2004). In broad terms, the methods differ in how complex the responses to environmental gradients can be. For example, simple and early methods, such as BIOCLIM, have fixed limits on each gradient that are independent of other gradients (see Guisan and Zimmermann 2000 for a comparison between methods that were commonly used at that time). Generalised Linear Models (GLMs) were developed later, and allowed the responses to be various parametric shapes, including Gaussian curves and β -functions (Austin et al. 1994). Generalised Additive Models (GAM; Hastie and Tibshirani 1990) are more flexible than GLMs, as they are non-parametric, and smoothing functions allow the data to determine the shape of the response functions. GLMs and GAMs were viewed more favourably than other methods when I began this research, and were the focus of international workshops conducted in 2001 and 2004 (Guisan et al. 2002; Guisan et al. 2006a).

During the course of this research, a number of new methods were developed, or gained popularity, including Maxent (Phillips et al. 2004; Phillips et al. 2006; Phillips and Dudík 2008), Generalised Dissimilarity Modelling (GDM; Ferrier et al. 2007) and Multivariate Adaptive Regression Splines (MARS; Leathwick et al. 2005). These methods became especially popular after they performed well in a comparative study against the more established methods, including GLMs and GAMs (Elith et al. 2006). Importantly, however, these methods have not been tested for ecological realism, and only evaluated using data from the same study area as the models were developed (Elith et al. 2006). This is not a truly independent dataset to make general inferences, and these complex models may be overfitting to the study area and have poor transferability (Peterson et al. 2007).

The initial chapters in this thesis are based solely on GAMs (Chapters 2–4), although methods did ‘evolve’ to include Maxent (Chapter 6). While GAMs only performed moderately in the comparative study (Elith et al. 2006), they are potentially less prone to overfitting than more complex methods, and may be superior when validations are performed with truly independent data (i.e. different study area and time). No method is unanimously considered superior, and GAMs and Maxent are still powerful and valid methods to produce SDMs.

The evolution of statistical methods is likely to continue. New methods have already been suggested, including quantile regression, structural equation modelling and geographically weighted regression (Austin 2007). Some scientists are combining the results of models produced with different methods to produce a consensus estimate (Araújo and New 2006). Further testing with truly independent data may shed new light on the balance between statistical complexity and overfitting, although obtaining truly independent data is impossible in many situations (Austin 2007), and testing using artificial data may be necessary (Austin et al. 2006).

1.1.4 Model evaluation

All ecological models are simplifications of reality, and hence contain errors. The magnitude of these errors needs to be quantified before models can be reliably used for conservation purposes. SDMs generally produce some sort of habitat suitability index as a continuous variable, and a threshold must be used if this is to be converted to a binary map of suitable or unsuitable habitat (or predicted presence or absence). Therefore, model evaluation methods can be broadly separated into those that use a threshold, and those that are independent of a threshold (Fielding and Bell 1997).

For the duration of this research, the ‘best-practice’ method for model evaluation was considered to be Area Under the receiver operating characteristic Curve (AUC;

Pearce and Ferrier 2000b; Austin 2007). This was largely because it is insensitive to the subjective selection of a threshold. However, towards the end of the research, doubt was raised over the effectiveness of the method. Austin (2007) suggested that models with similar AUC could produce vastly different predictions, and Lobo et al. (2008) showed it neglected spatial errors, ignored differences between errors in omission and commission, and was sensitive to the extent of the study area.

In this thesis I provided further evidence that AUC is an insufficient measure of predictive success, and developed new methods to address some of its shortcomings. In Chapter 4, I complimented AUC by also employing other validation techniques, including a novel method to consider the spatial context of errors. In chapter 6 I highlight the dangers of using subjective benchmarks to evaluate model performance, such as an AUC of more than 0.7. Most candidate models for a given species had similar AUC values, and therefore AUC was not sensitive to the predictor selection and model building process.

Chapters in this thesis that were focused on producing temperature surfaces (Chapters 2, 3 and 5) avoided the shortcomings of AUC by not performing a comprehensive model evaluation. They only demonstrating that correlations were improved (in terms of the deviance explained, or D^2). The approach taken throughout this thesis is to compare alternative models for the same species with each other, and provide evidence as to which models have stronger correlations than others. I accepted that evaluation against an 'independent' dataset from the same study area did not protect against overfitting, and instead avoided overfitting by limiting the complexity of response curves and the number of predictors in models. Model evaluation is an unresolved issue (Guisan et al. 2006a), and I do not claim that the evaluations in this thesis are complete. Nevertheless, the thesis makes a valuable contribution to improving

validation techniques by introducing new validation methods (Chapter 4) and highlighting the shortcomings of existing methods (Chapter 6).

1.1.5 Issues with species distribution modelling

Species distribution models are based on a number of assumptions. For example, models assume that species are at equilibrium with the environment, and that the distribution of species is causally determined by environmental factors such as climate (Guisan and Zimmermann 2000). These assumptions are not usually tested when models are developed, yet results are used to estimate extinction risk from climate change (Thomas et al. 2004), determine whether species are occupying their full range (Svenning and Skov 2004), determine if species are niche or dispersal limited (Moore and Elmendorf 2006), predict the extent of species invasions (Peterson 2003), or to select areas for species persistence in protected areas (Araújo and Williams 2000). The conclusions of all these studies may be incorrect if the assumptions on which the SDMs are based are invalid. There are numerous issues with the assumptions and simplifications surrounding SDMs (e.g. Loehle and LeBlanc 1996; Hampe 2004), and these have received increasing attention.

Some of the less critical issues that have received attention concern which datasets are suitable for modelling. For example, investigating the appropriate scale (Scott et al. 2002; Graf et al. 2005; Guisan et al. 2007; Trivedi et al. 2008), sample size (Stockwell and Peterson 2002; Coudun and Gégout 2006), and method for selecting pseudo-absences (e.g. Zaniewski et al. 2002; Engler et al. 2004) does not cast doubt over SDMs as a whole, but determines which datasets are most suited to use in SDMs.

Recent studies have cast more serious doubts over SDMs in general, and the remainder of this section will detail five key issues with SDMs that have arisen in the scientific literature. These issues are central to this thesis, and the remaining chapters

provide new insights into these issues and develop new methods to improve the robustness of SDMs.

The first issue concerns the assumption that correlations between species distributions and environmental factors are causal. Bahn and McGill (2007) demonstrated that randomly generated, but spatially structured, environmental variables were also able to satisfactorily explain species distributions, and this suggests that many species-environment correlations may be spurious. Species distributions are often clustered, and environmental conditions are usually more similar in locations in close proximity. It is possible that species distributions are correlated with environmental factors simply because they are both spatially structured (Currie 2007).

The second issue concerns the underlying assumption that SDMs capture the ecological niche of species. SDMs are based on actual distribution data, which is a result of biotic interactions as well as environmental conditions. Therefore, SDMs are usually reported to capture the realised niche of the species (Austin 2002). However, SDMs do not necessarily capture competition unless it is included as an extra predictor (Leathwick and Austin 2001), and some suggest that SDMs actually capture the fundamental niche (Soberón and Peterson 2005). Others have suggested that SDMs do not capture the niche *per se*, as there is no functional or physiological relationship (Kearney 2006). Instead, it has been suggested that simple SDMs, for example those based on environmental envelopes, may capture the potential distribution, while more complex methods capture the realised distribution (Jiménez-Valverde et al. 2008). Clarifying the niche theory on which SDMs are based is one of the biggest challenges (Araújo and Guisan 2006; Guisan et al. 2006a), as it is difficult to justify SDMs when they do not have a strong ecological basis. A strong ecological grounding may provide

confidence that correlations are not simply due to the spatially structured environmental predictors.

The third issue with SDMs is the uncertainty in predictions. Predictions from SDMs vary dramatically according to the statistical method used, the scale of the analysis, the climate data used, and the parameters selected (Thuiller et al. 2004a; Thuiller 2004; Araújo et al. 2005a; Araújo and New 2006; Beaumont et al. 2007). This variability can be large enough to compromise the usefulness of the methods (Araújo et al. 2005a). The outputs from multiple models can be combined to produce a narrower range of predictions (Araújo and New 2006), however Beaumont et al. (2007) argue that this merely hides the uncertainty, and increases the risk of overestimating the confidence that should be placed in SDMs. Further research is needed to determine if the range of uncertainty can be reduced, and to develop methods that are better at conveying the inherent uncertainty of SDMs to users.

The fourth issue with SDMs, and one that has received comparatively little attention, is that they are sensitive to the accuracy of the data used (Dormann et al. 2008). Data errors can affect many aspects of the SDMs, including the statistical significance, the shape of response curves, the predictors that are selected, prediction accuracy, and the spatial extent of predictions (Van Niel et al. 2004; Van Niel and Austin 2007). In some respects, this is associated with model uncertainty, however data uncertainty has other implications with respect to the complexity of models that can be justified. For example, some of the better performing methods in the study of Elith et al. (2006) produce complex response curves to environmental factors. Given that the shape of response curves is sensitive to data accuracy (Van Niel and Austin 2007), it may be difficult to justify complex responses to environmental gradients that have large errors.

The fifth issue with SDMs concerns spatial processes and spatial autocorrelation. Spatial autocorrelation is problematic for SDMs as it can affect the apparent significance of predictors (Legendre 1993). However, autologistic regression (Miller et al. 2007) or contextual indices (Ferrier et al. 2002; Wintle et al. 2005) can be used to include spatial context in SDMs, and this may be an opportunity to account for dispersal, home range size, or species interactions. Indeed, if SDMs are purely niche-based, then they can be criticised for failing to consider spatial processes such as these (Hampe et al. 2004). Autologistic regression is not a complete solution to this problem, as it uses data from the response variable as a predictor. Therefore, autologistic regression becomes interpolation rather than explanation, and it cannot be used to make predictions at locations or times where survey data is unavailable (Segurado et al. 2006). In addition, autologistic regression can lead to biased models, and underestimate the effect of environmental variables (Dormann 2007). There is consensus that spatial processes need to be considered when modelling species distributions (e.g. Dirnböck and Dullinger 2004; Dullinger et al. 2004; de Frutos et al. 2007), yet the manner in which this should be achieved is an unresolved issue (Guisan et al. 2006a).

This list of issues is not complete, but demonstrates increasing attention to the limitations of SDMs in recent years. Other issues that have not been listed here include genetic variation within species (Hampe 2004), the difficulty in extrapolating SDMs into novel climates, and the potential effects of rising CO₂ levels (Loehle and LeBlanc 1996).

1.2 The vegetation of the Illawarra Escarpment

The Illawarra Escarpment is situated approximately 80 km south of Sydney, Australia (Figure 1.1; 34.4 °S, 150.9 °E). It is oriented northeast to southwest, and rises 300 to 600 m above the city of Wollongong on the coastal plain (Figure 1.2). Its well-vegetated

slopes create a conspicuous and contrasting backdrop to the city, and this has led to it being listed as a “scenic landscape of statewide significance” by the National Trust of Australia (NPWS 2003). The escarpment is an important component of the local community, as demonstrated by the Commission of Inquiry that was conducted to address its long-term management issues.

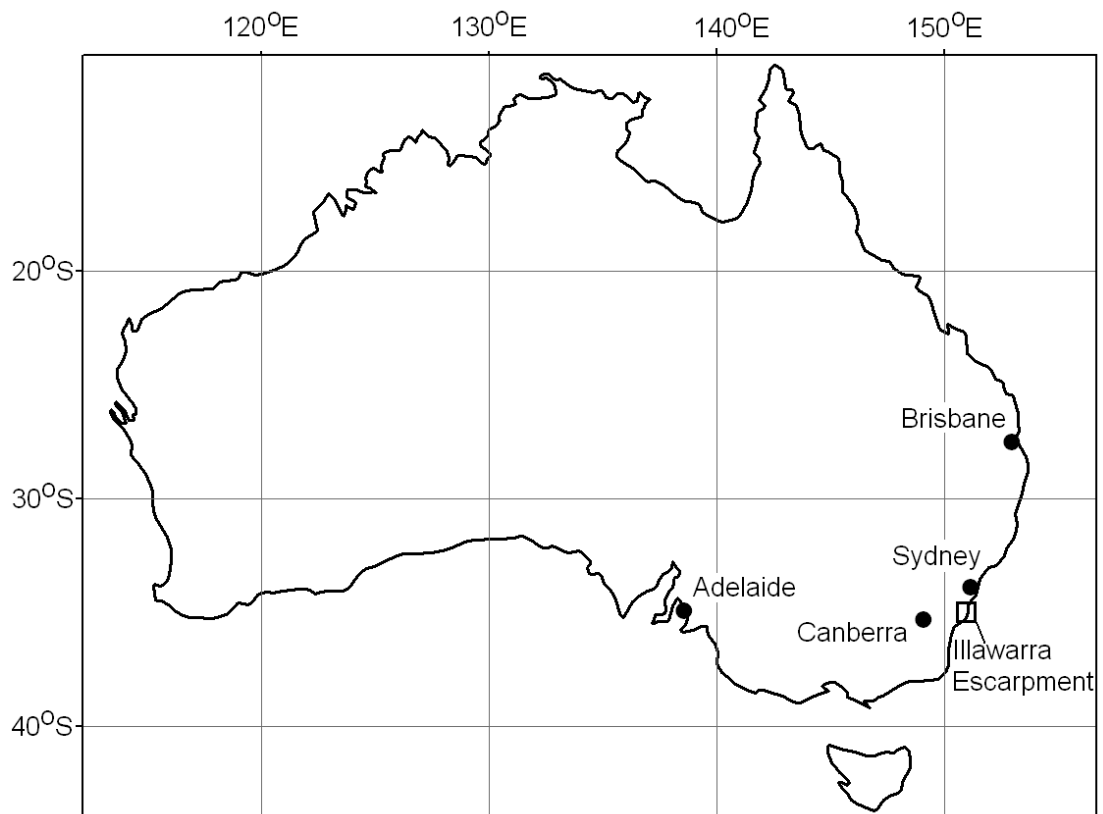


Figure 1.1: The location of the study area.

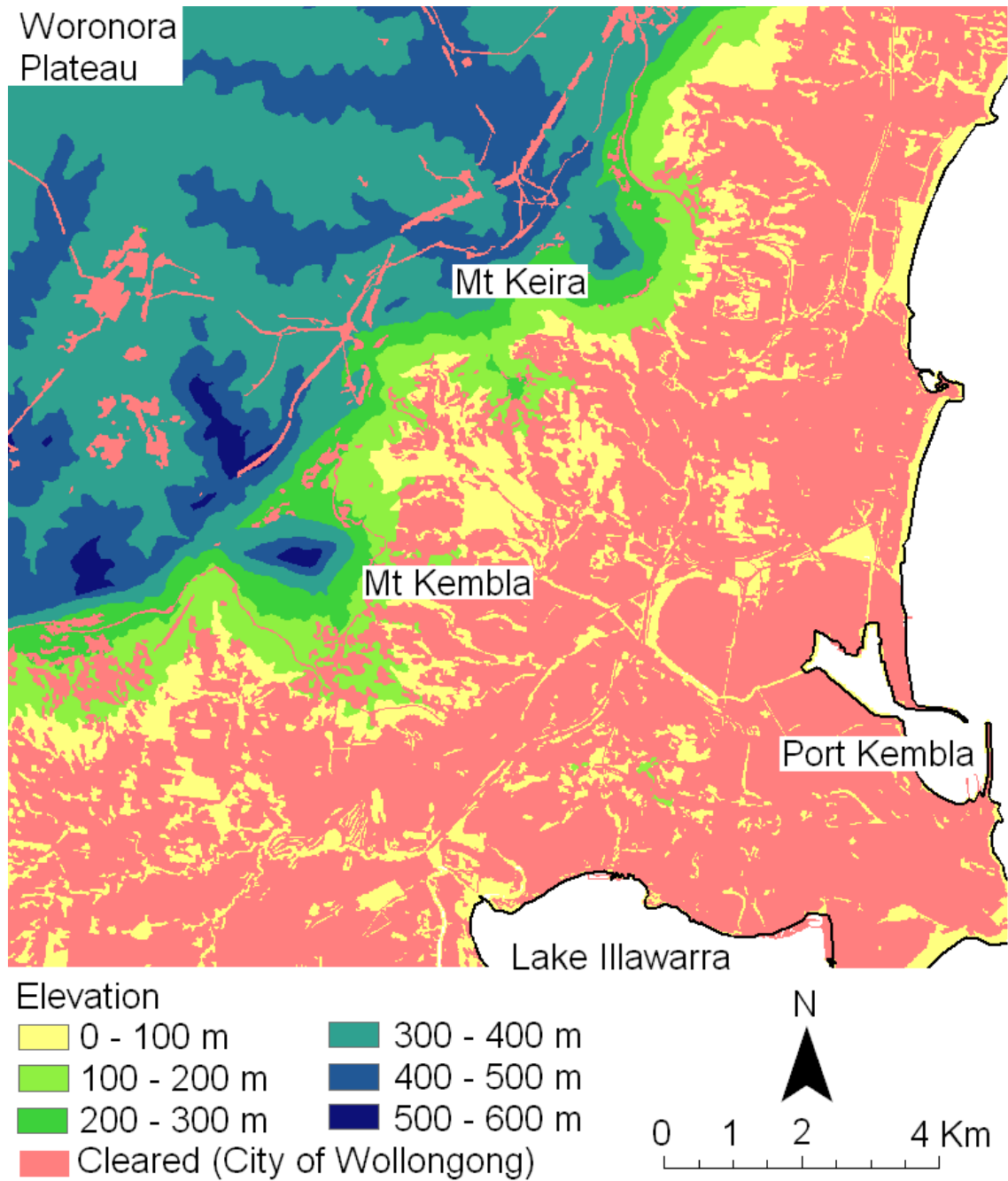


Figure 1.2: The elevation of the Illawarra Escarpment (yellow to blue), with the area cleared for the city of Wollongong shown in red.

Mt Keira and Mt Kembla protrude eastward from the escarpment (Figure 1.2) and create variations in aspect and shelter from winds and radiation. The combination of the topography and geology create an environment that is rare within the Sydney Basin

bioregion, and many of the vegetation communities are unique to the Illawarra (NPWS 2003).

Despite the uniqueness and importance of the local vegetation, there is still a poor knowledge of the factors that determine the distribution of species and communities. For example, there is a north-south trend in eucalypts, with Escarpment Blackbutt forest dominating the escarpment north of Mt Keira, Escarpment Moist Blue-Gum Forest dominating between Mt Keira and Mt Kembla, and Moist Coastal White Box forest dominating south of Mt Kembla (NPWS 2002; Figure 1.3). While local and regional literature provide distributional maps of communities (NPWS 2002), and a rough guide to where individual species are located (Fuller 1995), these do not explain this north-south trend in eucalypts, or other distributional patterns.

‘Please see print copy for image’



Figure 1.3: The north south trend in eucalypts that occurs on the Illawarra Escarpment (from NPWS 2002).

There are three types of rainforest on the Illawarra Escarpment, which have a relatively patchy distribution (Figure 1.4). Lowland Dry Subtropical Rainforest is an endangered ecological community that occupies exposed foothills and drier gullies at low elevations. Coachwood Warm Temperate Rainforest and Illawarra Escarpment Subtropical Rainforest occupy moister sites, typically at sites that are sheltered from hot, dry northwesterly winds (Fuller 1995). The rainforests of the Illawarra have received comparatively more attention than the eucalypt communities. Numerous studies at the University of Wollongong have examined the composition of rainforests (Mills 1986), and examined the effects of edaphic factors (Bywater 1985) and fire (Erskine 1984). These studies are consistent with broader scale studies that have examined the effects of these factors (e.g. Beadle 1954, 1966), but importantly, all are non-spatial. That is, they seek to determine whether or not a factor affects the composition or presence of rainforest communities or species, but they cannot be used to produce a map of rainforest distribution, or quantitatively explain or predict the spatial distribution in terms of environmental factors. This limits the ability to quantify distributional changes that will occur due to, for example, climate change or various disturbance events.

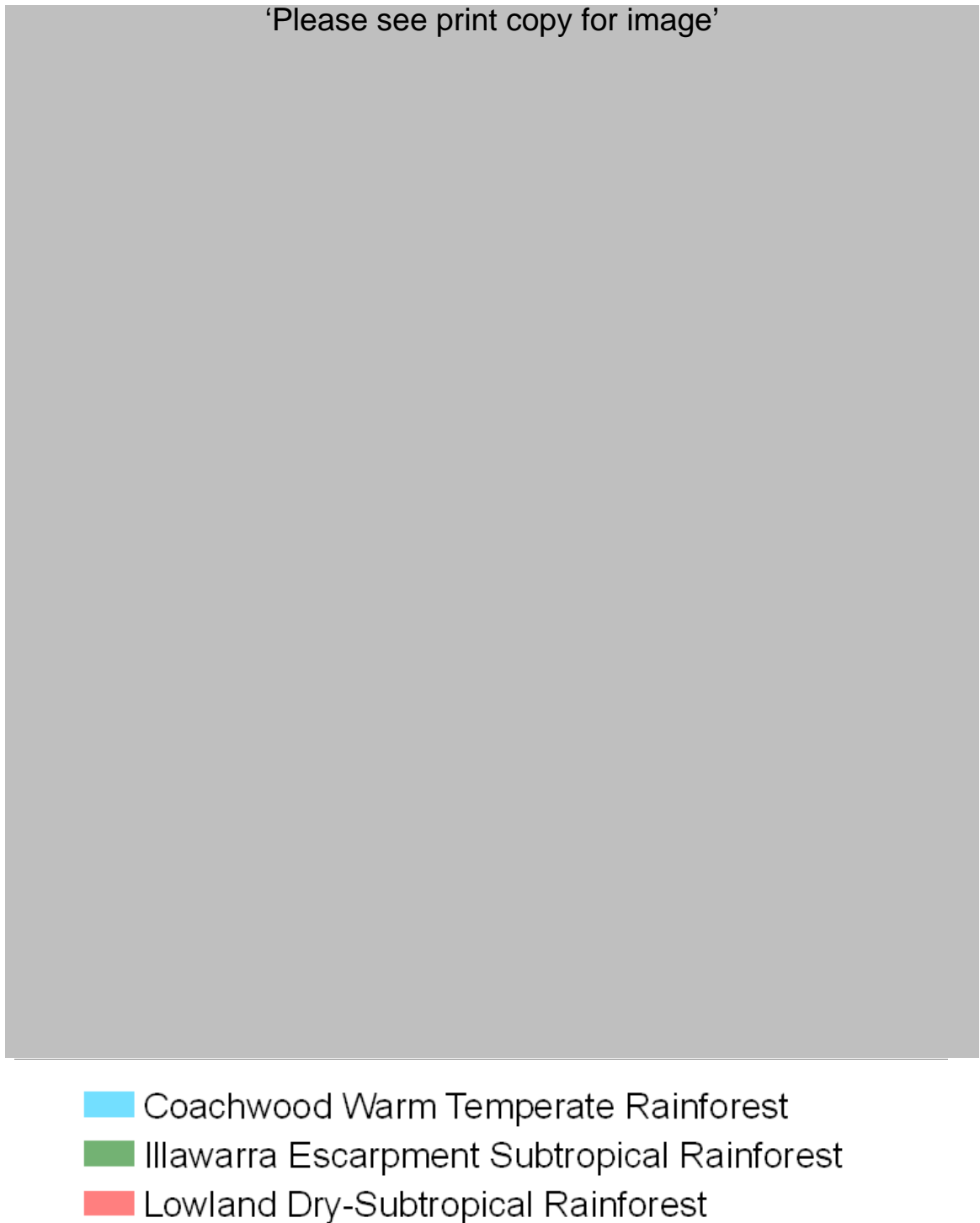


Figure 1.4: The patchy distribution of the moist (blue, green) and dry (red) rainforest communities that are located on the Illawarra Escarpment (from NPWS 2002). Elevation is depicted in the background as a grey-scale image (Figure 1.1).

1.3 Research aims and thesis outline

Many methods have been used to explain the relationships between species' distributions and environmental conditions. There have been times when scientists thought they had developed methods that could adequately identify these relationships, only to later realise that the assumptions behind these methods were not appropriate (Wiens 2002). To avoid this trap, I deliberately avoided trying to identify one 'true' model or 'correct' ecological relationship. Instead, the goal of this thesis was to assess and improve the methods used to produce SDMs, and contribute to the ongoing evolution of methods in the scientific literature. The remaining chapters focus on the five issues with SDMs discussed in Section 1.1.5—with a particular focus on data accuracy, spatial processes, and the variability of competing models. The specific aims were:

1. To produce accurate fine-scale temperature surfaces based on a variety of topographic and geographic factors, and determine if this improved the performance of SDMs. Most SDMs assume that elevation is an adequate surrogate for temperature, but I demonstrate that this assumption is not valid (Ashcroft 2006; Ashcroft et al. 2008; Chapters 2 and 3). I further demonstrate that increasing the accuracy of temperature surfaces improves the ability of SDMs to explain the distribution of vegetation.
2. To determine whether SDMs could be improved by considering the spatial context of predictors. In Chapter 4, I demonstrate the benefits of using contextual indices to incorporate spatial processes into SDMs, and how this overcomes the shortcomings of autologistic regression (see Section 1.1.5).

3. To produce accurate fine-scale surfaces of the climate change that has occurred over the last 35 years. I develop a novel approach to estimate landscape scale climate change, quantify the spatial variability in warming, and demonstrate that species differ in the amount of warming they have experienced (Chapter 5; Ashcroft et al. 2009). This research assists land managers to locate refugia from climate change and identify the species that are most at risk from climate change.
4. To investigate the uncertainty associated with species distribution models and how this relates to commonly used evaluation techniques such as AUC (see Section 1.1.4). In Chapter 6 I demonstrate that it is inappropriate to assess SDMs against benchmark performance standards, and present a method to compare competing models and increase confidence in the selected predictors.
5. To review how the landscape scale models produced in this thesis advance the scientific literature. The discussion chapter, Chapter 7, presents the challenges and benefits of producing SDMs at the landscape scale, using the methods developed in this thesis as examples. I discuss the benefits over models at other scales, and how they could be combined to improve the model building process.

The remaining chapters of this thesis have been published in scientific journals (Chapters 2, 3 and 5), or been formatted so that they can be submitted to journals in future (Chapters 4, 6 and 7). This formatting has resulted in some duplication of site descriptions and methods in each chapter, although this has been minimised through

cross references. The most comprehensive description of the study area is included in the supplementary material for Chapter 4, including the vegetation and geology. Section 4.5.2 provides the most detailed account of the vegetation survey and design. The most comprehensive assessment of DEM and data accuracy is included in the supplementary material for Chapter 3. The comparison between the sampling design I use to create temperature surfaces and existing coarse-scale methods is included in Section 7.4.

Chapter 2: A method for improving landscape scale temperature predictions and the implications for vegetation modelling¹

2.1 Introduction

Understanding the relationship between environmental factors and the distribution of vegetation can provide a meaningful contribution to environmental planning and management (Austin 2002; Ferrier et al. 2002). This is especially true at the landscape scale where environmental decisions are often made (Lookingbill and Urban 2003; Chuanyan et al. 2005). Quantifying the environmental niche of different species and communities can provide evidence as to what would occur if the land were used differently (Guisan and Zimmermann 2000), allows the estimation of past climate from fossils (Arundel 2005), can be used to aid ecological restoration (Chuanyan et al. 2005), and can be used to predict how future climate change will alter the distribution of vegetation (Hörsch 2003).

A common way of explaining the distribution of vegetation is through the static modelling of survey data (see Guisan and Zimmermann 2000 for a review). These models capture the realised niche of vegetation in terms of environmental variables, but are based on the assumption that the vegetation is a result of, and in equilibrium with, the current environment and not a relict from the past (Austin 2002).

Static models require detailed maps of the environmental factors that influence the vegetation, the majority of which are based on Digital Elevation Models (DEMs) and

¹ This chapter has been published as: Ashcroft MB (2006) A method for improving landscape scale temperature predictions and the implications for vegetation modelling. *Ecological Modelling*, 197, 394–404.

contain some level of error (Van Niel et al. 2004). For example, the DEM can be used to estimate the slope, aspect, hydrology and radiation for the study area.

Elevation is often used as an indirect predictor of temperature (Lookingbill and Urban 2003), or is used in techniques (e.g. BIOCLIM) that interpolate data from weather stations (Hughes et al. 1996; Lindenmayer et al. 1999, 2000; Dymond and Johnson 2002). These methods are prone to error when predicting local temperature variations because there is often a lack of weather stations on which to base the interpolation, they ignore the influence of the local topography (Guisan and Zimmerman 2000), and they fail to account for effects such as cold air drainage and evaporative cooling (Lookingbill and Urban 2003).

Mountainous terrain can be especially difficult to model because the high spatial variability of environmental factors leads to a complex mosaic of vegetation (Hörsch 2003). In Australia, there can be ten species of eucalypt (Myrtaceae: *Eucalyptus* spp.) in a small area (Florence 2004). Whilst it is accepted that there are changes in dominant canopy species associated with slope and aspect, the exact relationship with direct predictors is uncertain (Bell and Williams 1997). Environmental factors such as climate, phosphorus (Beadle 1954, 1966), fire (Florence 2004), and moisture (Wardell-Johnson et al. 1997) have been associated with the distribution of Australian vegetation, but no vegetation models have yet been able to satisfactorily explain the distribution of eucalypts (Austin et al. 1997), possibly because some scientists assume that the same environmental factors are limiting all the species (Arundel 2005). In addition, most studies have ignored the interaction between environmental factors at a location and those of neighbouring areas.

The complexity of Australian vegetation is well illustrated by the Illawarra Escarpment, approximately 80 km south of Sydney, Australia (Figure 2.1). There is a

complex mosaic of eucalypt forests, woodlands, and rainforests on the Woronora Plateau and the slopes of the escarpment. The city of Wollongong lies on the coastal plain and foothills to the south and east, but there are also some remnants of native vegetation. The climate of the Illawarra region is humid and mild, with average daily minimum temperatures of 9 to 18 °C and maximum temperatures of 17 to 26 °C throughout the year (Fuller 1995). Annual rainfall ranges from 1000–1200 mm on the coastal plain to 1500–1600 mm on the escarpment, with slightly more rain falling in February-May than in August-November (Fuller 1995).

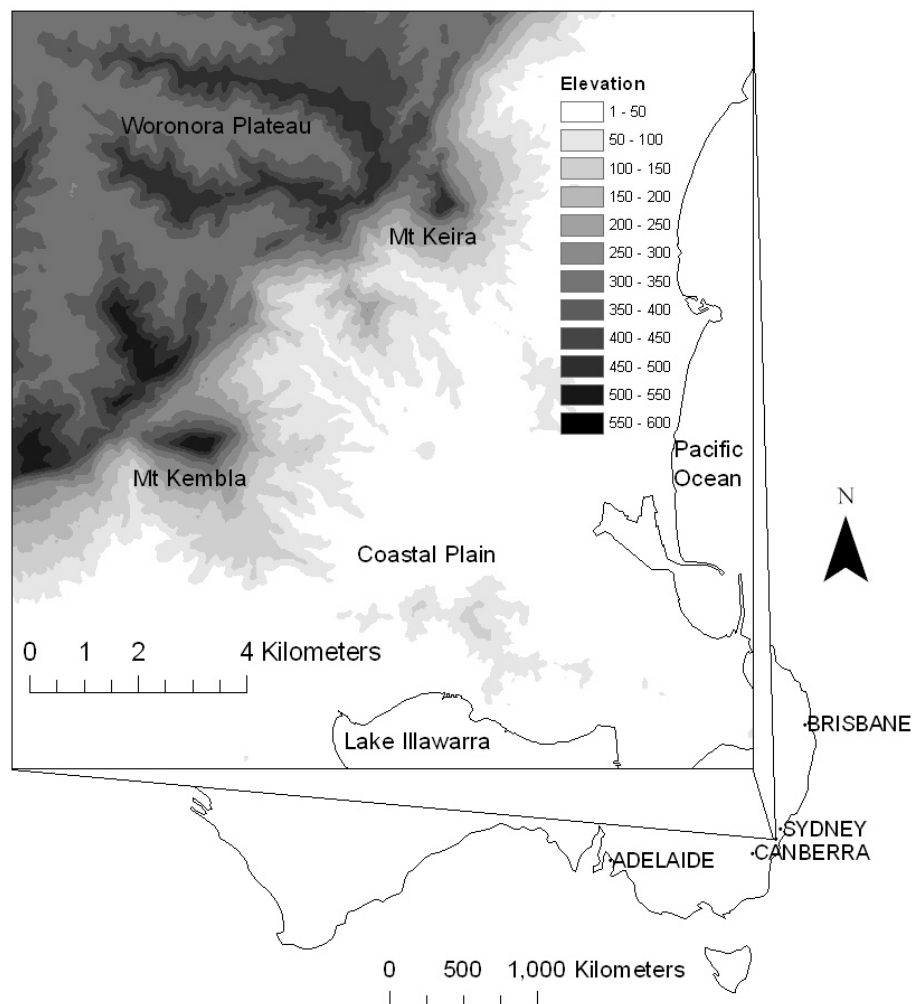


Figure 2.1: The topography of the Illawarra Escarpment in the vicinity of Wollongong, Australia. The inset is a Digital Elevation Model showing the rising elevation from the coastal plain to the Woronora Plateau, with Mt Keira and Mt Kembla protruding eastward.

Preliminary modelling of the study area using Generalised Additive Models (GAMs; Hastie and Tibshirani 1990) confirmed that it was difficult to quantitatively explain the current distribution of vegetation using the available predictors. It has been suggested that obtaining better quality predictor variables would be a good first step to address this problem (Guisan and Zimmermann 2000; Austin 2002). This could include replacing the elevation predictor with more accurate and/or direct maps of average maximum and minimum temperatures.

The aims of this study were to develop more accurate maps of average summer maximum and minimum temperatures, and to quantify the improvement in vegetation modelling performance when these were used instead of elevation. Whilst estimates of temperature have been developed for other study areas, this study aimed to determine whether the estimates could be improved by considering the interaction between a location and its surrounding environment.

It was hypothesised that wind and air movements would average out large differences in radiation and elevation over small distances and cause temperatures to be more strongly correlated with the average elevation and radiation in the surrounding area than they are with the actual elevation and radiation where the temperature was recorded. This was tested by comparing the linear regression of temperature against the local average radiation and elevation with the linear regression of temperature against the actual elevation and radiation where temperature sensors were located. A stronger relationship would indicate that the local averages were better correlated with temperature, and would therefore be more appropriate for any temperature prediction method including linear regression and elevation sensitive interpolation methods such as ANUSPLIN and GIDS (Price et al. 2000).

Chuanyan et al. (2005) have also suggested that elevation is an unsatisfactory predictor for capturing the environmental niche of vegetation, and recently compared a number of other temperature estimation techniques. It is hoped that using the locally averaged elevation and radiation will further increase the accuracy of any of these methods, and lead to more ecologically realistic vegetation models.

2.2 Materials and methods

2.2.1 Environmental predictor variables

Elevation data for the Illawarra Escarpment was available in the form of a digital elevation model (DEM) with a 10 m cell size. Whilst it is unknown how the DEM was created, it appears to have been derived from the contours of a topographic map and contains some noticeable imperfections. Airborne Laser Scanning (ALS) data that is available for a subset of the study area (courtesy of AAMHatch Pty Ltd) suggest that the errors in elevation are generally in the order of 5–10 m for most of the study area, but may be up to 30 m near the steep cliffs around Mt Keira (See Section 3.6.1 for a more comprehensive assessment of data accuracy).

Three maps were also obtained. A map of vegetation communities was available courtesy of the Department of Environment and Conservation (NPWS 2002). Spatial errors for community boundaries are within 24 m for 93% of the map area, but may be up to 70 m in some areas on the escarpment. Communities are described in terms of species composition, and the canopy cover of each structural layer is estimated. Cultural data (roads, walking trails, powerlines, and gas pipelines) was provided by the Department of Infrastructure, Planning and Natural Resources (DIPNR). A comparison with high-resolution aerial photos (courtesy of AAMHatch Pty Ltd) suggests that spatial errors for the cultural data may be up to 50 m in the vicinity of the escarpment. A

geology map was available for part of the study area courtesy of Phil Flentje at the University of Wollongong. This was in the form of categorical data, with one value for each of the 20 geological units. An extra categorical value, 'unknown', was added for those areas outside the geological map. The geology map was used as a surrogate for soil properties (such as phosphorus) that are known to influence the distribution of vegetation.

Streamlines were calculated from the DEM where the flow accumulation (as determined using ESRI ArcMap hydrology functionality) was greater than 500 cells. The distance to the streams was calculated using Euclidean distance, with values less than one being rounded up so that the log transformation would produce values greater than zero and the output would be more sensitive to areas that are near streams. Lookingbill and Urban (2003) used a similar log transformation of distance to streams in their estimations of temperature.

The distance to disturbance was estimated by calculating the minimum distance to either the lines in the cultural data or the 'Cleared' polygons in the vegetation data. Values less than one were rounded up and the data was also log transformed.

Exposure to winds was estimated by calculating the angle to the horizon for each azimuth that is a multiple of 15. This was done in ArcGIS using an AML script to calculate the shadow using 'hillshade' at altitudes of 0.125, 0.25, 0.5, 0.75, 1, 1.25, 1.5, 2, 2.5, 3, 3.5, 4, 5, 6, 7, 8, 9, 11, 13, 15, 18, 21, 25, 29, 34, 39, 45, 51, 58, 65, 73, and 81 degrees. The cells in the resulting raster grid contained the minimum angle that resulted in no shadow. This grid was incremented by one and log transformed so that the output was more sensitive to changes near low altitudes.

Exposure to warm and dry westerly to northwesterly winds was estimated by averaging the log-transformed angles for azimuths of 255, 270, 285, 300, 315, and 330

degrees. Exposure to cold, moist southerly winds was estimated by averaging the log-transformed angles for azimuths of 150, 165, 180, 195, and 210 degrees. Exposure to moist northeasterly winds was estimated by averaging the log-transformed angles for azimuths of 30, 45, and 60 degrees. These azimuths were chosen based on preliminary modelling and existing literature on the area. Dry westerly winds are dominant in winter and moist south and northeasterly winds are common in spring and summer (Erskine 1984; Bywater 1985; Mills 1986; Fuller 1995). Northwesterly winds are not as common, but are dry and warm in summer and can have a desiccating influence on the local rainforests (Fuller 1995).

Averages were employed to the wind directions because it had the effect of allowing wind to ‘bend’ around mountains, thus avoiding the long wind-shadows that stretch across the entire coastal plain when considering only one direction. Kramer et al. (2001) used the EXPOS model for a similar effect, but their model also allows wind to bend over the top of mountains. In any event, these are still approximations for exposure to wind, as wind is also influenced by valleys, mountaintops, and elevation (Raupach and Finnigan 1997; Finardi et al. 1998; Uchida and Ohya 1999; Ruel et al. 2001).

Incoming solar radiation was calculated using the DEM and the Solar Analyst (USDA Forest Service) extension for ESRI ArcView. The total direct radiation was calculated for January 18th 2005, and is referred to in this chapter as simply ‘radiation’. The 18th January was selected because it was near the middle of the observation period.

2.2.2 Predicting maximum and minimum temperatures

Geographic Information System (GIS) data was used to stratify the study area according to elevation, radiation, and distance to streams, as these factors have been identified as influencing maximum and/or minimum temperature (Moore et al. 1993; Lookingbill and Urban 2003). Forty locations for temperature loggers were selected based on the

stratification results to overcome a number of problems with random sampling. These problems include the clustering of high elevation and low radiation sites which would cause random sites to be so close that they may be spatially auto-correlated, access restrictions to privately owned lands, and other access problems caused by the steep topography and dense vegetation that could not be identified until the sites were visited.

Whilst it is recognised that non-random sampling can lead to bias, attempts were made to minimise this risk by ensuring the full range of each predictor was sampled, and by ensuring the environmental predictors were poorly correlated for the sampled locations. Radiation was poorly correlated with both elevation ($r^2 = 0.009$) and distance to streams ($r^2 = 0.015$) due to the stratification, whilst elevation was moderately correlated with distance to streams ($r^2 = 0.306$) because there were fewer streams near the drainage divides at high altitudes.

Temperatures were recorded using DS1921G iButton temperature loggers (Dallas Semiconductor/MAXIM). Recordings were made every 30 minutes from 29th November 2004 to 9th January 2005, and from 15th January 2005 to 25th February 2005. Sensors were placed on the surface of the ground with as much shelter from direct radiation as possible given the vegetation at each location. Each sensor was pinned to the ground inside a small, coarse meshed bag, however three sensors moved by 1–2 m during the study period due to disturbance from falling trees, erosion, and possibly lyrebirds (See Section 7.4 for a comprehensive discussion on sensor placement).

In previous studies, temperature sensors have been placed at a variety of heights including 10 cm and 5 cm below the surface, and 15 cm, 30 cm, 1.3 m and 2 m above the surface (Lookingbill and Urban 2003; Lemenih et al. 2004; Porté et al. 2004; Ritter et al. 2005). Some have used radiation screens to avoid direct radiation (Ritter et al. 2005), whilst others used the shade of the trees (Lookingbill and Urban 2003). It is not

evident which height provides the most useful predictor for the distribution of vegetation, but it has been shown that both soil and air temperatures influence the growth rate of eucalypts (Bell and Williams 1997). It is also not clear how well the surface temperature correlates with either the subsurface soil temperatures or canopy air temperatures, but it has been suggested that surface temperatures have the maximum diurnal variation and may be 5–10 °C different from the air temperature at 1.5 m – where meteorological measurements are made (Campbell and Norman 1998). Surface temperatures may be more spatially variable, because they are less subjected to the winds and advection that can mix air (Porté et al. 2004), and are obviously more exposed to solar radiation than subsurface measurements.

When the sensors were reprogrammed in mid January 2005, the percentage canopy cover of each site was visually estimated to the nearest 10% and recorded. The full range of canopy covers were observed (0-100%), and canopy cover was poorly correlated with elevation ($r^2 = 0.065$), radiation ($r^2 = 0.117$) and distance to streams ($r^2 = 0.072$). Therefore, canopy cover was considered for inclusion in models for predicting temperature. It should, however, be noted that the visual assessment of canopy cover is prone to error. The relative importance of different canopy and sub-canopy layers is not obvious, and it is unknown how much the canopy cover varies temporally, or whether measurements should be biased towards the path of the sun.

One site had to be discarded because the data on the temperature logger was lost. For each of the remaining 39 sites the daily maximum and minimum temperatures were recorded, and then averaged to determine the mean maximum and minimum temperatures for each of the 39 sites. Linear regression was used to determine how well elevation explained the average maximum and minimum temperatures, as done by Lookingbill and Urban (2003). The results of the regression were compared with the

regression using elevation in combination with the other predictors (radiation, log distance to streams, and percentage canopy cover).

Partial response graphs and residuals were examined to ensure that the regression was, as expected, linear, and that the residuals appeared to be normally distributed. Linear relationships have already been established (Lookingbill and Urban 2003).

In order to establish whether or not the relationship between elevation and temperature varies during the course of the day, the average temperature for each 30-minute period was calculated for each site. Regression was used to calculate the relationship for each 30-minute interval, and the correlation coefficients recorded. The regression was conducted using elevation alone, and elevation in combination with radiation and canopy cover. The lapse rate was estimated from the coefficient of the elevation parameter in the regression.

2.2.3 Using a low pass filter to improve estimates

It was hypothesised that wind and air movements would cause the maximum temperature at a given site to be more strongly correlated with the average elevation and radiation over the surrounding area than with the actual elevation and radiation where the temperature was recorded. For example, a site surrounded by areas of consistently high radiation would be warmer than a site surrounded by a mosaic of low and high radiation.

In order to test this hypothesis, the radiation and elevation predictors were transformed using low pass filters. This process averaged the values of each predictor over a circular region around each pixel in the predictor map, and calculated the moving average of elevation and radiation. The low pass filters were performed using the neighbourhood functionality of ESRI ArcMap, with radii of 100 m, 200 m, 500 m, 750 m, 1000 m, 1250 m, and 1500 m.

For each radius, linear regression was used to examine the effect of the low pass filter on the correlation coefficient between maximum temperature and elevation and/or radiation (in comparison to the regression with the unfiltered predictors). The radius that maximised the r^2 of the regression was used to estimate the optimal radius for the low pass filter. The low pass filter was also used to examine the effect on the correlation coefficient of the regression for the average temperature for each 30-minute period during the day.

As the minimum radius (100 m) is ten times the cell size of the DEM (10 m), the resolution of the DEM is not expected to have a significant influence on the optimal radius. As averaging will cancel out random errors, it is also expected that DEM accuracy will become less significant once the low pass filter is used. Problems may be encountered in future if the cell size approaches the radius of the filter, but there is no reason to believe that the optimal radius in terms of distance would change, even though the radius would obviously be less in terms of the number of cells.

2.2.4 Vegetation modelling

A dataset of random points (defined by an easting and northing in the study area) was created and the vegetation community was determined for each point from the vegetation map (NPWS 2002). Rare and non-vegetated communities were discarded, which left 23 different communities. There were also a small number of points (<1%) that had to be discarded because there were spatial or other inconsistencies between the different data layers. For each remaining point, the environmental predictors were extracted from the appropriate themes in ArcMap, and the data set was randomly split into a training data set of 4995 points and a validation data set of 2306 points.

A GAM was produced in SPlus (Insightful Corp.) for each of the 23 communities using the training dataset and the predictors (see Section 2.2.1 for more details) of

elevation, geology, log distance to streams, log distance to disturbance, and exposure to the three wind directions (EGWD models). Each of the models was then applied to the validation data set and each point classified into one of the 23 communities according to the model that produced the highest predicted probability of occurrence.

The GAMs were then repeated using the predicted average summer maximum and minimum temperatures instead of elevation. In the first instance, the maximum was predicted using the filtered elevation and radiation (ER model), and in the second instance the maximum was predicted using the filtered elevation and radiation and the canopy cover (CER model). It was necessary to estimate the canopy cover from the vegetation map, with each community assumed to have a constant summer canopy cover of between 30% and 90%, estimated according to the community descriptions by the NPWS (2002). This is in contrast with the visual estimates of canopy cover that were used to derive the formula for maximum temperature. In future, the canopy cover could be estimated more accurately using remote sensing (Wang et al. 2003).

No attempts were made to trim insignificant predictors from any of the GAMs, or to vary the degrees of freedom for each predictor. This ensures that the comparison between models is only comparing the effect of the maximum temperature predictor, but runs a risk of over-fitting. It has also been suggested that excessive absences past the recorded distribution of a species need to be culled (Austin and Meyers 1996; Leathwick et al. 1996). This was not done because the output was the “most probable entity” rather than the “probability of occurrence” (Guisan and Zimmermann 2000). Whilst it is recognised that a GAM can predict a non-zero probability of occurrence outside the observed range, this will not result in it being the most probable entity as long as another community is more likely to occur in that location.

2.3 Results

2.3.1 Predicting minimum temperature

Elevation was highly significant for predicting the average summer minimum temperatures ($r^2 = 0.763$, $t = -10.92$, d.f. = 37, $P < 0.001$). Distance to streams was also significant ($r^2 = 0.196$, $t = -3.000$, d.f. = 37, $P < 0.01$), but this must be treated with caution due to the moderate correlation between elevation and distance to streams. This is emphasised by the fact that when distance to streams and elevation were both used to model the average minimum temperature, the distance to streams was no longer significant ($r^2 = 0.766$, $t_{\text{elev}} = -9.355$, d.f. = 36, $P_{\text{elev}} < 0.001$, $t_{\text{stream}} = 0.603$, d.f. = 36, $P_{\text{stream}} > 0.05$), and there was negligible improvement in correlation from the regression with elevation alone. The average minimum temperature was not significantly influenced by either radiation ($r^2 = 0.006$, $t = -0.461$, d.f. = 37, $P > 0.05$) or canopy cover ($r^2 = 0.015$, $t = 0.754$, d.f. = 37, $P > 0.05$), nor were they significant when combined with elevation and/or distance to streams.

Based on these results, the minimum temperature was predicted based on elevation, with the equation:

$$T_{\min} = 17.3 - 0.0052 * \text{Elevation}$$

Where T_{\min} is the predicted average summer minimum temperature ($^{\circ}\text{C}$) at each location, and the Elevation (m) is taken from the DEM. The graph of the predicted average minimum temperature against the recorded average minimum temperature is shown in Figure 2.2a.

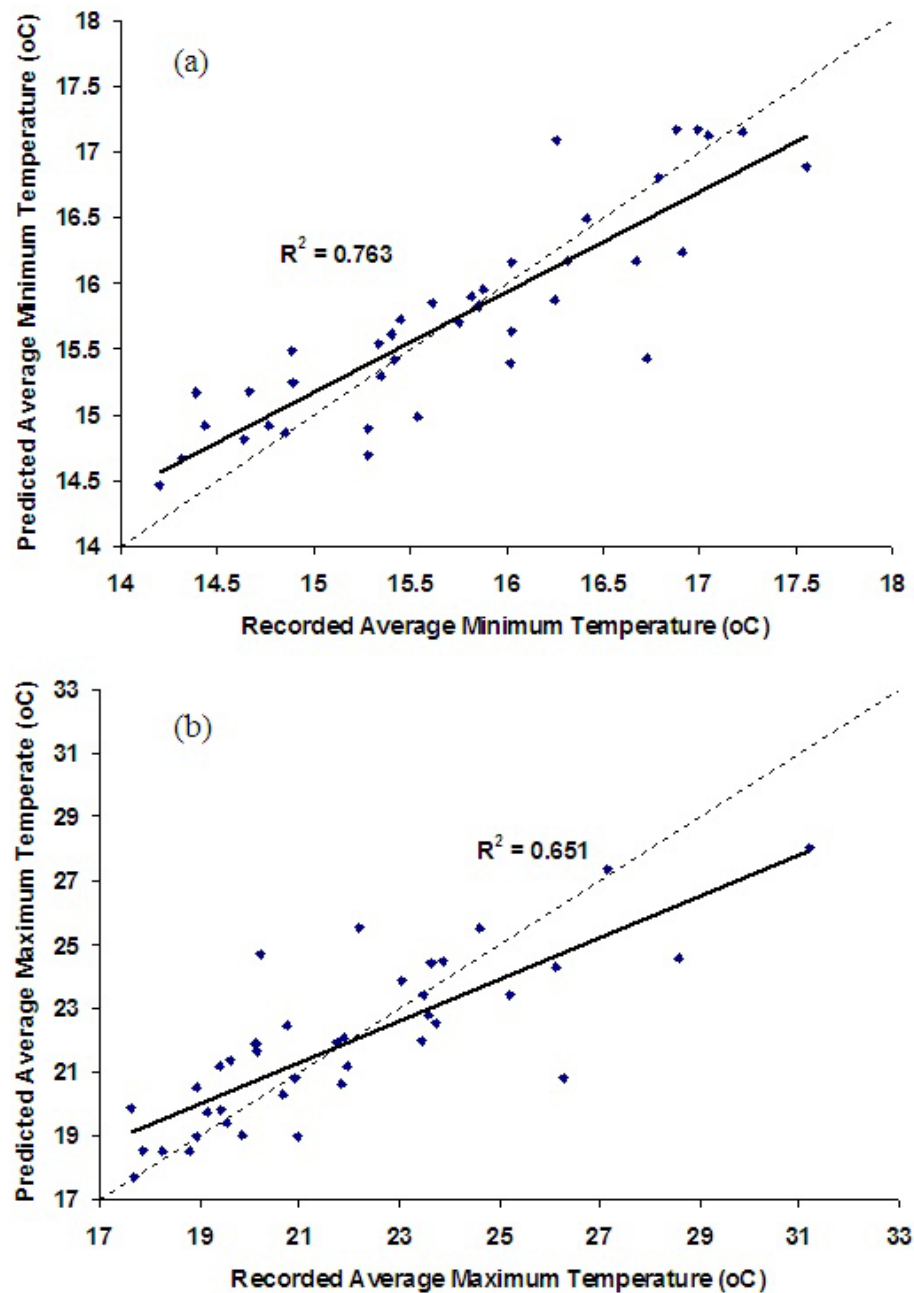


Figure 2.2: The relationship between the predicted average summer minimum (a) and maximum (b) temperatures and the corresponding minimum and maximum temperatures that were recorded by 39 temperature sensors placed at ground level on the Illawarra Escarpment. The predicted average minimum temperature is based on the regression of the actual recorded minimum temperatures against the elevation of each site, whilst the predicted average maximum temperature is based on the regression of the actual recorded maximum temperature against the elevation and canopy cover of each site.

2.3.2 Predicting maximum temperature

Linear regression with each predictor individually showed that both canopy cover ($r^2 = 0.303$, $t = -4.014$, d.f. = 37, $P < 0.001$) and elevation ($r^2 = 0.185$, $t = -2.896$, d.f. = 37, $P < 0.01$) were significantly correlated with maximum temperature, but radiation ($r^2 = 0.055$, $t = 1.473$, d.f. = 37, $P > 0.05$) and distance to streams ($r^2 = 0.005$, $t = -0.433$, d.f. = 37, $P > 0.05$) were not. When both elevation and canopy cover were used in the regression the correlation improved substantially ($r^2 = 0.651$, $t_{\text{canopy}} = -6.934$, d.f. = 36, $P_{\text{canopy}} < 0.001$, $t_{\text{elev}} = -5.988$, d.f. = 36, $P_{\text{elev}} < 0.001$). When all the parameters were included in a multiple regression the r^2 improved to 0.680, but elevation and canopy cover were the only predictors that were significant. The equation for predicting the average maximum temperature using the canopy cover and elevation was:

$$T_{\text{max}} = 28.9 - 6.6 * \text{Canopy} - 0.0127 * \text{Elevation}$$

Where T_{max} is the predicted average maximum temperature ($^{\circ}\text{C}$), Canopy is the visually estimated canopy cover as a ratio between 0 and 1, and Elevation (m) is taken from the DEM. The graph of the predicted average summer maximum temperature against the recorded average summer maximum temperature is shown in Figure 2.2b.

2.3.3 The effects of low pass filters

When a low pass filter was used to average the radiation over various radii, the correlation with the average maximum temperature increased substantially, reaching a maximum value at a radius of 1000 m ($r^2 = 0.199$, $t = 3.029$, d.f. = 37, $P < 0.01$, Figure

2.3a). It can also be seen that the relationship became more significant, and transformed radiation from an insignificant parameter ($P > 0.05$) into a significant one ($P < 0.01$).

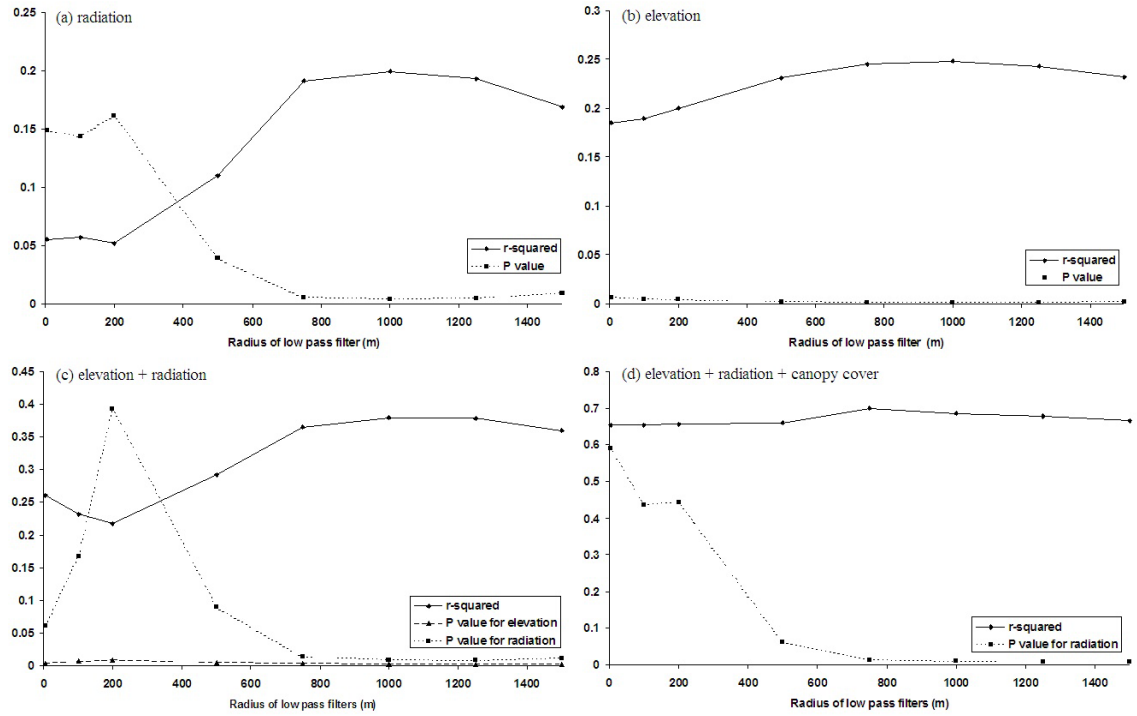


Figure 2.3: The r^2 and P values that resulted from the regression of the average summer maximum temperature against various predictors for 39 temperature sensors placed at ground level on the Illawarra Escarpment. Predictors were the low pass filtered radiation (a), the low pass filtered elevation (b), the low pass filtered radiation and the low pass filtered elevation (c), and the canopy cover, low pass filtered radiation and low pass filtered elevation (d).

There was also an improvement in the significance of the relationship between maximum temperature and elevation when using a low pass filter, with the best result also at a radius of 1000 m ($r^2 = 0.248$, $t = -3.493$, d.f. = 37, $P < 0.01$, Figure 2.3b). Elevation was significant at every radius ($P < 0.01$), but there were slight improvements in the correlation coefficient and significance.

The correlation coefficient was also improved when average summer maximum temperature was regressed against both filtered elevation and filtered radiation (Figure 2.3c). In this case, there was a slight degradation in correlation with radii of 100 m to 200 m, but the best results were once again with a radius of 1000 m ($r^2 = 0.379$, $t_{\text{elev}} = -3.236$, d.f. = 36, $P_{\text{elev}} < 0.01$, $t_{\text{rad}} = 2.760$, d.f. = 36, $P_{\text{rad}} < 0.01$). This represented a substantial improvement from when the unfiltered elevation was used alone (r^2 from 0.185 to 0.379). Elevation was significant for each radii ($P < 0.01$), but radiation was only significant for radii between 750 m and 1500 m ($P < 0.05$).

When canopy cover was included as a predictor in the linear regression, along with the low pass filtered elevation and radiation, the best correlation was at a radius of 750 m ($r^2 = 0.699$, $t_{\text{canopy}} = -6.233$, d.f. = 35, $P_{\text{canopy}} < 0.001$, $t_{\text{elev}} = -5.517$, d.f. = 35, $P_{\text{elev}} < 0.001$, $t_{\text{rad}} = 2.617$, d.f. = 35, $P_{\text{rad}} < 0.05$, Figure 2.3d). Both elevation and canopy cover were highly significant at every radii ($P < 0.001$), but radiation only became significant with radii greater than 500 m ($P < 0.05$). There was a marginal improvement in correlation between the low pass filtered result and the unfiltered result (r^2 from 0.654 to 0.699).

2.3.4 Intra-day trends

Regression of the average temperature for each 30-minute period against elevation emphasised the poor relationship between elevation and daytime temperatures. Not only is elevation an inadequate predictor of the average maximum temperature ($r^2 = 0.185$), but it is also a poor predictor of the average temperature for every time interval from 10:00am to 4:30pm inclusive ($r^2 < 0.4$). In contrast, the estimate using the low pass filtered elevation and radiation maintained a reasonable correlation throughout the day ($r^2 > 0.43$). The result using the filtered predictors was higher than the result based on

unfiltered elevation for every time interval from 8:00am to 9:00pm inclusive (Figure 2.4a).

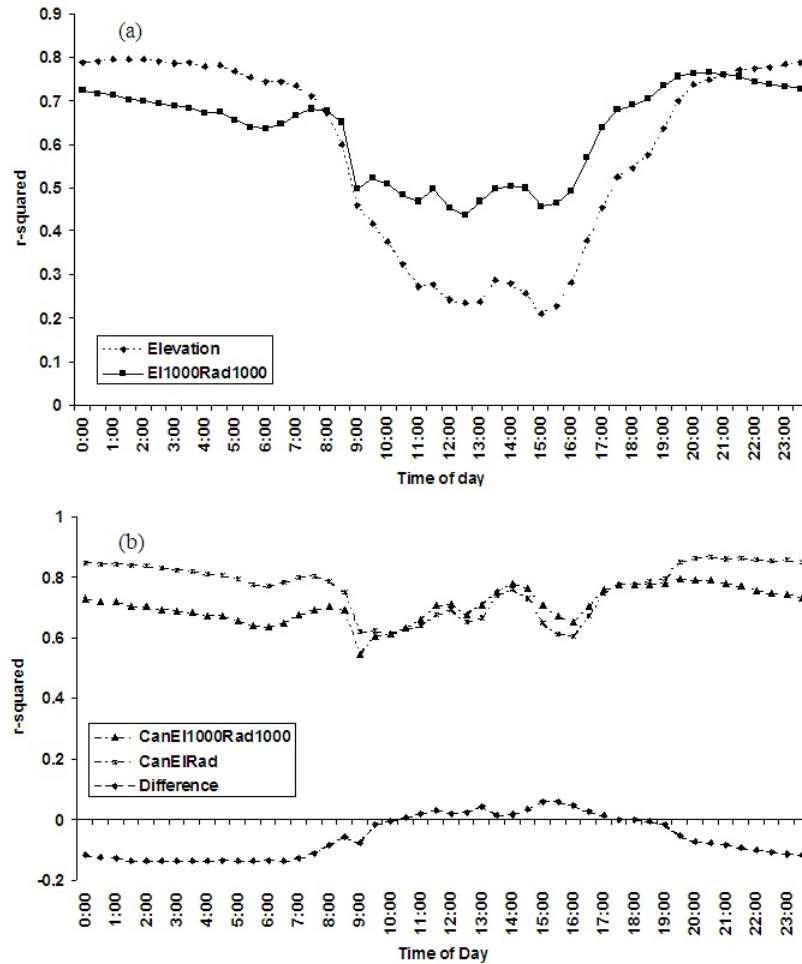


Figure 2.4: A comparison between the r^2 values for the different regressions predicting the average temperature for each 30-minute period throughout the day (based on 39 temperature sensors at ground level). Elevation refers to the regression using the unfiltered elevation (m) taken directly from the DEM. EI1000Rad1000 refers to the regression using the elevation (m) and radiation (W/m^2) as averaged over a radius of 1000 m. CanEI1000Rad refers to the regression using the canopy cover, unfiltered elevation (m) and unfiltered radiation (W/m^2). CanEI1000Rad1000 refers to the regression using canopy cover, the low pass filtered elevation (m) and the low pass filtered radiation (W/m^2) as averaged over a radius of 1000 m. Difference refers to the difference between the r^2 values of the two regressions involving canopy cover, with positive values indicating that the filtered result is higher.

When canopy cover, radiation and elevation are included in a multiple linear regression, then once again the low pass filtered results for day time average temperatures (10:30am to 5:00pm inclusive) are an improvement over the unfiltered results, but the night-time temperatures are substantially better using the unfiltered elevation (Figure 2.4b).

The intra-day results emphasise that whilst the low pass filtered radiation and elevation improve daytime temperature predictions, they are not as good at predicting the nighttime temperatures. This is consistent with using unfiltered elevation to predict the minimum temperature, but using the low pass filtered elevation and radiation to predict the maximum temperature.

2.3.5 Vegetation modelling

The overall accuracy of the GAM model using elevation, geology, distance to streams, distance to disturbance, and exposure to the 3 wind directions (EGWD model) was quite poor at 46.4%. When the average summer maximum temperatures were predicted using the low pass filtered elevation and radiation (ER model), the overall accuracy of the GAM model improved to 49.0%. The formula used was:

$$T_{\max} = -32.7 + 0.01327 * \text{Rad1000} - 0.0115 * \text{Elev1000}$$

Where T_{\max} is the predicted average summer maximum temperature ($^{\circ}\text{C}$), Rad1000 is the radiation (W/m^2) averaged over a 1000 m radius, and Elev1000 is the elevation (m) averaged over a 1000 m radius.

Alternatively, when the maximum temperature was predicted using the canopy cover and the low pass filtered radiation and elevation (CER model), the overall accuracy improved substantially to 61.8%. The formula used was:

$$T_{\max} = -13.3 + 0.00955 * \text{Rad1000} - 0.0137 * \text{Elev1000} - 5.3 * \text{Canopy}$$

Where T_{\max} is the predicted maximum temperature ($^{\circ}\text{C}$), Rad1000 is the radiation (W/m^2) averaged over a 1000 m radius, Elev1000 is the elevation (m) averaged over a 1000 m radius, and Canopy is the canopy cover as a fraction between 0 and 1.

Figure 2.5 illustrates the estimated average summer maximum temperature based on the low pass filtered elevation and radiation. The distributional patterns of maximum temperatures are shown to be vastly different to the distributional patterns of the average summer minimum temperature that are based on elevation alone. These differences are maintained when the average summer maximum temperatures are predicted based on canopy cover, and the low pass filtered elevation and radiation.

2.4 Discussion

2.4.1 Temperature prediction

The average summertime maximum temperature could not be accurately predicted using elevation alone ($r^2 = 0.185$), but a much better estimate could be made using the percentage canopy cover, the low pass filtered elevation and the low pass filtered radiation ($r^2 = 0.685$). The estimates based on the low pass filtered predictors also outperformed elevation for predicting the average temperature for each 30-minute period from 8:00am to 9:00pm inclusive. In all cases the optimal radius for the low pass

filter was between 750 m and 1000 m. Elevation was a better predictor for the average summer minimum and overnight temperatures.

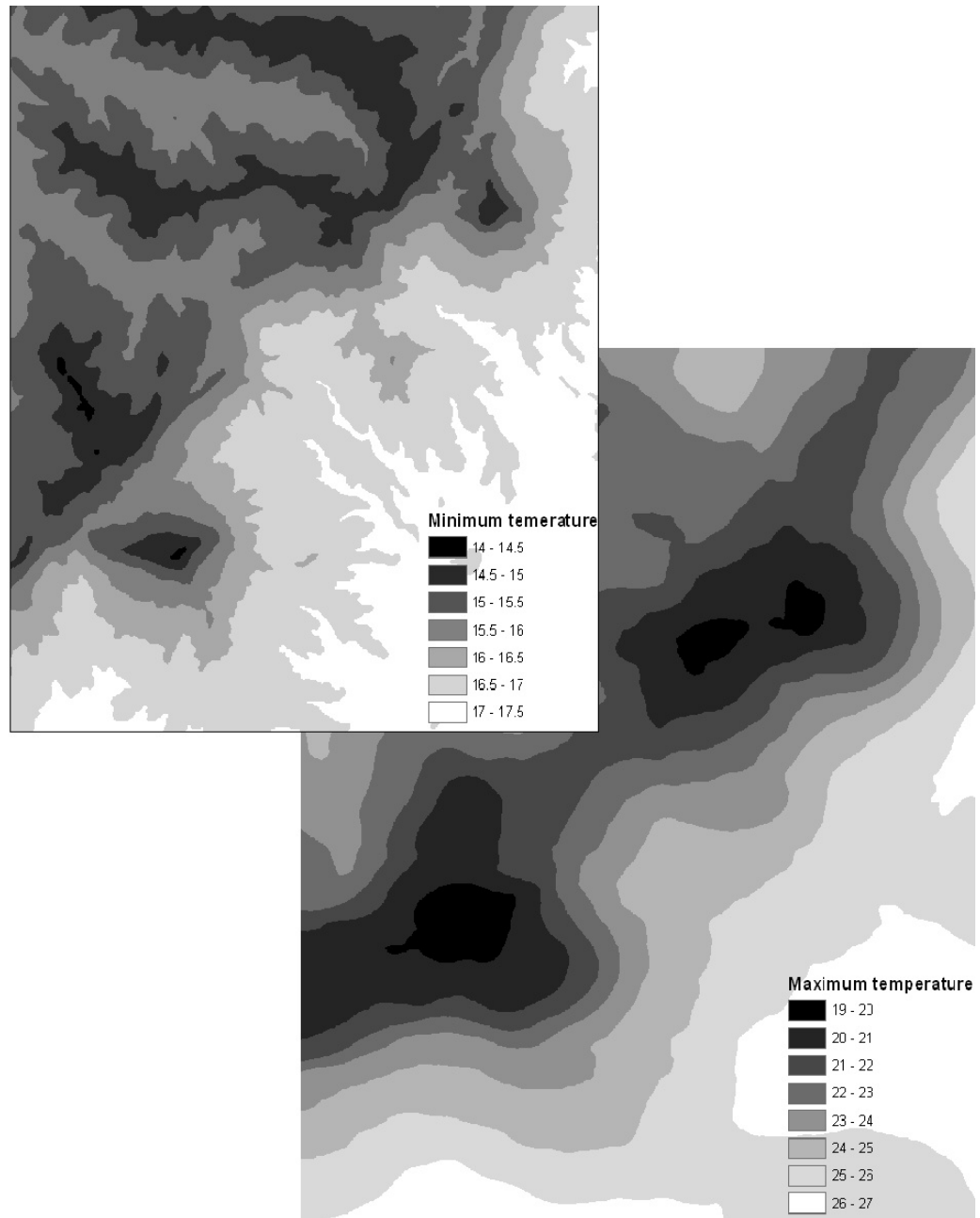


Figure 2.5: A comparison between the predicted average summer minimum temperature based on elevation (top left) and the predicted average summer maximum temperature based on the low pass filtered elevation and low pass filtered radiation (bottom right). Whilst the minimum temperature south of Mt Kembla is similar to the minimum temperature north of Mt Keira (15.5 °C – 16.5 °C), the maximum temperature appears to be 1 °C – 2 °C cooler (20 °C – 21 °C versus 21 °C – 23 °C).

Using a low pass filter is a crude method to consider the interaction between a site and its neighbours, but appears to be effective for this study area. This is possibly because it reflects the movement of hot and cold air to and from surrounding locations, and allows locations where there is consistently higher radiation to be hotter than those locations where there is a mosaic of high and low radiation. It remains to be tested whether this local phenomenon can be replicated at other sites, and how the relationship varies according to latitude and relief. In areas with more constant canopy cover and radiation it would be expected that elevation would become more dominant.

Likewise, as the diurnal variation in temperature decreases at locations deeper into the soil or higher off the surface (Campbell and Norman 1998), it is possible that the influence of radiation and/or canopy cover may be reduced when the temperature sensors are placed at different locations, or if radiation screens are used. Under these circumstances, it is also possible that the effect of the surrounding environment is reduced, and hence the effect of low pass filtering may be diminished or absent.

It is possible that the filtering method could be improved by weighting the elevation/radiation in the surrounding locations according to distance from the centre (similar to Price et al. (2000) and Ferrier et al. (2002)), predominate wind directions, or the shape of the topography. It also needs to be confirmed that the low pass filtered elevation and radiation would also improve the results of other temperature estimation techniques such as ANUSPLIN and GIDS (Price et al. 2002). At the continental scale, where the pixel size is roughly the size of the filter used in this study (e.g. 1000 m in Price et al. 2000), a low pass filter may not have any effect. At the regional scale where the pixel size is 100 m or less (e.g. Ferrier et al. 2002), filtering the elevation used by ESOCIM may slightly improve the accuracy of the model, but this also depends on

whether the effect is valid for non-surface temperatures, and whether it is valid for other study areas.

The pixel size is obviously important for studies in mountainous areas, because the elevation may vary by hundreds of meters within one pixel when the pixel size is 100–1000 m. This could translate to a within pixel temperature variations of 2–3 °C. It is also worth noting that in the Illawarra region, there are many vegetation patches that occupy small areas that could not be captured using a pixel size of 100–1000 m, and so a landscape scale model is necessary to capture to fine scale changes in temperature and vegetation. It is at the landscape scale that low pass filtering probably has the greatest potential to improve temperature estimates and modelling results.

Previous studies have shown that the lapse rate of temperature is in the order of 6 °C/1000 m, with daily variations from 3.8 °C/1000 m near the minimum temperature to 7.0 °C/1000 m near the maximum (Lookingbill and Urban 2003). In this study, the lapse rate varied from a minimum of 4.9 °C/1000 m at 6am to a maximum of 8.4 °C/1000 m during the day. A lapse rate of 5.2 °C was determined for the minimum temperature based on elevation alone, and values of approximately 9–13 °C/1000 m were used in the various formulas for maximum temperature, depending on the parameters that were included in the regression. Therefore, the results of this study appear to be consistent with previous research.

The relationship between canopy cover and temperature is also consistent with previous studies, with the effect ranging from 2 °C to 10 °C depending on the location of the sensor and the range of canopy cover examined (Porté et al. 2004; Lemenih et al. 2004; Ritter et al. 2005). In this study, a 50% difference in canopy cover accounted for a 2.6 °C to 3.3 °C difference in maximum temperature, depending on the other parameters included in the regression.

The effect of radiation is more difficult to compare with previous studies because it was not significant as a predictor unless the low pass filter was used. Lookingbill and Urban (2003) found that radiation slightly improved the estimation of the maximum temperature for their mountainous area (r^2 from 0.41 to 0.48), but the effects may vary according to latitude and the time of year.

It is possible that the predictions of maximum temperature could be improved further. Qualitatively, moist sites appeared to be cooler than dry sites. This is consistent with a study by Bywater (1985), which suggested that diurnal variations are lower during rainy periods, and the study by Ritter et al. (2005), which considered soil moisture as one factor leading to lower temperatures.

Note that temperatures were only recorded for the summer period from December 2004 until February 2005. These recordings may not be representative of the long-term average for this time of year, and may differ from the temperatures recorded at any other height. Therefore, all temperatures discussed in this chapter should be treated as relative temperatures rather than absolute temperatures. In addition, it is unknown whether this time of year and these sensor locations have the most predictive power for modelling species distribution. It is possible that another season or sensor height may be more biologically important.

2.4.2 Vegetation modelling

When the vegetation on the Illawarra Escarpment (Figure 2.1) is modelled, using a GAM for each of the 23 communities, and the validation data set classified according to which community has the highest probability of occurrence, the results are quite poor. Using the predictors of elevation, radiation, distance to streams, distance to disturbance, and exposure to 3 wind directions the overall accuracy is only 46.4%. If the elevation predictor is replaced with the predicted average summer minimum temperature (using

elevation) and the predicted average summer maximum temperature (using canopy cover, and the low pass filtered elevation and radiation at a radius of 1000 m) the overall accuracy improves substantially to 61.8%.

Caution must be used, however, since this improvement may be artificially high because the vegetation map was used to estimate canopy cover and therefore predict maximum temperature. This may have introduced feedback into the system when the maximum temperature was subsequently used to predict the vegetation community.

There are still improvements in the overall GAM accuracy from 46.4% to 49% when canopy cover is not used, and the maximum temperature is predicted using the low pass filtered elevation and radiation alone. Therefore, the better estimates of maximum temperature improve vegetation models regardless of whether or not canopy cover is used. Whilst the improvement without canopy cover is not large, this would be expected given that the estimate of maximum temperature is much poorer than when canopy cover is included ($r^2 = 0.379$ versus $r^2 = 0.685$).

The vegetation modelling results will also have been affected by the accuracy of the vegetation maps that were used. Subsequent surveying has highlighted some limitations of these maps, and the vegetation models in subsequent chapters of this thesis give a more reliable assessment of performance. However, even survey data has inaccuracies, and vegetation models are inherently trying to explain inaccurate distributional data. The training and test samples that were generated in this chapter were very large given the small size of the study area, and this potentially introduced spatial autocorrelation. Under these circumstances, the training and test datasets are not truly independent, and the performance may be overstated. Once again, the models in subsequent chapters are not affected by this issue, and are a more reliable indication of model performance.

In this study, the standard deviation of the regression residuals decreased from 2.8 °C in the formula for maximum temperature using elevation alone, to 2.5 °C in the formula that also included radiation, to 1.8 °C in the formula that also included canopy cover. Whilst these may not be indicative of the whole study area due to the stratification and selection of sensor locations, they imply the prediction errors are reduced by possibly 30–40%. Errors of this magnitude cannot be ignored as it has been suggested that 41% of eucalypts have a mean annual temperature range of less than 2 °C (Hughes et al. 1996). Including the canopy cover feedback into the prediction of maximum temperatures and vegetation is not ideal, but it may be necessary to include canopy cover in the models if the errors are going to be reduced to a satisfactory level. The best solution might be to obtain the canopy cover from alternative sources such as remote sensing (Wang et al. 2003).

Qualitative analysis of the GAM partial response graphs from this study suggests that the actual improvement in modelling results depends on which environmental factors are limiting each community. If a community is being limited by minimum temperature then including maximum temperature into the model may have less effect than a community that is being limited by maximum temperature. Also, the improvement depends not only on the predictive power for that community, but also on the degree to which that community can be distinguished from other communities, and which factors are limiting their distribution.

Previous studies have suggested that eucalypt communities and species may be significantly influenced by summer maximum temperatures (Passioura and Ash 1993) and/or winter minimum temperatures (Moore et al. 1993). These studies, and the results of this research, suggest that the limiting factors vary from species to species and

community to community, and it is unlikely that any one temperature predictor can differentiate the 23 communities found in this study area.

One problem with using multiple temperature predictors is that they can be highly correlated with each other – leading to problems in the GAMs. Lehmann et al. (2003) solved this problem by utilising the difference between the mean annual temperature and the winter average, however, in the Illawarra the average summer minimum and maximum temperatures were poorly correlated and could both be included in the models without introducing any problems with the GAMs. Moore et al. (1993) suggest that the relative contributions of elevation and radiation vary from summer to winter, and therefore it is possible that winter temperatures could be included as well if they are poorly correlated with the summer temperatures.

Austin (2002) suggests that it is difficult to determine whether poor vegetation models result from an unidentified environmental variable, or from other factors such as competition or poor dispersal. This study highlights that it could also be due to using inaccurate predictors, as suggested by Ferrier et al. (2002). This study has shown that the performance of vegetation models can be improved by developing more accurate estimates of seasonal maximum and minimum temperatures.

Whilst the improvement in prediction from using direct predictors has been noted (Austin and Meyers 1996; Guisan and Zimmermann 2000), it has also been suggested that they allow the model to be applied to wider areas (Guisan and Zimmermann 2000; Austin 2002). Two possible issues were seen during this study that may cause this to not always be the case. Firstly, the temperature range of each community and the overlap between them varied according to which formula was used to predict the maximum temperature. Clearly, it would be dangerous to apply the models to another area unless the maximum temperature had been calculated in the same manner as when the model

was developed—especially if the temperatures were based around air, soil, or canopy temperatures at a different height, taken during a different year, or during a different season. This is similar to the findings of Weiss and Hays (2005).

Secondly, the relationships in this study area are not necessarily applicable to the broader region. For example, *Eucalyptus sieberi* is known to be dominant on the Hawkesbury sandstone (Fuller 1995), which is found along the top of the escarpment in this study area. The models produced in this study area suggest it can only occur where the minimum temperature is low (due to high elevation), however it does grow where the Hawkesbury sandstone occurs at lower elevations to the north of the study area. It is unknown how many of the models reflect the ‘local’ rather than the ‘global’ environment, or whether other factors such as competition with new species would change the relationship in other areas.

2.5 Conclusions

Using the low pass filtered radiation and elevation for temperature predictions improved the estimates of maximum temperature for all combinations of elevation, radiation and/or canopy cover. The low pass filter also improved the temperature estimates for all 30-minute periods during the day from 8am to 9pm inclusive. The optimal radius was 750 m to 1000 m in all cases, but this may change in other study areas, in other seasons, or for other temperature sensor locations.

Including the improved estimates of maximum temperature in vegetation community models substantially improved the overall classification accuracy from 46.4% to 61.8%. This suggests that the effort spent to produce more direct or accurate predictors can reap large rewards, and it should not always be assumed that elevation is a good surrogate for temperature.

Chapter 3: The effect of exposure on landscape scale soil surface temperatures and species distribution models²

3.1 Introduction

Species distribution models (SDMs) provide information that is valuable for environmental planning activities (Ferrier et al. 2002), however better management outcomes may be achieved if landscape scale models are improved (Lookingbill and Urban 2003; Chuanyan et al. 2005; Lookingbill and Urban 2005). Current predictions may be inadequate if data used to produce models has insufficient spatial resolution or thematic accuracy (Guisan and Zimmermann 2000), or there are unsuitable predictors (Austin et al. 2006).

Spatial variations in temperature have a large influence on the distribution of vegetation (Lookingbill and Urban 2003; Pearson and Dawson 2003; Lookingbill and Urban 2005) and are therefore a vital component of SDMs. Temperature influences many ecological and physiological processes such as photosynthesis, energy and carbon balances, water and nutrient cycles, decomposition and mineralisation (Lookingbill and Urban 2003; Peng and Dang 2003; Bond-Lamberty et al. 2005). Both air and soil temperatures affect the growth rate and survival of plants (e.g. Peng and Dang 2003; Rocha Corrêa and Fett-Neto 2004).

Many temperature dependant ecological processes are non-linear in nature, which means that small differences in temperature can have a large influence on vegetation

² This chapter has been published as: Ashcroft MB, Chisholm LA, French KO (2008) The effect of exposure on landscape scale soil surface temperatures and species distribution models. *Landscape Ecology*, 23, 211–225.

(Wang et al. 2003; Bond-Lamberty et al. 2005; Weiss and Hays 2005). Hughes et al. (1996) found that 41% of eucalypts in Australia had a mean annual temperature range of less than 2 °C, and 25% less than 1 °C. Therefore, obtaining accurate temperature estimates is vital, and has become increasingly relevant due to climate change.

Predicting temperature variations at the landscape scale (extent of 10 – 200 km (Pearson and Dawson 2003)) poses different problems than at the global (> 10 000 km), continental or regional (200 – 2000 km) scales. Elevation may be the dominant factor at coarse scales, but variations in radiation, wind, moisture and aspect may be of greater importance when examining finer scale regions with low elevational ranges (Lookingbill and Urban 2003; Chuanyan et al. 2005). Consider a hypothetical situation where temperatures decrease at a rate of 6 °C/1000 m with fluctuations (mean = 0 °C, s.d. = 2 °C) that reflect variations in exposure (see Figure 3.6 in Supplementary Material). Variations of this magnitude are consistent with, for example, the 300 m elevational differences in the tree line in the European Alps (as noted by Frank and Esper (2005)). Despite that fact that these variations would be critical for many species, such as the eucalypts discussed above, the correlation between elevation and temperature is high ($r^2 = 0.893$) over an elevational range of 3000 m. This is because elevation is causing an 18 °C difference in temperature over its range, and variations in the order of 2 °C barely affect the correlation. As the elevational range of the study area becomes smaller, the variations in exposure become more comparable with the effect of temperature. For example, over an elevational range of 1000 m, elevation is only causing a 6 °C difference in temperature, and the variations in exposure reduce the correlation substantially ($r^2 = 0.539$, Figure 3.6). This illustrates that the range of conditions in a study area can affect how important each factor appears to be. If the ‘influence’ (coefficient * range) of exposure became greater than that of elevation, then

this would be expected to have dramatic effects on the strength of the correlation (r^2), and alter the predicted distribution of temperatures. This concept of ‘influence’ will be discussed further throughout this chapter.

SDMs often use elevation as a surrogate for temperature (Guisan et al. 1999; Hörsch 2003), or use elevation and location to improve interpolations from weather stations (e.g. Hughes et al. 1996; Coudun et al. 2006; Randin et al. 2006). Interpolations from weather stations may improve bioclimatic models at global to regional scales, but smoothing is not desirable at the landscape scale as complex heterogeneous patterns are common and need to be explained (Wu and Hobbs 2002; Wu 2006). Interpolation methods may also be prone to error in mountainous landscapes because weather stations are often sparse, and may not sample the full range of microclimates (Guisan et al. 1999; Guisan and Zimmermann 2000; Lookingbill and Urban 2003).

As a range of environmental factors determines the temperature of each site, predictions of landscape scale soil surface temperatures should incorporate as many of these influences as possible. The aim of this chapter was to improve the accuracy of landscape scale soil surface temperature estimates by utilising models with four environmental factors: elevation, radiation, moisture and exposure. The first three factors have already been shown to be important in other study areas (e.g. Lookingbill and Urban 2003), but the impact on SDMs needs further investigation. The potential influence of winds has been noted elsewhere without being quantified (Lookingbill and Urban 2003; Lassueur et al. 2006), and this chapter introduces exposure predictors that may serve as a surrogate for winds. These have the potential to not only quantitatively improve soil surface temperature estimates, but to differentiate temperature patterns at different times of the year as wind directions change. This may be especially beneficial

for species that are limited by conditions during a specific part of their lifecycle, such as flowering, growth, or germination.

This chapter focuses on soil surface temperatures rather than air temperatures. Soil temperatures have higher diurnal variations (Campbell and Norman 1998), are influenced more by radiation and canopy cover, and spatial differences are not affected as much by convection and mixing of air (Porté et al. 2004). Therefore soil surface temperatures have the potential to be more spatially heterogeneous than air temperatures, and may be better able to explain patchy vegetation patterns. Minimum and maximum temperatures are especially of interest because they can be less correlated with elevation than mean temperatures (e.g. Lookingbill and Urban 2003), and mean temperatures can give misleading indications of ecological processes (Bond-Lamberty et al. 2005).

3.2 Methods

3.2.1 Study area

The study was conducted on approximately 12000 ha of the Illawarra Escarpment and Woronora Plateau (34.4 °S, 150.9 °E), approximately 80 km south of Sydney, Australia (Figure 3.1). The study area contains a complex mosaic of moist and dry rainforests, moist eucalypt forests, tall open eucalypt forests, upland swamps and woodlands (NPWS 2002), with the City of Wollongong on the coastal plain. The canopy cover of trees ranges from 0% in upland swamps to almost 100% in moist rainforests. The elevation ranges from sea level to 573 m. Temperatures in the region have been associated with exposure to radiation, winds and coastal influences (Bywater 1985; Mills 1986; Fuller 1995), but these effects have not been quantified.

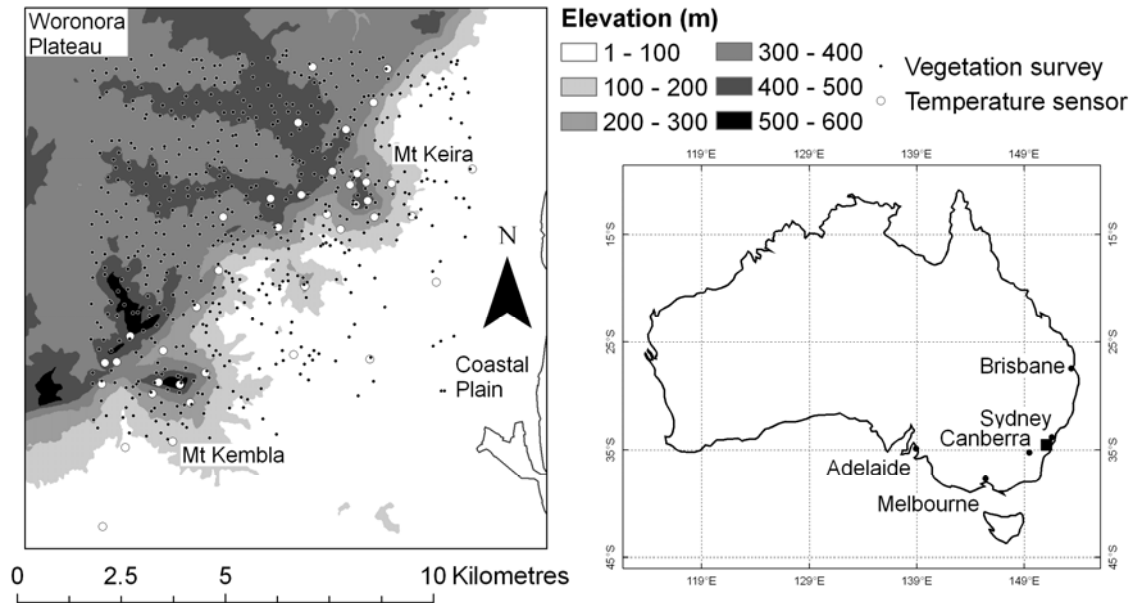


Figure 3.1: The Illawarra Escarpment, 80 km south of Sydney, Australia. The inset shows the topography of the study area as a digital elevation model, with the 600 vegetation survey sites and 40 temperature-monitoring sites shown by black and white circles respectively.

The most common wind directions at 9am (Table 3.2 in supplementary material) are from the NE and SW, while the most common at 3pm are from the NE and SE (Table 3.3 in supplementary material). The highest average wind speeds are associated with winds from the W to NW, and these are also associated with higher temperatures and lower humidities. For example, 3pm winds from the WNW in summer are associated with an average humidity of only 31% - less than half that from most other directions, and the maximum temperature is up to 8.5 °C warmer (Table 3.3). The wind direction at 3pm is associated with variations of 3.1 °C to 6.9 °C in maximum temperature during other seasons, whilst 9am wind direction is associated with variations of 1.4 °C to 3.4 °C in seasonal minimum temperatures. These statistics illustrate that wind direction can dramatically affect the temporal variations in maximum and minimum temperatures at a given site, however the purpose of this

chapter is to test whether differences in exposure will affect the spatial distribution of average maximum and minimum temperatures. That is, if WNW winds are hot and dry, are sites that are exposed to this direction hotter because they receive the winds more frequently, or because direct exposure to the winds has more effect? The influence winds have on local soil surface temperatures will be determined by both the frequency of winds arriving from various directions and the associated attributes of those winds (such as the ‘drying power’ caused by the wind speed and vapour pressure deficit). Therefore, WNW winds may have a large influence on local temperatures, even though they are rare.

3.2.2 Data

A vegetation survey of 600 sites (Figure 3.1) was completed between July 2005 and March 2006 (See Section 4.5.2 for more details on survey design). Each site (20 m by 20 m) was surveyed for presence or absence of 37 common species, of which the majority were trees (Table 3.1). All species will be referred to by their abbreviation hereafter.

Table 3.1: The 37 species from the Illawarra Escarpment that were modelled as part of this study. * marks an introduced species.

Abbreviation	Common name	Scientific name
TVH	Two-veined hickory	<i>Acacia binervata</i>
GW	Green wattle	<i>A. mearnsii</i>
LP	Lilly pilly	<i>Acmena smithii</i>
BS	Black she-oak	<i>Allocasuarina littoralis</i>
ROP	Red olive plum	<i>Cassine australis</i>
CW	Coachwood	<i>Ceratopetalum apetalum</i>
HC	Hairy clerodendrum	<i>Clerodendrum tomentosum</i>
RB	Red bloodwood	<i>Corymbia gummifera</i>
NC	Native cascarilla	<i>Croton verreauxii</i>
JW	Jackwood	<i>Cryptocarya glaucescens</i>
MG	Murrogun	<i>C. microneura</i>
PT	Prickly tree fern	<i>Cyathea leichhardtiana</i>
GST	Giant stinging tree	<i>Dendrocnide excelsa</i>
SF	Sassafras	<i>Doryphora sassafras</i>
MGG	Mountain grey gum	<i>Eucalyptus cypellocarpa</i>
YS	Yellow stringybark	<i>E. muellerana</i>
BB	Blackbutt	<i>E. pilularis</i>
SPM	Sydney peppermint	<i>E. piperita</i>
CWB	Coast white box	<i>E. quadrangulata</i>
SG	Scribbly gum	<i>E. racemosa</i> X <i>Haemastoma</i>
BGH	Blue gum hybrid	<i>E. saligna</i> X <i>botryoides</i>
SA	Silvertop ash	<i>E. sieberi</i>
GG	Gully gum	<i>E. smithii</i>
FRG	Forest red gum	<i>E. tereticornis</i>
BWR	Bolwarra	<i>Eupomatia laurina</i>
CSF	Creek sandpaper fig	<i>Ficus coronata</i>
LT	Lantana	<i>Lantana camara</i> *
CTP	Cabbage tree palm	<i>Livistona australis</i>
HLD	Hairy-leaved doughwood	<i>Melicope micrococca</i>
VMO	Veined mock-olive	<i>Notelaea venosa</i>
SP	Sweet pittosporum	<i>Pittosporum undulatum</i>
FW	Featherwood	<i>Polyosma cunninghamii</i>
WB	Whalebone tree	<i>Streblus brunonianus</i>
TT	Turpentine	<i>Syncarpia glomulifera</i>
SR	Scentless rosewood	<i>Synoum glandulosum</i>
BP	Brush pepperwood	<i>Tasmannia insipida</i>
RC	Red cedar	<i>Toona ciliata</i>

Soil temperatures were collected at 40 sites (Figure 3.1) from November 2004 until August 2006 using DS1921G iButton temperature loggers (Dallas Semiconductor/MAXIM). For the first 3 months (summer 2004/5) soil temperatures were recorded every 30 minutes, and the sensors were placed on the surface of the soil with as much protection from direct radiation as possible given the vegetation at the site. For the next 6 months (autumn and winter 2005) recordings were made every hour with the sensors located 1 cm below the soil surface to eliminate the effect of direct radiation, but potentially introducing errors when deposition and erosion changed the burial depth, or where differences in soil properties altered the effect of burial depth. For the final 12 months recordings were made every hour and sensors were located both on the surface and 1 cm below. Duplication of sensors was performed as consistent results would imply that neither direct radiation nor burial depth was causing the results to differ substantially from the other sensor placement (see Section 7.4 for a discussion on sensor placement). For this study I was primarily concerned with the spatial pattern of temperature estimates, as this was what was used to predict the distribution of vegetation. It was not critical if there was a consistent bias between subsurface and surface temperatures, unless the spatial pattern of temperatures changed.

The data was collected every three months at the end of the four seasons (February for summer, May for autumn, August for winter and November for spring). There was no data in the last week of every season as the data had to be downloaded and the sensors reprogrammed. Some sensors failed, were discarded due to spurious or outlying data, or could not be relocated. The sample size varied between 35 and 40 (median = 39).

The data was divided into 28 periods of 21 or 22 days (four periods within each season), but there were 44 datasets due to the duplication of sensors during the final 12 months. Dividing into months was not performed as the missing week at the end of each season meant that there was more data for some months than others. Three-week periods were selected as a trade-off between longer periods (more data but variable conditions) and shorter periods (less data but more constant conditions). Future studies may be able to use shorter periods if there are more sensors.

Twenty-four exposure predictors were developed based on the log transform of the angle to the horizon in selected directions. Kramer et al. (2001) showed that the angle to the horizon was a strong predictor of windthrow, and the exposure predictors are designed to capture the influence of winds. For the 24 azimuths that were a multiple of 15 (0 ° to 345 °), the ArcMap hillshade command was used with altitudes of 0.125, 0.25, 0.5, 0.75, 1, 1.25, 1.5, 2, 2.5, 3, 3.5, 4, 5, 6, 7, 8, 9, 11, 13, 15, 18, 21, 25, 29, 34, 39, 45, 51, 58, 65, 73, and 81 degrees. The ‘shadows’ that were produced for each altitude were combined into a raster grid for that azimuth so that each cell contained the minimum angle that resulted in no shadow. The value in each cell was effectively the angle to the horizon in that direction (azimuth). A low angle indicated that horizontal winds from that direction could blow almost directly onto the site, whereas a high angle indicated that they would need to ‘bend’ over or around the topography, as the site was sheltered. The angles in the raster grid were incremented by one and log transformed so that the final exposure predictor would be more sensitive to lower angles. For example, an angle of 0.125 ° became an exposure of 0.05 (very exposed), 2 ° became 0.48 (exposed), 9 ° became 1.00 (moderate), 25 ° became 1.41 (sheltered), and 81 ° became 1.91 (very sheltered). In effect, the exposure predictors are a measure of topographic ‘shading’, and this is very different from the aspect of the site. For example, a site can have a

northwest aspect, but be sheltered from that direction if it is behind a mountain (See Section 3.6.2 in the supplementary material for more details on the difference between exposure and aspect).

3.2.3 Temperature analysis

For each of the 44 datasets, the mean daily maximum, daily minimum and average temperature were calculated for each site. The spatially averaged minimum, maximum and average temperatures were then calculated for each dataset. This provided a general indication of the seasonal variation in soil temperatures.

Linear regression was used to determine the relationship between elevation and temperature. This involved 132 separate regressions, with 44 datasets of minimum, maximum and average temperatures. Multiple regression was then used to explain minimum, maximum and average temperatures for each dataset in terms of elevation, exposure and moisture. For average and maximum temperatures radiation was also used, but this was excluded from minimum temperatures as these usually occur during the night. Therefore, the equations used to predict temperature were:

$$T_{\min} = a_n + b_n.\text{Elevation} + c_n.\text{Moisture} + d_n.\text{Exposure}$$

$$T_{\text{ave}} = a_a + b_a.\text{Elevation} + c_a.\text{Moisture} + d_a.\text{Exposure} + e_a.\text{Radiation}$$

$$T_{\max} = a_x + b_x.\text{Elevation} + c_x.\text{Moisture} + d_x.\text{Exposure} + e_x.\text{Radiation}$$

The elevation predictor was the same in all regressions (see Section 3.6.1 in the supplementary material for more detail on the source and accuracy of predictors). A different radiation predictor was used in each regression, which was estimated for the central day of the recording period using the Solar Analyst (USDA Forest Service 2007) extension for ArcView (ESRI). The estimates were only based on topographic factors

— not cloud cover or actual measurements. The moisture predictor was either distance to streams (DS), log of distance to streams (LD), distance to coast (DC) or a topographic wetness index (TWI). The four temporally invariant moisture parameters (DS, LD, TWI, DC) and 24 exposure parameters (0° to 345° in 15° increments) were tested in each of 96 combinations, and the regression with the highest correlation (r^2) selected. The moisture predictors are designed to reflect soil moisture, proximity to coastal influences or creeks, but DS may also reflect cold air drainage at night. Note that no measurements of wind, moisture or radiation were made as part of this study, and I only examined the correlation between temperature and topographically derived surrogates.

For each regression, the coefficient of each predictor was multiplied with its observed range over the 40 sites in order to estimate the influence of that predictor in the study area. For example, if the coefficient for elevation was $6^{\circ}\text{C}/1000\text{ m}$ and the sensor locations spanned an elevational range of 500 m, then the predicted influence was 3°C . Converting the coefficients in this manner allows the influence of elevation to be compared with the influence of the other factors, as well as allowing seasonal comparisons. This is not meant to replace r^2 and P-values, but to complement them by looking at how much variation in temperature each predictor can cause in the study area. Not all predictors were significant in all regressions ($P < 0.05$), but none were removed because it was desirable to have the same number of parameters when making comparisons between different seasons.

3.2.4 Species distribution models

The distribution of each species was modelled using 134 Generalised Additive Models (GAMs, Hastie and Tibshirani (1990)). The first model contained only the geology predictor — obtained from a geology map (NSW Department of Primary Industries) and

used as a surrogate for soil properties (see Section 4.5.1 for more information on the geology of the study area). The second model contained the geology and elevation predictors. The remaining 132 models contained geology and either the predicted minimum, maximum or average temperature from one of the 44 datasets. The two-predictor models would explain more deviance than a model that only contained geology, but only the models with geology/elevation and geology/temperature were compared. All comparisons are between models with two predictors and so any improvement in deviance is not because of extra predictors.

The data was analysed using the 16 recording periods throughout the year (four per season). For each recording period, there were 1–2 above ground estimates and 1–2 below ground estimates. Spurious results could be detected if there was a lack of consistency between 2004/5 and 2005/6, or between surface and subsurface sensor placements. Temperature estimates were considered to perform better than elevation if they consistently explained more deviance, at least during certain seasons.

The results of the SDMs are included only to indicate which predictors may be more useful, and do not represent a final model. Developing such a model would require a more considered choice of statistical method, a predictor selection algorithm and more thorough independent validation. Multiple seasonal temperature estimates may be included in any SDM, and the significance of each may vary according to the statistical method and the other predictors in the model. Comparing models where each has geology and one other predictor ensures that any detected differences are due to the predictors themselves — not due to other factors.

3.3 Results

As predicted by Campbell and Norman (1998), the surface sensors had marginally higher maximums and lower minimums than the subsurface sensors. However, the

correlation coefficients were similar (low standard deviation) for each time of the year (Figure 3.2), and the factors affecting the spatial distribution of temperatures did not differ substantially. For example, the selected exposure directions at any time of the year were similar between 2004/5 and 2005/6 as well as subsurface and surface sensor placements (Figure 3.3). Therefore all results for the same time of year were averaged, as I was primarily interested in spatial patterns and the influence of different factors – not the absolute magnitude of temperatures.

3.3.1 Maximum temperatures

The mean seasonal maximum temperatures for the study area ranged from 10.7 °C to 21.8 °C (Figure 3.2a). The spatial distribution of summer maximum temperatures could not be adequately explained by elevation, with a mean (s.d.) r^2 of only 0.083 (± 0.037) in early summer. There was a clear seasonal trend that elevation had a lower correlation in early summer and higher in late autumn (Figure 3.2b), however the r^2 only reached a mean of 0.541 (± 0.127) at its peak. Including radiation, exposure and moisture in the regressions caused some large improvements in the ability to predict maximum temperatures. The r^2 for early summer maximums increased by 0.381 (± 0.052) to 0.464 (± 0.025) and winter maximums by 0.207 (± 0.044) to 0.598 (± 0.089) (Figure 3.2b).

Elevation had a fairly consistent influence (coefficient * range) on the study area of 4.0 °C (± 1.1) throughout the year (Figure 3.3a), representing a rate of change of approximately 7.7 °C/1000 m. The influence was highest in early spring and summer. The mean influence (coefficient * range) of radiation, wetness and exposure were 1.3 °C (± 1.3), 1.6 °C (± 1.0), and 2.3 °C (± 1.5) respectively (Figure 3.3a). These were much smaller than the influence of elevation, but were significant ($P < 0.05$, Tables 3.4, 3.5 in supplementary material) during a number of seasons. In particular, the mean influence

of exposure in early summer was 5.2°C (± 1.6), and this was higher than elevations mean influence of 4.9°C (± 1.8) during the same period.

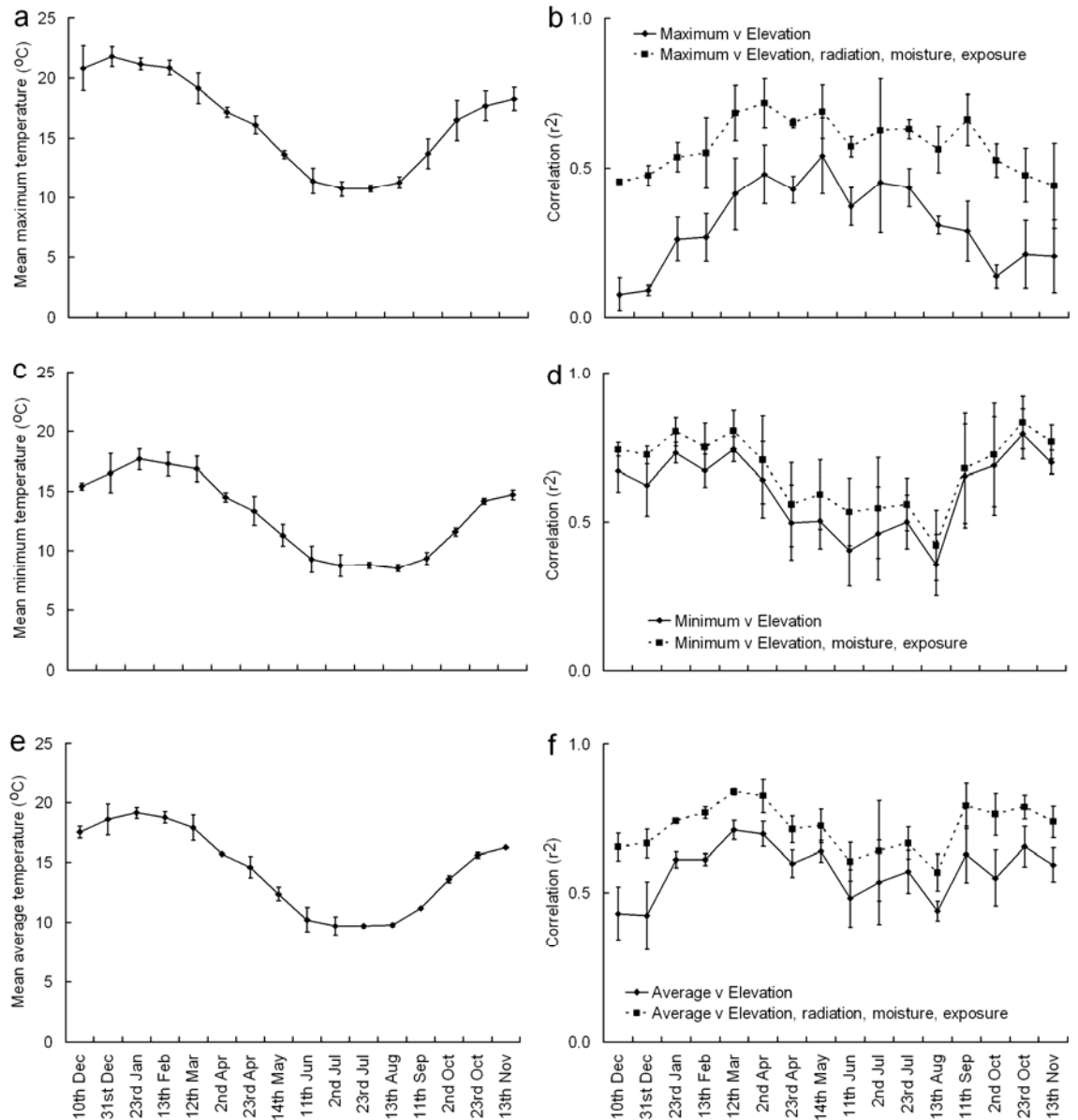


Figure 3.2: From top to bottom are the statistics for maximum (a-b), minimum (c-d) and average (e-f) temperatures. The left hand panels (a, c, e) illustrate the spatially averaged temperatures from 40 temperature monitoring sites. The right hand panels (b, d, f) illustrate the correlation (r^2) between the temperatures at each site and elevation, and the correlation between temperatures and a model containing elevation, moisture, wind and radiation. Each point is the mean (s.d.) that was produced by averaging the data from that time of year (2004/5 and 2005/6, surface and subsurface measurements).

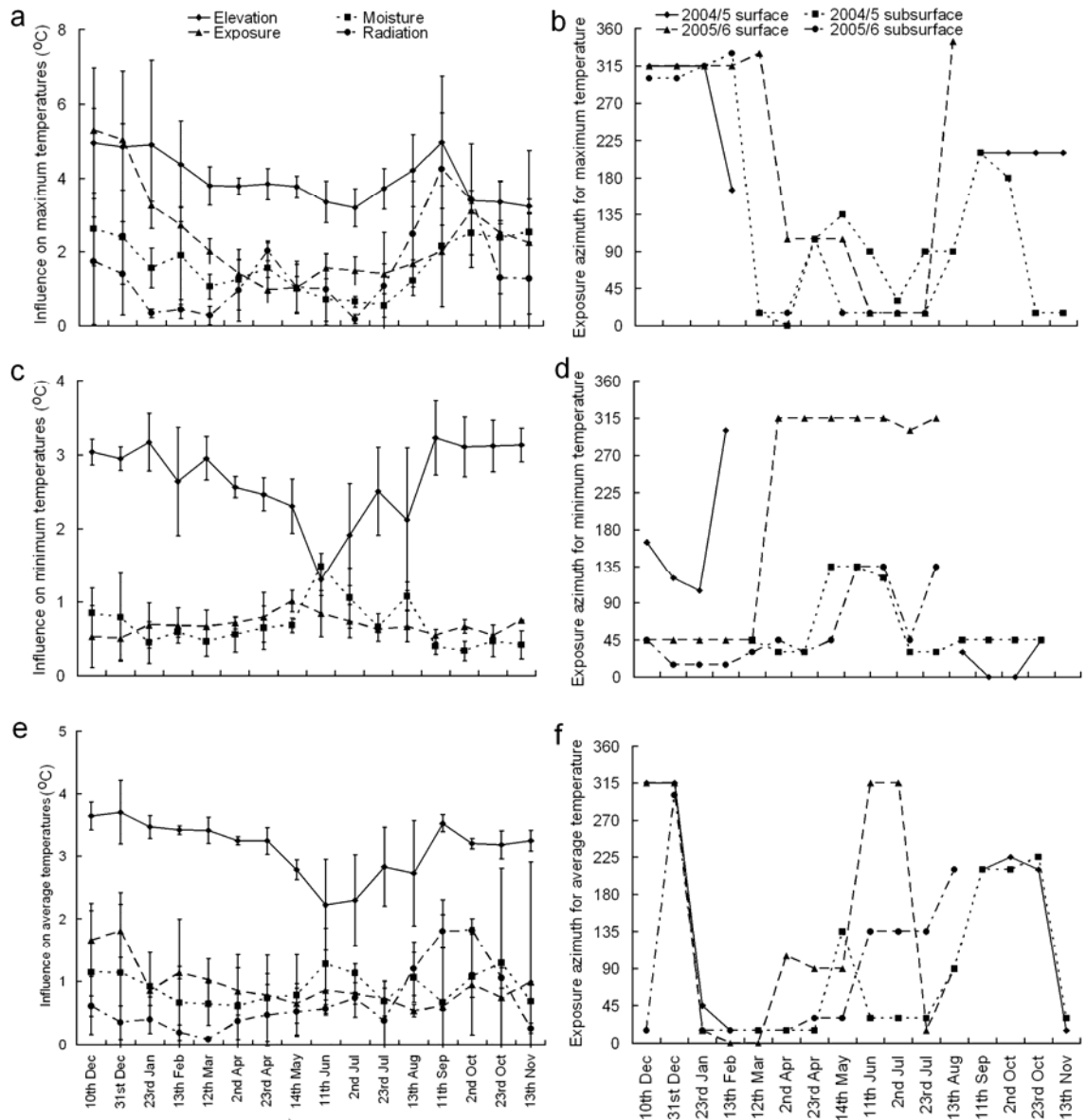


Figure 3.3: From top to bottom are the statistics for maximum (a-b), minimum (c-d) and average (e-f) temperatures. The left hand panels (a, c, e) illustrate the influence of radiation, elevation, moisture and exposure in the temperature regressions. Each point is the mean (s.d.) that was produced by averaging the data from that time of year (2004/5 and 2005/6, surface and subsurface measurements). The influence is determined by multiplying the coefficient of each predictor by the range of values observed over the 40 temperature monitoring sites. The right hand panels (b, d, f) illustrate the direction of the exposure predictor that was selected to maximise the correlation with elevation.

The selected directions for the exposure predictors corresponded somewhat closely with the wind directions that were most influential for temperatures at the BOM station (Table 3.3). Winds from the WNW were associated with the highest temperatures and lowest humidities in summer, and this corresponds somewhat with the selected NW exposure variables. Winds from the SSW in spring were associated with the lowest maximum temperatures, although SW winds were also cold and more common (Table 3.3). The selected exposure parameter in spring was SW in 63% of regressions and it always had a significant ($P < 0.05$) cooling effect. Winds from the N to NE were often amongst the warmest and most frequent directions during winter and autumn, and the selected exposure predictor was often from the NNE.

Radiation and moisture were more often significant ($P < 0.05$, Table 3.4) during autumn, winter and spring. The selected moisture parameter was DS in more than 50% of models during summer, autumn and winter, while LD was more frequently selected in spring (Table 3.4). In all of the models that included DS it was predicted that sites near streams would be cooler than those further away, but the parameter was only significant ($P < 0.05$) in 36% of models. Overall, moisture was significant ($P < 0.05$) in 23% of regressions for maximum temperatures, radiation 32%, exposure 75% and elevation 100%.

3.3.2 Minimum temperatures

The mean seasonal minimum temperatures for the study area ranged from 8.5 °C to 17.7 °C (Figure 3.2c). Elevation was correlated more with minimum temperatures than maximum temperatures, with the opposite seasonal trend (Figure 3.2d). That is, the correlation was highest in spring ($r^2 = 0.708 (\pm 0.112)$) and lowest in winter ($r^2 = 0.429 (\pm 0.117)$). The use of moisture and exposure predictors caused the r^2 for winter

minimums to increase by 0.085 (± 0.034) to 0.514 (± 0.122) and for summer minimums by 0.088 (± 0.060) to 0.732 (± 0.026).

There appeared to be little difference in the factors influencing the minimum temperatures for spring, summer and autumn, but winter temperatures presented a different pattern for both years and sensor placements. For spring, summer and autumn the mean influence (coefficient * range) of elevation was 2.9 °C (± 0.4), but this dropped to 2.0 °C (± 0.7) in winter (Figure 3.3c). These represented rates of change of approximately 5.5 °C/1000 m and 3.8 °C/1000 m respectively. The seasonal variations in the coefficients for elevation were consistent with previously reported results for air temperature in other regions (e.g. McVicar et al. 2007).

The influence of moisture was 1.1 °C (± 0.4) in winter, but this dropped to 0.6 °C (± 0.3) during the rest of the year (Figure 3.3c). In 75% of cases DC was the selected moisture parameter in winter, but it was only significant ($P < 0.05$, Table 3.4) in 44% of cases. Nevertheless, this was a clear trend, as DC was only selected in 19% of models during other times of the year, and was never significant. The coefficients suggested that sites nearer the coast had consistently higher winter minimums than those further away, confirming the continentality effect. DS was often significant ($P < 0.05$) for summer minimum temperatures, while LD was occasionally significant for summer and spring minimums.

The mean influence of exposure was only 0.7 °C (± 0.2) throughout the year. Exposure to the NE, and to a lesser extent N, were consistently selected and significant ($P < 0.05$) in summer, autumn and spring, whilst SE and NW exposures were often significant ($P < 0.05$) in winter (Table 3.5). These directions for the exposure predictors closely matched the directions of winds that were most influential for minimum temperatures at the BOM weather station (Table 3.2). These directions, as well as the

use of the DC predictor, suggest that coastal influences play a role in moderating minimum temperatures. Overall, elevation was significant ($P < 0.05$) in 93% of regressions for minimum temperature, moisture 25% and exposure 50%. Interestingly, exposure was noticeably more significant ($P < 0.05$) for subsurface temperatures than surface temperatures (71% versus 25%).

3.3.3 Average temperatures

The results for average temperatures were similar to the results for both minimum and maximum temperatures and are presented in Section 3.6.3 in the supplementary material.

3.3.4 Species distribution models

For seven of the 37 species (19%) elevation consistently explained more deviance than temperature over the whole year (Figure 3.4a). For another five species (14%), temperatures explained marginally more deviance at inconsistent times during the year (Figure 3.4c), or marginally more during winter (Figure 3.4e).

For the remaining 25 species (68%) temperature consistently explained much more deviance than elevation during at least part of the year. For five species, elevation explained approximately the same amount of deviance as temperature during most of the year, but temperature explained more during certain periods – usually winter, and usually minimum temperatures (Figure 3.4b). For another 13 species, temperatures explained more deviance than elevation over almost the entire year, with large improvements in summer and for maximum temperatures in particular (Figure 3.4d). These species were mostly moist rainforest species. For the final group of seven species, temperature explained more deviance than elevation over almost the whole year, but the

largest improvements were usually seen in winter and/or with minimum temperatures (Figure 3.4f).

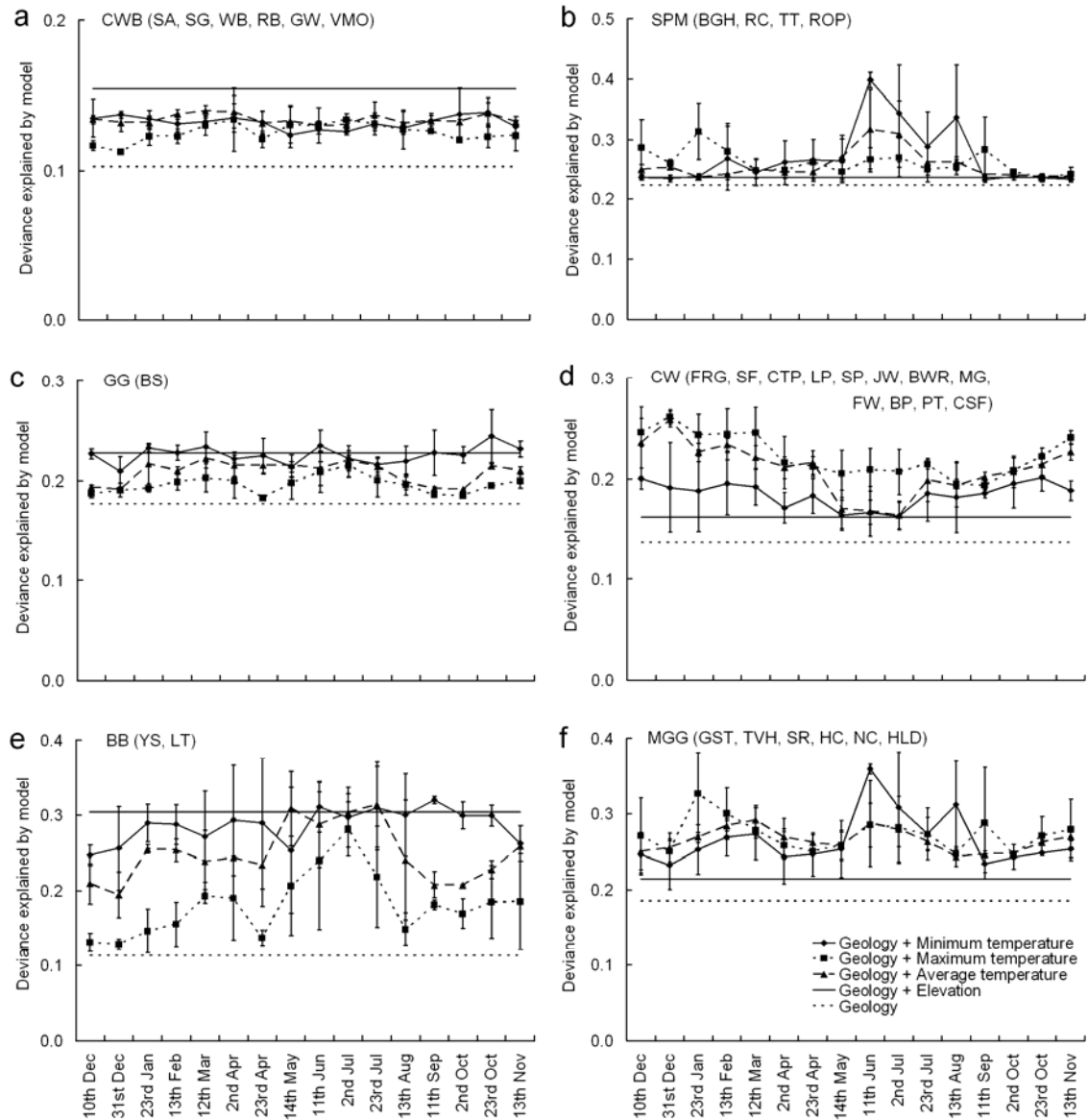


Figure 3.4: 132 SDMs were produced for each species based on different temperature estimates. The performance was assessed in terms of the amount of deviance explained (y-axis), and a model containing elevation and geology was compared against models containing temperature and geology. Each point is the mean (s.d.) that was produced by averaging the data from that time of year (2004/5 and 2005/6, surface and subsurface measurements). Data in each panel is for the first species listed, while the species in parentheses produced similar results.

The increase in the amount of deviance explained by models with summer maximums is not surprising given their low correlation with elevation. However the extra deviance explained when using winter minimums is somewhat surprising given that the r^2 values were only increased by a mean of 0.085 when moisture and exposure were included, and the predictors were not always significant ($P < 0.05$). The difference was that summer maximums and winter minimums were the two occasions when the influence (coefficient * range) of elevation was predicted to be less than that of the exposure and moisture predictors respectively (Figure 3.3a, c). This dramatically changed the spatial distribution of temperatures (Figure 3.5).

During seasons when elevation had a much higher influence than the other factors (e.g. summer minimums, Figure 3.3c), the predicted distribution of temperatures resembled that of elevation (Figure 3.5). This is despite the fact that DS, LD and exposure to the N and NE were often significant ($P < 0.05$). For winter minimums, the moisture and exposure parameters were not as often significant ($P < 0.05$), but the high influence of DC relative to elevation dramatically changed the expected temperature distribution. This new distribution matched the distribution of species that were only found at cooler locations away from the coast (e.g. MGG, Figure 3.5), and species that were only found on the escarpment slopes occurred at locations with a minimum of approximately 9 °C. For summer maximum temperatures the coolest locations no longer occurred at high elevations, but in locations that were at moderate elevations and sheltered from the NW. The map of low summer maximums closely matched the distribution of moist rainforest communities (NPWS 2002), and therefore it is not surprising that there was more deviance explained by the models for many rainforest species (e.g. CW, Figure 3.5).

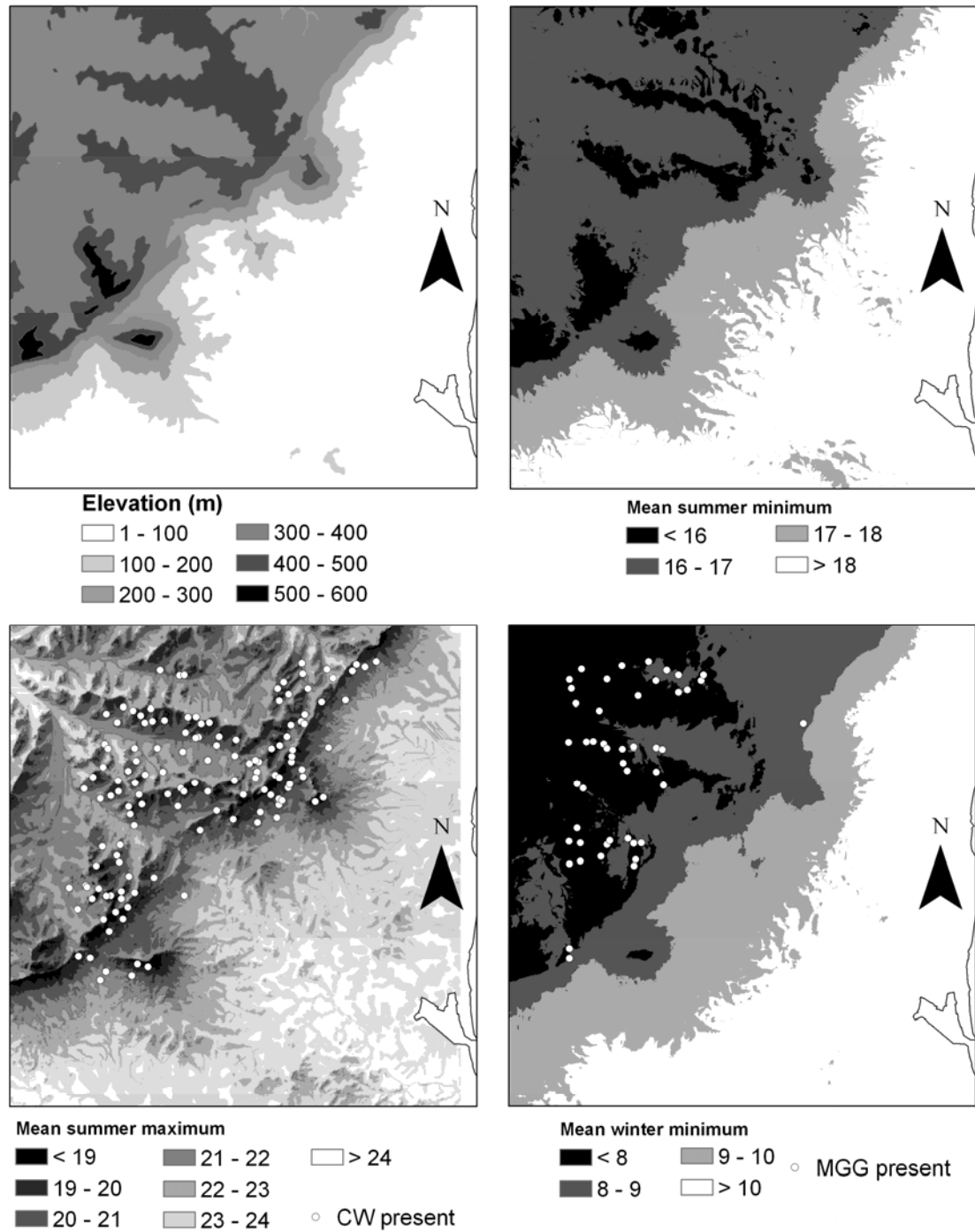


Figure 3.5: The surfaces of elevation, summer minimum, summer maximum and winter minimum temperatures. The temperature surfaces ($^{\circ}\text{C}$) were calculated by averaging all surfaces for each season (2004/5 and 2005/6, surface and subsurface measurements, 4 recording periods). Presence data for two species has been included on the surfaces that explained the most amount of deviance in their models.

3.4 Discussion

It is commonly suggested that climate is the dominant factor in SDMs at coarse scales, but factors such as topography and land-use are more influential at smaller extents (Pearson and Dawson 2003; Thuiller et al. 2004b; Coudun et al. 2006). This opinion may be proposed because of the widespread use of coarse-grained temperature grids produced by the interpolation of weather station data, which do not adequately capture higher resolution variability. With a 50 km grain size the within-cell temperature variation can be as much as 33.8 °C (mean 1.8 °C), as assessed using 1 km cells (Hijmans et al. 2005). The results of this study suggest there could still be large within-cell variation within the 1 km cells. Temperature variations may be equally important for SDMs at finer scales, but innovative methods are needed to develop high-resolution temperature surfaces that are sufficiently accurate to explain landscape scale vegetation mosaics. Technology now exists to create accurate estimates of direct environmental predictors at the landscape scale, and this could lead to a new generation of vegetation models and a better understanding of species' response to environmental gradients (Lookingbill and Urban 2005).

3.4.1 Factors affecting landscape scale temperatures

Elevation was clearly the most consistent, and usually most influential, determinant of temperature, but other factors were also found to be important. Elevation was particularly well suited to predicting the more moderate temperatures such as spring and autumn averages, summer minimums and winter maximums. Exposure and coastal influences were more important for the extreme temperatures such as winter minimums and summer maximums, and these are likely to be physiologically limiting the distribution of many species.

Coastal moderators had a large influence on winter minimums, with sites that were exposed to the NE or SE warmer than sheltered sites, and sites that were closer to the coast warmer than those further away. This is consistent with continentality effects, and interpolations from weather stations can detect similar trends at coarser scales (McVicar et al. 2007). The increase in r^2 was small and the predictors were not always significant ($P < 0.05$), but the influence (coefficient * range) of DC in particular exceeded that of elevation, and this dramatically changed the expected spatial distribution of winter minimum temperatures. Similarly, NW exposure was found to be more influential than elevation for summer maximums, and this also drastically altered the expected spatial distribution of temperatures. In each case, the new spatial distribution of temperatures was better able to explain the distribution of vegetation. The novel use of influence (coefficient * range) therefore appears to complement the r^2 and P-value statistics, as it provides an indication of which factors dominate the predicted spatial distribution of temperatures. High r^2 values do not necessarily mean that all the important predictors have been included (see introduction), and small improvements in r^2 can have a big impact on the predicted distribution of temperatures if the extra predictors are influential.

The exposure predictors that were developed for this study were very successful for landscape scale temperature prediction. They were significant ($P < 0.05$) in 60% of regressions, and this was more than both radiation (26%) and moisture (30%) predictors. Specific exposure directions were consistently selected in each season, and these closely matched the directions that were identified as the most influential wind directions based on BOM observations (Tables 3.2, 3.3) and literature on the study area (Mills 1986; Fuller 1985).

The exposure predictors were designed to capture winds but may instead be, for example, reflecting the meteorological conditions at the time-of-day most radiation is received. For example, if all else is equal, an east-facing site receives its radiation earlier in the day when the temperatures are lower, and hence vapour pressure deficits are lower. This means that the east-facing site will have lower rates of evapotranspiration than the west-facing site, all else being equal, and the east-facing site would be moister. This explanation would not, however, explain why a range of exposure directions were significant over the year, why the BOM station experiences vastly different temperatures under different wind directions (Tables 3.2, 3.3) or why the exposure directions in this study matched those from the BOM observations.

Other researchers are encouraged to test the exposure predictors in other study areas to determine if they are widely applicable, or whether they only become important in coastal mountain ranges. Larger study areas may benefit from techniques such as Geographically Weighted Regression (e.g. Foody 2004) so that the wind direction and/or influence can vary spatially.

The exposure predictors may also prove beneficial when placed directly into SDMs, as winds may have direct influences on plants as well as indirect influences through temperature. They are potentially much less sensitive to DEM errors than aspect because they are based on topographic shading (Van Niel et al. 2004). Therefore, a variable that represents shelter from both the east and west may be able to capture N-S gullies without the accuracy problems noted by Van Niel and Austin (2007).

The topographic wetness index (TWI) and log of distance to streams (LD) predictors were rarely significant ($P < 0.05$) in the regressions, although LD was often significant in spring. Distance to streams (DS) was the selected moisture predictor in most regressions, but distance to coast (DC) became increasingly important in winter -

especially for winter minimums. DC was also selected in some estimates of summer maximum, and so regressions with both DS and DC may prove beneficial. One outstanding question is whether the DC predictor should be modified to measure distance to coast in the direction of the prevailing winds rather than Euclidean distance.

Radiation was rarely significant ($P < 0.05$), but it might be more so if canopy or cloud cover was considered. Canopy cover is known to influence surface temperatures (Paul et al. 2004; Porté et al. 2004; Bond-Lamberty et al. 2005; Ashcroft 2006; Chapter 2), but it was not used in this study as it could potentially confound results when SDMs are produced. That is, it was undesirable to use vegetation to predict temperatures, and then use temperature to predict the distribution of vegetation. The effect of canopy cover may be higher during summer days and winter nights (Paul et al. 2004). This may indicate why summer maximums and winter minimums were not explained as well in this study as more moderate temperatures.

3.4.2 Which environmental factors influence the distribution of species?

Many of the SDMs in this study were dramatically improved by using soil surface temperatures, however further research is needed to determine why this occurs. It is possible that the distribution of some species is directly limited by seasonal temperatures associated with the optima for germination, growth, flowering or seed production, however temperature is not necessarily the limiting factor. Another possibility is that high soil temperatures are providing an indicator of low soil moisture. While topographic wetness indices are commonly used in SDMs, topography is only the dominant factor for moisture during periods of high rainfall (Moore et al. 1993; Ridolfi et al. 2003; Lookingbill and Urban 2004). However, under wet conditions, plants are less likely to find moisture the limiting factor. During dry periods, radiation,

temperature, soils and canopy cover can become more important (Ridolfi et al. 2003; Lookingbill and Urban 2004). Therefore higher soil temperatures may provide a better indication of where soil moisture will be low during dry periods. Lookingbill and Urban (2004) found that soil moisture increased with increasing elevation, decreasing distance to streams and decreasing radiation. All these factors were identified as leading to lower temperatures in this study. Moisture availability has already been identified as a strong determinant of species distribution (Leathwick and Whitehead 2001; Lookingbill and Urban 2003, 2004) and has been suggested as an important factor in other studies on the Illawarra Escarpment (Erskine 1984; Bywater 1985; Mills 1986).

Soil moisture may directly influence species distribution or it may indirectly influence distributions through mediation of fire frequency and/or intensity. The links between temperature, soil moisture and fire regime have been recognised (Lindenmayer et al. 1999; Schumacher et al. 2006), and fire has been cited as a factor affecting the distribution of vegetation in the Illawarra (Erskine 1984; Bywater 1985). Therefore, if soil temperatures provide an indication of moisture, they may also be providing an indication of fire regime. Further research is needed to determine whether temperature, moisture, and/or fire regime are influencing the distribution of each species. The results of this study only prove a correlation between temperature and the distribution of vegetation, and this could also be due to chance.

3.4.3 Why is elevation a better predictor for some species?

The fact that elevation explained more deviance than almost every temperature estimate for a group of species (Figure 3.4a, c, e) is worthy of further examination. This result may be because the species are limited by a number of different seasons throughout the year, and elevation provides a good indication of mean annual temperatures.

Alternatively, it could be that for these species the seasonal temperature of highest

importance is well correlated with elevation. The data collected during 2004–2006 may not represent the conditions present when the trees became established, and the distribution may be associated with past soil temperatures. Elevation may be associated with rainfall or other environmental factors, and the limiting factor may not be temperature. Finally, convection and mixing average out spatial variations in air temperature relative to soil surface temperatures. This would mean that the high spatial variability that is caused by differences in radiation, moisture and exposure are less prevalent in air temperatures, and species that are limited by air temperatures may be better modelled using elevation. There is some evidence to support this last possibility, as exposure was often more significant for the subsurface sensors than the surface sensors.

3.5 Conclusions

The direction of prevailing winds makes a large difference to the temperatures and humidities observed at a given site on a given day (Tables 3.2, 3.3). The exposure of each site to the most influential directions (e.g. NW) affects its average seasonal maximum and minimum temperatures. Incorporating exposure and coastal influences into temperature estimates can change the predicted spatial distribution of temperatures, and these are better able to explain the distribution of many plant species.

3.6 Supplementary material

3.6.1 Source and accuracy of predictors

Elevation was determined from a 25 m grain sized Digital Elevation Model (DEM, courtesy of NSW Department of Environment and Climate Change). Lidar data that is available for a subset of the study area showed the DEM generally contained elevational

errors of less than 10 m, but they were 30–50 m where there were steep slopes, and this was partly caused by the relatively coarse grain size. To reduce the ‘stepped’ appearance, the DEM was interpolated to a grain size of 5 m, although it is acknowledged that this did not eliminate all elevational errors.

The DEM errors were thoroughly assessed and will not affect the main results or conclusions of this or other chapters in this thesis. For example, elevational errors of 30–50 m are rare, but even these will only result in errors of 0.2–0.3 °C given that temperature changes at ~6 °C per 1000 m. These errors are trivial compared with the spatial and temporal changes in temperature, and the overall errors in the models specified (e.g. 1.8–2.8 °C in Chapter 2). The effect of DEM errors on derived environmental factors (e.g. aspect, slope, wetness indices) is more difficult to quantify (Van Niel et al. 2004), and the overall importance of each factor may be underestimated if it has high errors (see also Chapter 7). However, environmental factors are less useful if they are sensitive to errors, and should be avoided anyway. Indeed, we highlight that one of the benefits of the exposure predictors introduced in this chapter is that they are less sensitive to DEM errors. The temperature surfaces developed in this chapter are predominately based on variables that are relatively insensitive to DEM errors (elevation, distance to coast, distance to streams, exposure, radiation), and therefore the DEM has little impact on the models in this thesis that use those surfaces.

Total (direct plus diffuse) radiation was calculated for the central date of each of the 28 periods using the Solar Analyst (USDA Forest Service 2007) extension for ArcView (ESRI). No measurements of cloud cover or radiation were made during this study, and so the radiation predictor is an estimate of the insolation that each site would receive on a clear day based only on the topography of the study area.

Streams were located using the hydrology functionality of ArcMap (ESRI), where the flow accumulation was more than 500 cells. The distance to streams (DS) was calculated using Euclidean distance, and the log (distance to streams + 1) (LD) was determined. Streams may contain spatial errors, as the DEM was not hydrologically corrected, although the magnitude of the errors is expected to be small due to the typically steep gullies (see Figure 3.7). The topographic wetness index (TWI, Moore et al. 1993) was calculated using the formula:

$$TWI = \log(A/\tan(\beta))$$

where A is the upstream area (flow accumulation plus one times the area of each cell), and β is the slope. Distance to coast (DC) was calculated using a vector map of the coastline. Spatial errors of the Distance to coast (DC) predictor were negligible relative to the range. Of greater concern is whether lakes and harbours have the same effect as the open ocean, and whether the distance should be measured in the direction of the prevailing winds rather than Euclidean distance. The vegetation survey covered a broader range of DC (1.3 to 11.3 km) than the temperature sensors (1.9 to 7.6 km), and this means that the influence (coefficient * range) of DC may be underestimated in this study.

The only available soil map for the area was at a coarse scale (1:100,000), and was deemed to be insufficiently accurate. For example, it did not differentiate the sandy soils atop Mt Keira from the nutrient rich soils on the escarpment slopes. Therefore, a geology map (NSW Department of Primary Industries) was used in SDMs as a surrogate for soil properties, and contained spatial errors in the order of 50 to 200 m.

Whilst it was also at a scale of 1:100,000, it was more accurate than the soil map, probably due to the presence of economically viable coal deposits in the study area.

Due to poor GPS reception, spatial errors for the vegetation survey locations were estimated to be within 15 m for most sites, but possibly up to 50 m for sites in close proximity to cliffs and/or with a dense canopy cover. Some sites where the surface temperatures were recorded contained no trees, and so it was not possible to ensure consistency in microclimate by placing all sensors, for example, on the shady side of trees (as done by Lookingbill and Urban (2003)). Some sites used shrubs, groundcovers or debris for shelter while others used trees.

3.6.2 The differences between exposure and aspect

Aspect suffers from a number of problems, and exposure may offer solutions to many of these. Firstly, as aspect is measured on a circular scale, it cannot be easily interpreted (Pfeffer et al. 2003; Lassueur et al. 2006). It can be transformed using sine or cosine to yield predictors that reflect north-south or east-west trends (Hörsch 2003; Hengl et al. 2004), but this still does not cater for topographic shading (Graf et al. 2005; Pierce Jr. et al. 2005). Finally, errors in aspect can be large depending on the slope of the site, and how the DEM was derived (Wise 2000; Hengl et al. 2004; Van Niel et al. 2004). Pierce Jr. et al. (2005) conclude that aspect is unsuitable for mapping landscape scale vegetation patterns, and local studies (Bywater 1985) have stressed that local rainforest communities can have any site orientation as long as they have suitable shelter.

The differences between aspect and exposure are readily seen in the selected study area. Even if aspect is transformed (using $\cos(\text{aspect} - 315^\circ)$) to yield NW aspects, it differs substantially from the exposure to NW predictor (Figure 3.7). The aspect predictor is noisy in places where the slope is low, while the exposure predictor does not show any effects of DEM errors. Different predictors that are derived from the same

DEM can vary in their sensitivity to errors, and aspect is highly prone to error (Van Niel et al. 2004). It is clear (Figure 3.7) that the exposure predictor indicates the foothills of the escarpment are all sheltered from the NW, while they have variable aspect.

3.6.3 Results for average temperatures

The average seasonal temperatures for the study area ranged from 9.7 °C to 19.1 °C (Figure 3.2e). Elevation was a reasonably consistent representation of average temperatures throughout the year (Figure 3.2f), with a mean r^2 of 0.569 (± 0.109). The correlation was marginally better in autumn and spring than in summer and winter. The inclusion of radiation, moisture and exposure (Figure 3.2f) increased the correlation by a mean of 0.143 (± 0.053) to 0.712 (± 0.095). Once again, autumn and spring temperatures had marginally higher correlations.

The factors influencing average temperatures were a combination of factors affecting minimum and maximum temperatures. Elevation had an influence between that of minimum and maximum temperatures, and similar to the effect on minimum temperatures, was noticeably lower in winter (Figure 3.3e). Similarly to maximum temperatures, radiation was more significant ($P < 0.05$) in late winter and early spring, exposure to the SW had a cooling influence in spring, and exposure to the NW had a warming influence in early summer (Table 3.4, 3.5). Similarly to minimum temperatures, exposure to the N and NE was often significant ($P < 0.05$) throughout the year, and always had a warming influence. DS was often significant ($P < 0.05$) throughout the year, but DC was significant in winter and LD in spring. Overall, elevation was significant ($P < 0.05$) in 100% of regressions for average temperatures, radiation 20%, moisture 43% and exposure 55%.

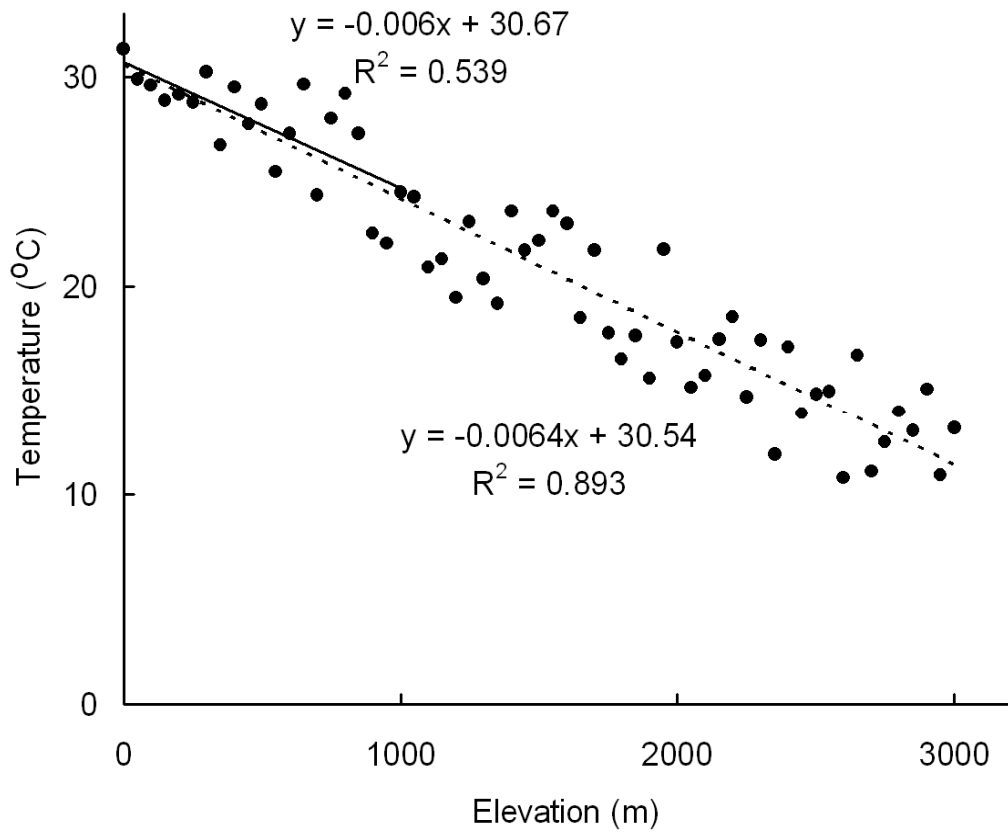


Figure 3.6: A hypothetical dataset created by assuming temperatures decrease at a rate of 6 °C/1000 m, but with fluctuations (mean = 0, s.d. = 2 °C) introduced to simulate variations in exposure. Two regressions have been performed – one over the full range of 3000 m, and one over only 1000 m.

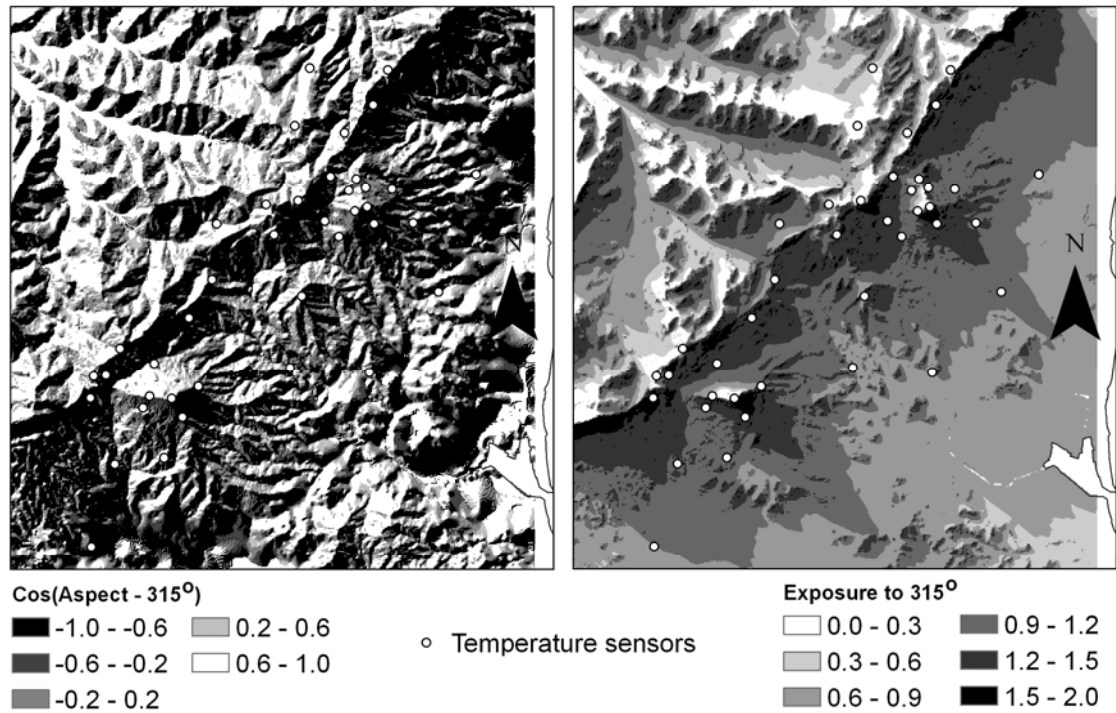


Figure 3.7: A comparison between northwesterly aspect (left) and northwesterly exposure (right). Aspect is more sensitive to DEM errors and does not consider topographic shading.

Table 3.2: The frequency (Freq) of winds at 9am over 35 years from 16 main compass directions. Data was obtained from the Bureau of Meteorology (they must be contacted to purchase it) and relates to daily observations from the University of Wollongong station (068188), which is located in the foothills east of Mt Keira (Figure 3.2). Recording period was March 1972 to February 2007 inclusive, and there were more than 3000 observations per season. Note that Summer is December to February, and Winter June to August. The Humidity (Humid), wind speed (Speed, km/h) and minimum temperature (Min, °C) were averaged over the days when winds from each direction occurred.

Wind	Summer			Autumn			Winter			Spring							
	Direction	Freq	Humid Speed Min	Freq	Humid Speed Min	Freq	Humid Speed Min	Freq	Humid Speed Min	Freq	Humid Speed Min						
N		5.0%	69%	7.0	17.6	3.3%	66%	7.3	15.2	3.2%	64%	7.0	9.2	4.3%	60%	8.9	12.7
NNE		4.2%	68%	8.7	18.2	2.5%	65%	8.7	15.8	2.1%	61%	10.0	9.0	3.4%	62%	11.3	13.4
NE		10.9%	70%	7.8	18.1	6.6%	67%	8.6	15.6	3.9%	63%	6.9	8.8	8.3%	62%	8.5	13.4
ENE		2.8%	71%	7.3	17.6	1.5%	69%	8.5	16.1	0.7%	64%	8.6	8.4	2.6%	67%	8.0	13.3
E		5.5%	70%	5.6	17.1	2.5%	70%	6.7	15.1	1.1%	64%	7.0	8.2	5.4%	63%	7.2	12.4
ESE		2.5%	71%	5.6	17.6	1.5%	71%	11.2	16.2	0.8%	65%	11.0	8.4	2.7%	65%	7.7	12.6
SE		7.1%	70%	7.5	17.3	3.7%	73%	7.4	15.4	1.9%	64%	6.5	8.8	6.9%	65%	8.0	12.9
SSE		2.3%	68%	10.8	17.4	1.8%	72%	9.8	15.0	0.7%	70%	14.2	10.8	2.7%	64%	11.2	13.0
S		7.8%	72%	8.7	17.2	7.4%	70%	8.7	14.3	4.4%	65%	8.8	9.2	6.8%	64%	11.1	12.3
SSW		6.5%	69%	12.4	16.9	7.1%	68%	12.4	14.5	3.8%	62%	13.9	9.5	7.1%	61%	13.4	12.0
SW		13.9%	69%	10.2	17.0	20.6%	66%	10.2	13.9	21.5%	60%	11.9	9.1	14.5%	59%	13.7	11.9
WSW		2.5%	66%	10.2	16.8	5.7%	61%	12.2	13.3	10.1%	56%	15.6	9.1	4.7%	51%	14.1	11.5
W		2.2%	59%	10.5	16.9	4.7%	55%	13.0	12.7	9.6%	51%	16.8	8.6	4.9%	47%	16.8	12.2
WNW		1.4%	52%	14.7	17.5	2.1%	53%	17.9	13.3	6.0%	49%	26.9	8.5	2.5%	46%	22.7	11.3
NW		4.6%	59%	12.3	17.7	5.4%	60%	12.0	14.1	8.5%	56%	16.2	9.0	7.3%	50%	15.0	13.0
NNW		2.2%	62%	12.9	18.1	2.5%	61%	10.5	14.5	2.2%	56%	11.4	9.4	3.5%	55%	13.3	12.6
Calm		18.7%	75%	0.0	17.8	21.1%	72%	0.0	13.9	19.6%	70%	0.0	8.2	12.5%	67%	0.0	12.6

Table 3.3: The frequency (Freq) of winds at 3pm over 35 years from 16 main compass directions. Data was obtained as per Table 3.2. The Humidity (Humid), wind speed (Speed, km/h) and maximum temperature (Max, °C) were averaged over the days when winds from each direction occurred.

Wind Direction	Summer			Autumn			Winter			Spring						
	Freq	Humid	Speed Max	Freq	Humid	Speed Max	Freq	Humid	Speed Max	Freq	Humid	Speed Max				
N	4.0%	69%	12.8	26.9	3.5%	64%	9.8	24.0	3.0%	58%	7.6	18.9	3.6%	61%	11.7	23.5
NNE	6.6%	68%	16.9	27.2	4.3%	66%	13.2	23.9	2.8%	55%	9.4	19.4	4.8%	66%	16.1	23.2
NE	23.6%	66%	16.1	26.6	14.7%	66%	11.7	24.4	8.2%	60%	9.4	18.9	19.5%	63%	14.7	23.1
ENE	7.2%	67%	15.0	25.9	6.3%	66%	10.9	23.6	3.6%	57%	9.1	18.1	6.2%	63%	13.2	22.0
E	8.2%	66%	9.7	25.3	8.1%	65%	7.1	22.6	5.2%	57%	5.7	18.4	7.7%	62%	8.8	21.9
ESE	5.4%	66%	10.3	24.5	4.0%	63%	9.3	22.2	1.8%	56%	7.5	17.3	5.2%	62%	10.3	20.4
SE	15.1%	68%	11.8	24.0	12.7%	65%	9.9	21.8	8.0%	61%	7.6	17.0	13.1%	63%	12.3	20.2
SSE	6.2%	66%	16.4	23.4	6.2%	65%	12.7	21.4	3.7%	59%	13.1	16.5	5.6%	64%	15.5	19.8
S	6.7%	69%	17.0	24.0	7.7%	66%	14.4	20.8	7.3%	60%	12.1	16.6	5.8%	66%	17.0	19.9
SSW	3.8%	70%	19.1	23.5	5.6%	67%	18.0	20.3	5.8%	57%	15.7	16.3	3.2%	65%	22.5	18.5
SW	5.0%	67%	18.4	24.7	8.0%	63%	14.6	20.2	13.9%	55%	14.1	16.8	7.0%	56%	19.5	20.5
WSW	1.0%	44%	22.7	27.1	2.7%	53%	14.8	21.5	6.0%	46%	16.8	17.0	3.0%	38%	21.8	21.7
W	1.1%	38%	17.8	29.5	3.0%	44%	15.4	22.2	8.7%	43%	15.8	17.9	4.0%	38%	20.1	23.5
WNW	0.6%	31%	22.6	32.0	1.8%	20%	20.0	21.7	4.5%	41%	25.3	17.6	2.2%	34%	28.2	24.0
NW	1.4%	51%	14.0	28.0	2.4%	51%	13.0	23.5	6.4%	45%	14.7	19.0	4.1%	45%	18.8	24.3
NNW	0.7%	57%	14.5	30.1	1.6%	58%	11.3	23.4	2.2%	48%	12.3	19.3	1.7%	47%	14.3	25.4
Calm	3.5%	78%	0.0	24.8	7.6%	66%	0.0	22.1	8.9%	62%	0.0	17.8	3.1%	73%	0.0	21.6

Table 3.4: Seasonal temperatures were regressed against elevation, radiation, moisture and exposure predictors. The numbers in this table indicate the percentage of regressions in which each predictor (other than exposure) was selected. The numbers in parentheses indicate how often they were significant ($P < 0.05$). Min, Max and Ave are short for minimum, maximum and average temperatures respectively. Blank cells indicate predictor was never selected.

	Elevation	Radiation	DC	DS	LD	TWI
Summer Max	100% (100%)	100% (0%)	33% (0%)	58% (0%)	8% (0%)	
Autumn Max	100% (100%)	100% (50%)	8% (0%)	75% (66%)	8% (0%)	8% (0%)
Winter Max	100% (100%)	100% (33%)	25% (0%)	50% (33%)	8% (0%)	17% (0%)
Spring Max	100% (100%)	100% (50%)		13% (0%)	50% (50%)	38% (0%)
Summer Min	100% (100%)		8% (0%)	75% (56%)	17% (50%)	
Autumn Min	100% (100%)		33% (0%)	33% (0%)	8% (0%)	25% (0%)
Winter Min	100% (75%)		75% (44%)	25% (0%)		
Spring Min	100% (100%)		13% (0%)	13% (0%)	25% (50%)	50% (0%)
Summer Ave	100% (100%)	100% (0%)		83% (60%)	17% (50%)	
Autumn Ave	100% (100%)	100% (8%)	25% (0%)	58% (43%)		17% (0%)
Winter Ave	100% (100%)	100% (17%)	67% (38%)	33% (50%)		
Spring Ave	100% (100%)	100% (75%)		25% (50%)	38% (100%)	38% (0%)

Table 3.5: Seasonal temperatures were regressed against elevation, radiation, moisture and exposure predictors. The numbers indicate the percentage of regressions in which each exposure predictor was selected (each column represents the combined total from three directions). The number in parentheses indicates how often they were significant ($P < 0.05$). Min, Max and Ave are short for minimum, maximum and average temperatures respectively. Blank cells indicate predictor was never selected.

	Direction of exposure predictor							
	345 - 15	30 - 60	75 - 105	120 - 150	165 - 195	210 - 240	255 - 285	300 - 330
Summer Max					8% (100%)			92% (91%)
Autumn Max	42% (80%)		42% (80%)	8% (0%)				8% (100%)
Winter Max	58% (57%)	8% (0%)	33% (25%)					
Spring Max	25% (100%)				13% (100%)	63% (100%)		
Summer Min	25% (100%)	42% (60%)	8% (0%)	8% (0%)	8% (0%)			8% (0%)
Autumn Min		66% (75%)		8% (100%)				25% (0%)
Winter Min		25% (0%)		42% (40%)				33% (25%)
Spring Min	25% (50%)	75% (83%)						
Summer Ave	50% (100%)	8% (0%)						42% (80%)
Autumn Ave	50% (67%)	17% (50%)	25% (0%)	8% (0%)				
Winter Ave	8% (0%)	25% (33%)	17% (0%)	25% (0%)		8% (0%)		17% (0%)
Spring Ave	13% (100%)	13% (100%)			75% (83%)			

Chapter 4: Addressing spatial autocorrelation in species distribution models without using survey data as a predictor

4.1 Introduction

The geographic distribution of a species is determined by factors such as its environmental niche, its dispersal ability, and interspecific competition (Pulliam 2000). Species Distribution Models (SDMs, Guisan and Zimmermann 2000; Rushton et al. 2004; Guisan and Thuiller 2005), however, usually focus on the environmental niche and neglect competition and dispersal (Guisan et al. 2006a). Where competition, dispersal or other spatial factors are considered, it is usually by utilising survey data in the surrounding neighbourhood (e.g. autologistic models), and taking advantage of spatial autocorrelation in the response variable or that of other species (Leathwick 1998; Araújo and Williams 2000; Leathwick and Austin 2001).

Spatial autocorrelation is a problem with respect to independent observations and can therefore affect the statistical significance of predictors in models (Legendre 1993; Keitt et al. 2002; Segurado et al. 2006). However, the statistical significance of predictors is not the only factor that is utilised to develop SDMs. Statistical skill, judgement and the ecological model utilised are now accepted as being important—maybe even more so than the statistical method (Austin 2002, 2007; Keitt et al. 2002; Guisan et al. 2006a).

Incorporating ecological theories into models can help develop new insights or questions (Olden et al. 2006), and combining niche and spatial factors has the potential to lead to models that are more useful and realistic than purely statistical and

niche oriented methods. There are limited ecological insights that can be gained from autologistic models because spatial clustering can be caused by a number of alternative ecological processes (Legendre 1993; Keitt et al. 2002; Dirnböck and Dullinger 2004; Wintle and Bardos 2006) and autologistic models cannot be used to test alternative hypotheses (analogous to McGill 2003). In addition, autologistic models cannot be used when there are sparse presences (Araújo and Williams 2000), and may simply be compensating for missing environmental factors (Leathwick 1998; Austin 2002; Guisan and Thuiller 2005). In an extreme situation, autologistic models can ‘explain’ a clumped distribution without any environmental factors as predictors (Araújo and Williams 2000; Bahn and McGill 2007), although this is interpolation rather than an actual explanation or prediction. Autologistic models may be useful for predicting the current distribution of species when the spatial distribution of direct predictors is poorly known and the ecological reason for the distribution is unimportant. However, they cannot be applied to different regions or times (Guisan and Thuiller 2005; Randin et al. 2006; Segurado et al. 2006), and are therefore inappropriate for climate change studies or predicting unknown populations in unsurveyed areas.

An alternative method to incorporate spatial factors into SDMs is by utilising contextual indices, such as the amount of rainforest within a 500 m radius (Ferrier et al. 2002; Wintle et al. 2005). In effect, they include spatial information using environmental factors (predictor variables) instead of the survey data (response variable). This reduces circularity in the model building process, and allows the models to be applied in unsurveyed regions and times. They have already been recommended for fauna with large home ranges (Ferrier et al. 2002), but the concept is also applicable to trees. This is because if there is more favourable habitat for a

species in the surrounding area, there are likely to be more trees of that species in the surrounding area, and therefore more seeds on the ground that can take advantage of any gaps that open in the canopy.

To illustrate the potential benefits of contextual indices, consider a hypothetical example where presence/absence data from 20 sites was available for a species in a study area containing two geological units, G1 and G2 (Figure 4.1). A traditional niche-based Generalised Additive Model (GAM; Hastie and Tibshirani 1990) assigns the same probability of existence (POE) to every location on each geological unit based on the ratio of presences to absences on that substrate (Figure 4.1a). This ignores the fact that, in this example, all presences are located near the large patch of G2 geology. A contextual model produced based on the amount of G2 geology within 150 m (for example) offers an alternative explanation for the observed distribution (Figure 4.1b), and produces a result more consistent with spatial or dispersal-oriented ecological theories such as island biogeography (MacArthur and Wilson 1967), source-sink models (Pulliam 1988, 2000) and fragmentation models (Fahrig and Merriam 1994; Hill and Curran 2003). Small patches of habitat have lower POE than the large patch, and the edges of the large habitat patch have lower POE than the centre. In addition, the contextual model provides some protection from spatial errors in the predictor maps and survey locations, as contextual indices create smoother transitions between discrete classes. This solves the problem whereby locations near soil boundaries may incorrectly be recorded on the wrong side of a boundary (as noted by Mummery and Battaglia 2002).

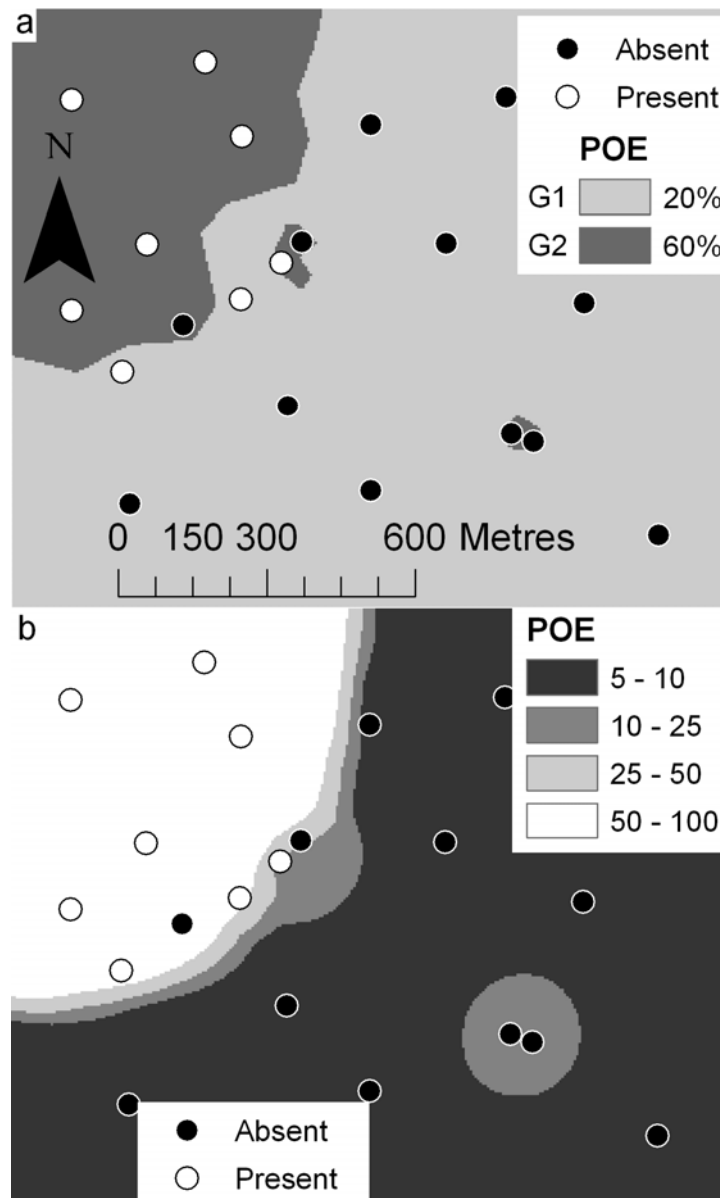


Figure 4.1: A hypothetical example for a study area with two geological units, G1 and G2, and presence/absence data for a selected species. Panels indicate the predicted probability of existence (POE) produced by a Generalised Additive Model without (a) and with (b) contextual indices.

Contextual indices are not a perfect representation of niche theory or spatial processes, but are a simple method to approximate the combined effects. It has already been recognised that environmental conditions need to be included in source-sink models (Pulliam 2000) and fragmentation studies (Hobbs and Yates 2003;

Debinski 2006), and including spatial processes in SDMs allows them to be more consistent with these other areas of ecology. Contextual indices are an extension of SDMs and are therefore not dynamic or process-based. However, process-based models often require potentially erroneous assumptions to be made about dispersal abilities or other ecological traits, as there is little information available on most species (Guisan and Zimmermann 2000; del Barrio et al. 2006). Contextual indices have the advantage that they can be fitted to presence/absence data without making *a priori* assumptions about the ecological traits of a species.

Throughout this chapter I use terms such as ‘patches’ for areas of favourable environmental conditions—even though they may be within a continuous forest. This differs from traditional fragmentation studies that often involve a matrix of cleared or disturbed land. Note also that I use the term ‘habitat quality’ as a measure of the ability to maintain a population under specific environmental conditions when spatial factors are ignored. The POE also considers the spatial context. Species can have high POE in ‘sink’ areas with low habitat quality if there is sufficient nearby ‘source’ habitat. Similarly, an isolated and small patch of high quality habitat can have a low POE. This terminology further illustrates the differences between contextual models and traditional niche-based models—where the habitat quality and POE are synonymous.

4.1.1 Extending the contextual indices methodology

The manner in which contextual indices have been applied to date (Ferrier et al. 2002; Wintle et al. 2005) is typically in fauna models, with a fixed and pre-specified radius (e.g. 500 m) and Boolean type predictors (e.g. rainforest or not rainforest). That implementation is a specific example of a more general methodology. For example, all cells could contribute equally to neighbourhood averages, or they could be

weighted such that distant points contribute less to the contextual indices (e.g. see Canham and Uriarte 2006). However, such comparisons are outside the scope of this chapter, and I focused on extending the contextual indices methodology by using continuous and categorical predictors, allowing the data to determine the optimal radius for each contextual model, and applying the models to flora rather than fauna.

I assessed these extensions using both single-predictor and multiple-predictor models. Single-predictor models were used to investigate alternative methods for incorporating continuous and categorical predictors, and to determine which factors influenced the optimal radius for contextual indices. Based on the best results from the single-predictor models, multiple-predictor models were developed to ensure the improvements were still valid when interactions between predictors were included. A more comprehensive comparison between contextual and non-contextual models was performed for one species.

4.2 Methods

The study was conducted on the Illawarra Escarpment and Woronora plateau, 80 km south of Sydney, Australia (34.4 °S, 150.9 °E). The escarpment runs NE to SW, with Mt Keira and Mt Kembla rising over the city of Wollongong on the predominately cleared coastal plain in the south and east (Figure 4.2). The uppermost geology in the study area, Hawkesbury sandstone, forms the summit of both mountains as well as the top of the escarpment. This geology supports vegetation communities that are vastly different from other substrates, and is dominated by eucalypt woodlands and upland swamps. The gullies on the Woronora plateau are predominately Bald Hill claystone and Bulgo sandstone, and support tall-open eucalypt forests, moist eucalypt forests, and rainforests (NPWS 2002). The escarpment slopes and foothills also contain moist eucalypt forests and rainforests, but have a different composition of eucalypts from

the gullies on the plateau. See the supplementary material (Section 4.5.1) for a more detailed description of the geology and vegetation.

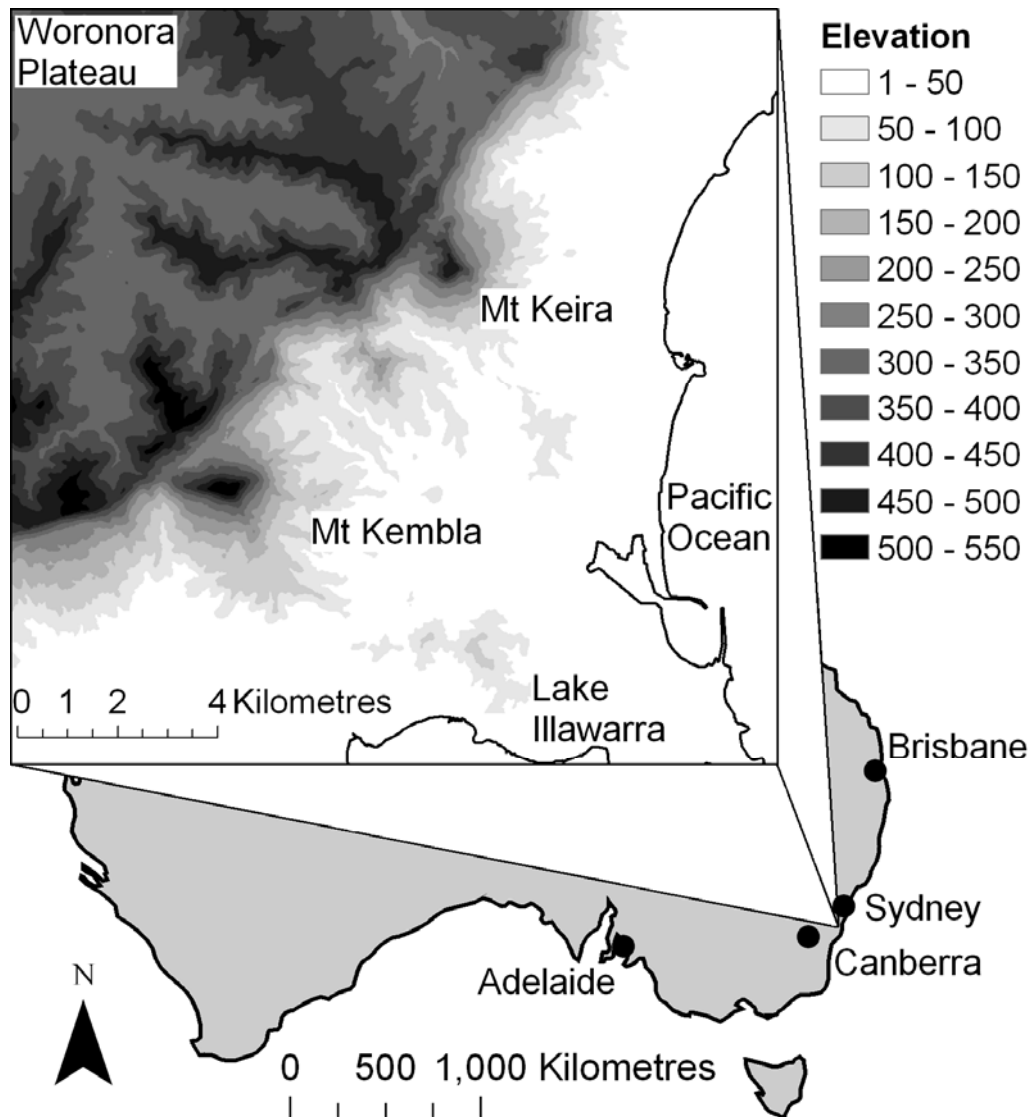


Figure 4.2: The topography of the Illawarra Escarpment in the vicinity of Wollongong, Australia (34.4 °S, 150.9 °E). The inset is a Digital Elevation Model showing the rising elevation (m) from the coastal plain to the Woronora Plateau, with Mt Keira and Mt Kembla protruding eastward. See Supplementary Material (Section 4.5.1) for maps of geology and vegetation.

Presence-absence data was collected (see Section 4.5.2 in supplementary material for details on survey design) for eight canopy tree species between July 2005 and

March 2006 at 600 sites (20 m by 20 m). All models were developed using the data from the first 360 sites, with additional sites used to test predictions in a simulated ‘unsurveyed’ area (see below). Silvertop ash (*Eucalyptus sieberi*) and red bloodwood (*Corymbia gummifera*) were selected because they were common in the woodland communities on the Hawkesbury sandstone; coastal white box (*E. quadrangulata*) and blackbutt (*E. pilularis*) were selected because they were common on the escarpment slopes; gully gum (*E. smithii*) and mountain grey gum (*E. cypellocarpa*) were selected because they occurred in plateau gullies; and, red cedar (*Toona ciliata*) and lilly pilly (*Acmena smithii*) were selected as two rainforest species. The number of presences for each species in the initial 360 sites ranged from 26 to 186. Two species had less than the recommended number of 50 presences (Stockwell and Peterson 2002; Coudun and Gégout 2006), but were still analysed as the fine spatial resolution in this study had the potential to produce better results with fewer presences than that recommended at coarser resolutions (Engler et al. 2004).

Seven environmental factors were used: elevation; average summer maximum temperature; geology; log of distance to streams; and, an estimation of the exposure to northeasterly, southerly, and west to northwesterly winds. Elevation was based on a 10 m-resolution Digital Elevation Model (DEM; see Section 3.6.1 for more information on accuracy). The average summer maximum temperature was based on the established relationship with low pass filtered elevation and radiation (Ashcroft 2006; Chapter 2). A geology map (NSW Department of Primary Industries) was used as a surrogate for soil properties, such as phosphorus, known to influence the distribution of Australian vegetation (Beadle 1954, 1966), and known to be influential for fine scale vegetation models (Coudun et al. 2006). Spatial errors of boundaries were 50 to 200 m (see Section 3.6.1 for more information on accuracy). The log of

distance to streams and the exposure to winds were derived using the DEM. Streams were located using the flow accumulation functionality of ArcGIS (ESRI, Redlands, US), whilst exposure to winds was based on the log transformed angle to the horizon. The complete derivation of the maximum temperature, distance to streams, and exposure to wind parameters has been provided by Ashcroft (2006; Chapter 2). The exposure to winds predictors were included as indirect predictors of landscape scale temperatures (see Ashcroft et al. 2008; Chapter 3), and were the best predictors that were available at the time the analysis was conducted. A comparison with other predictors and previous studies on the selected species is included in the discussion.

4.2.1 Single-predictor models

GAMs with single-predictors were used to compare the statistical performance of models with contextual indices with those produced using the same predictor without contextual indices. I also used these models to investigate how the optimal radius for contextual indices varied between different species and predictors, and how continuous and categorical predictors should be incorporated into contextual indices. It was not possible to simply apply contextual indices to the raw predictors. Categorical predictors cannot be averaged, and the average of continuous predictors, such as elevation, gives no indication of how much of the surrounding area is favourable habitat. Therefore, I converted the raw environmental predictors to 'habitat quality' before using contextual indices. I tested three methods by which this may be done, although other methods are also possible.

The first method used a GAM (using S-PLUS, Insightful Corp., Seattle, US) to produce a response curve for each species in terms of each environmental factor. This is effectively using the non-contextual model to generate the habitat quality for input to the contextual model. The outputs were exported to ArcGIS using a lookup table

(similar to Ferrier et al. 2002), and the predictor maps reclassified to produce habitat quality maps. These habitat quality maps were averaged using neighbourhood statistics to produce contextual indices. This approach may lead to excessively broad niches if there were numerous presences in sink areas (Austin 2002). Therefore, a second approach was used for each continuous predictor whereby 100% habitat quality was assigned to all locations where the environmental conditions were between the 10th and 90th percentile of the presences detected for that species, with 0% habitat quality assigned to everywhere else. Similar to using the response curves, the habitat quality was then averaged over the surrounding region to produce contextual indices. The third approach was identical to the second, but the 25th and 75th percentile were used. Each model was tested using contextual indices with radii of 30 m, 60 m, 100 m, 150 m, 200 m, 300 m, 400 m, 600 m, 800 m, and 1000 m, as well as the unfiltered predictor.

For each of the eight species and seven predictors the best GAM with contextual indices was selected by determining the radius and habitat quality selection method that maximised the deviance explained by the model (D^2 , Guisan and Zimmermann 2000). Each of the 56 single-predictor models without contextual information was compared with the best contextual model using that same species and predictor. Both the contextual and non-contextual models only had one predictor, as the contextual index was used instead of the raw environmental factor.

4.2.2 Multiple-predictor models

Multiple-predictor models were developed to ensure that the improvements in single-predictor models were not simply compensating for missing predictors. The best two-predictor models for each species were determined by testing each combination of the seven predictors. In each case the contextual and non-contextual models that

explained the most deviance were selected and compared. The multiple-predictor models were repeated using three to seven predictors. It is important to note that all comparisons are between models with contextual indices and models using corresponding raw predictor(s) instead of the contextual indices. Both models in all comparisons contain the same number of predictors and are based on the same environmental factors, although varying the radius of the contextual index could be viewed as an extra degree of freedom. For the contextual models, the seven candidate predictors were selected using the best radius and habitat quality selection method from the single-predictor models.

One species (*E. cypellocarpa*) was selected as a case study to do a more comprehensive investigation of a multiple-predictor model, and the contextual model for this species was assessed more comprehensively. The four most significant predictors in models for *E. cypellocarpa* (elevation, geology, and exposure to southerly and west to northwesterly winds) were selected for the comparison. The non-contextual model was produced using the four environmental variables as predictors. The contextual model converted each to habitat quality, calculated the average habitat quality within 500 m, and then used the contextual index in the model instead of the environmental variables. The radius was selected based on the average radius of the four predictors in the single-predictor contextual models. The habitat quality was determined using environmental envelopes, with the niche made as narrow as possible as long as all presences were within a reasonable geographical distance of favourable habitat. For example, *E. cypellocarpa* was actually observed on four geological units, but the presences on two of these, including Hawkesbury sandstone, were rare and only occurred in locations near more favourable units.

The statistical performances of the contextual and non-contextual models were compared using D^2 and Area Under the receiver operating characteristics Curve (AUC). The spatial autocorrelation of the residuals was examined using Moran's I (de Frutos et al. 2007) in ArcMap to determine if the contextual indices eliminated or reduced the problems discussed earlier. The Minimum Predicted Area (MPA) method was modified such that instead of using a fixed (e.g. 90% in Engler et al. 2004) threshold for occurrences, the full range of thresholds was examined (similar to AUC) and the average percentage difference between the two alternative models compared. In short, a lower MPA suggests a better model because a smaller predicted area contains the same proportion of presences. A new evaluation method was used which also examined the full range of thresholds. For each threshold, the map was classified into locations where the species was predicted to be present or absent. The average geographic distance of errors was estimated by calculating the minimum distance from each false positive to a recorded presence, and from each false negative to a recorded absence. A low average distance would indicate that most incorrect classifications were close to a location where they would be consistent with the survey data. Finally, the predictions of both the contextual and non-contextual models were compared using AUC in a simulated unsurveyed area. This was done using 167 of the 240 sites that were not used to develop the models. Not all sites could be used, as the original 360 sites did not cover the full range of environmental conditions, and I wanted to avoid testing extrapolation.

4.3 Results

4.3.1 Single-predictor models

Of the 56 single-predictor models developed (eight species, seven predictors), 55 (98%) were improved by including contextual information (Figure 4.3). The average magnitude of the improvement was 7.3% (s.d. = 6.4%), increasing the average deviance explained by the models from 13% (s.d. = 10.4%) to 20% (s.d. = 11.3%). This increase was significant ($P < 0.001$) according to a paired t-test. Only the best radius was selected to make these comparisons, but overall, 80% of the contextual models (response curve habitat selection method) performed better than the corresponding models without contextual indices. Choosing the optimal radius maximised the magnitude of the improvement—not merely select a rare case where an improvement occurred.

In 35 of the 48 models with continuous predictors (73%), at least one of the environmental envelopes with contextual information performed better than the models without contextual information. However, for 32 of the 48 models (67%) the best contextual result was obtained using the response curve for determining the habitat quality. This compares with 12 cases (25%) where the 50-percentile environmental envelope method performed best and four cases (8%) where the 80-percentile environmental envelope method performed best.

The optimal radius for contextual indices varied from 30 m to 1000 m. The average optimal radius for the seven predictors varied from 366 m for geology, up to 769 m for exposure to southerly winds (See Figure 4.9 in supplementary material). The general trend was that predictors that explained more deviance in the models, such as geology, elevation, and exposure to NE winds, tended to favour lower radii

than the predictors that explained less deviance. There were large differences in the optimal radius between species, with the median radii varying from 200 m for silvertop ash and lilly pilly up to 1000 m for coastal white box, blackbutt, and red cedar (See Figure 4.10 in supplementary material). The mean radii showed a similar trend to the medians, but one or two predictors that explained little deviance skewed results. The best predictors (highest D^2) for each species typically peaked at approximately the same radius (Figure 4.4), suggesting that the optimal radius of the best models was determined more by the species than the predictor.

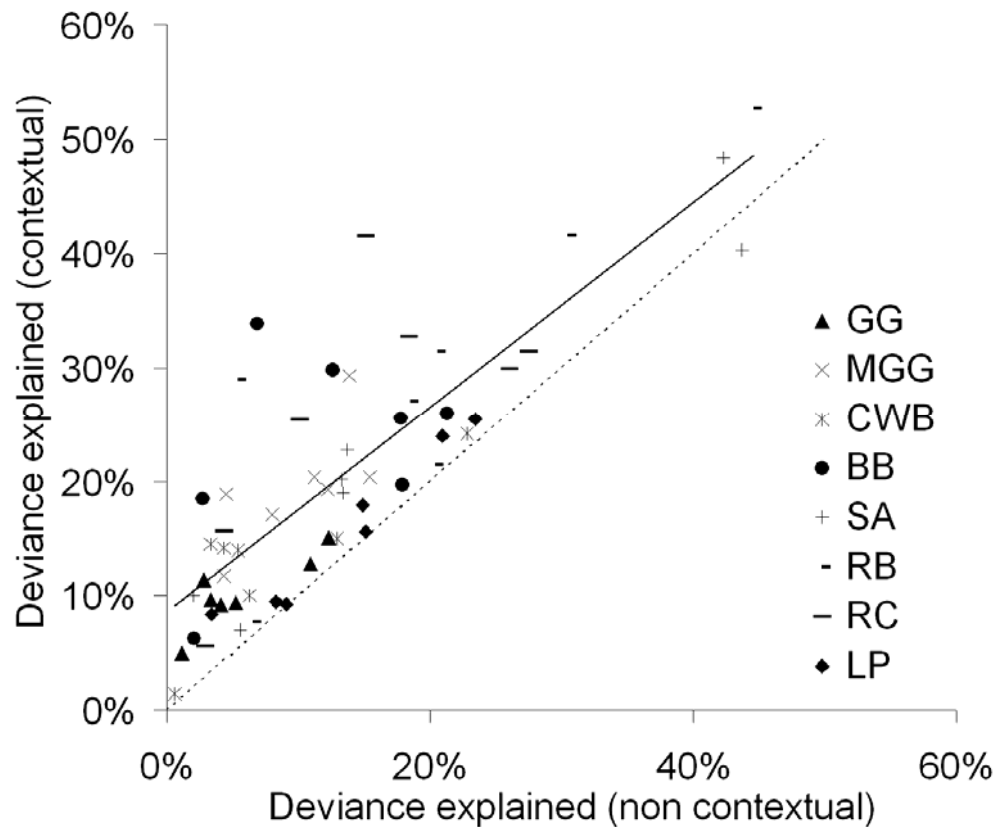


Figure 4.3: A comparison between the deviance explained by single-predictor models with and without contextual indices. The different symbols represent the eight species used in the study: gully gum (GG), mountain grey gum (MGG), coastal white box (CWB), blackbutt (BB), silvertop ash (SA), red bloodwood (RB), red cedar (RC) and lilly pilly (LP). There are seven points for each species – one for each of the seven environmental factors.

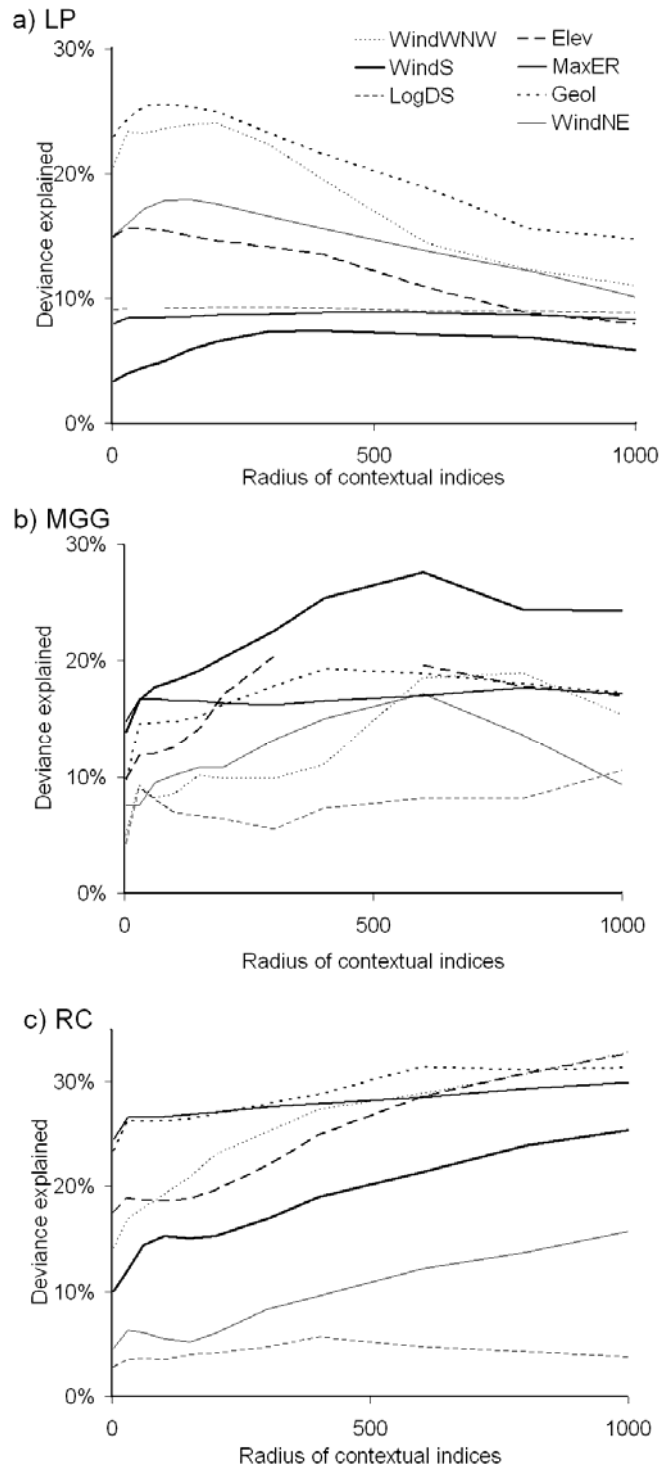


Figure 4.4: The deviance explained by the contextual models for lilly pilly (a), mountain grey gum (b) and red cedar (c) as a function of the radius of the contextual indices. The seven predictors are geology (Geol), elevation (Elev), summer maximum temperature (MaxER), shelter from WNW, S and NE winds (WindWNW, WindS, WindNE), and log of distance to stream (LogDS). Missing values occur due to problems with model convergence.

The two eucalypts from the gullies on the Woronora Plateau, gully gum and mountain grey gum, had similar optimal radii (median 600 m). This was also true for the two eucalypts from the escarpment slopes, coastal white box and blackbutt (1000 m). There was, however, a large difference between silvertop ash (200 m) and red bloodwood (600 m), even though both species occupied similar areas of Hawkesbury sandstone. The only noticeable difference between the distributions of the two species was that silvertop ash was observed atop small Hawkesbury sandstone ‘islands’ (Mt Keira and Mt Kembla), whereas red bloodwood was not. Red cedar had a larger radius (1000 m) than the other rainforest species, lilly pilly (200 m). Once again the species with the larger radius was only found in large fragments, while the species with lower radius was also found in smaller fragments.

4.3.2 Multiple-predictor models

Multiple-predictor models produced using contextual predictors performed 4.7% to 7.5% better than the equivalent models without contextual information (Figure 4.5), demonstrating that the statistical improvement in single-predictor models was mostly maintained when extra predictors were considered. To further investigate multiple-predictor models, a non-contextual GAM was produced for *E. cypellocarpa* using the predictors of elevation, geology, and exposure to west to northwesterly and southerly winds. The predicted probability of existence (POE) map that was produced by this model showed that although there were areas of moderate (25-50%), and occasionally high (>50%), POE that coincided with the observed occurrences of *E. cypellocarpa* (Figure 4.6a), similar POE values were also found in the central and northeastern portions of the study area—even though only one presence was recorded in these areas.

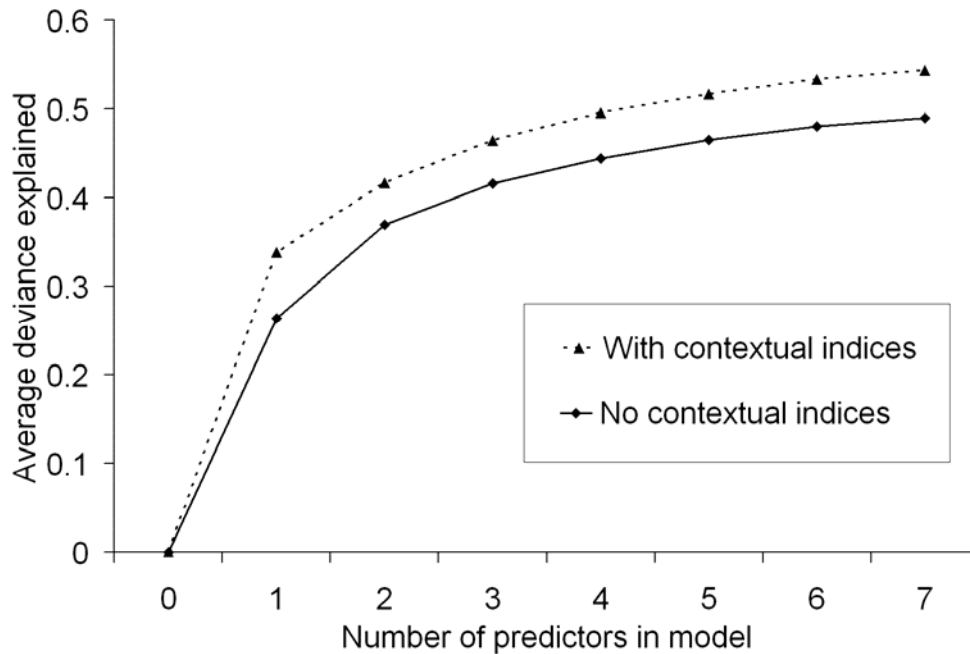


Figure 4.5: A comparison between the average deviances explained by multiple-predictor models, with and without contextual indices, for eight tree species in the Illawarra region.

The environmental envelopes that were selected for the contextual model for *E. cypellocarpa* were: elevation of 320 m to 420 m; exposure to southerly winds of 0.8 to 1.4 (moderately sheltered); exposure to west to northwesterly winds of 0.5 to 1.0 (exposed); and, geology of Bulgo sandstone or Bald Hill claystone. The GAM that was produced using the four contextual indices produced four similarly shaped partial response curves (See Figure 4.11 in supplementary material). According to each, the species was less likely to exist in locations where there was less than 40% favourable habitat in the surrounding area.

The model produced using the contextual indices explained more deviance (39%) than the model without contextual information (28%), and produced a dramatically different POE map (Figure 4.6b). In comparison to the non-contextual model, more

areas were classified as very unlikely (0% POE) or highly likely (>50% POE). There were three large patches of high POE, the northernmost two of which corresponded closely with two clusters where *E. cypellocharpa* was observed during the initial survey of the 360 sites that were used to develop the model, whilst the third was in the simulated unsurveyed area. The predicted POE in the central and northeastern portions of the study area was lower (0–25% instead of 10–50%), and this fitted the low observed occurrences in these areas better.

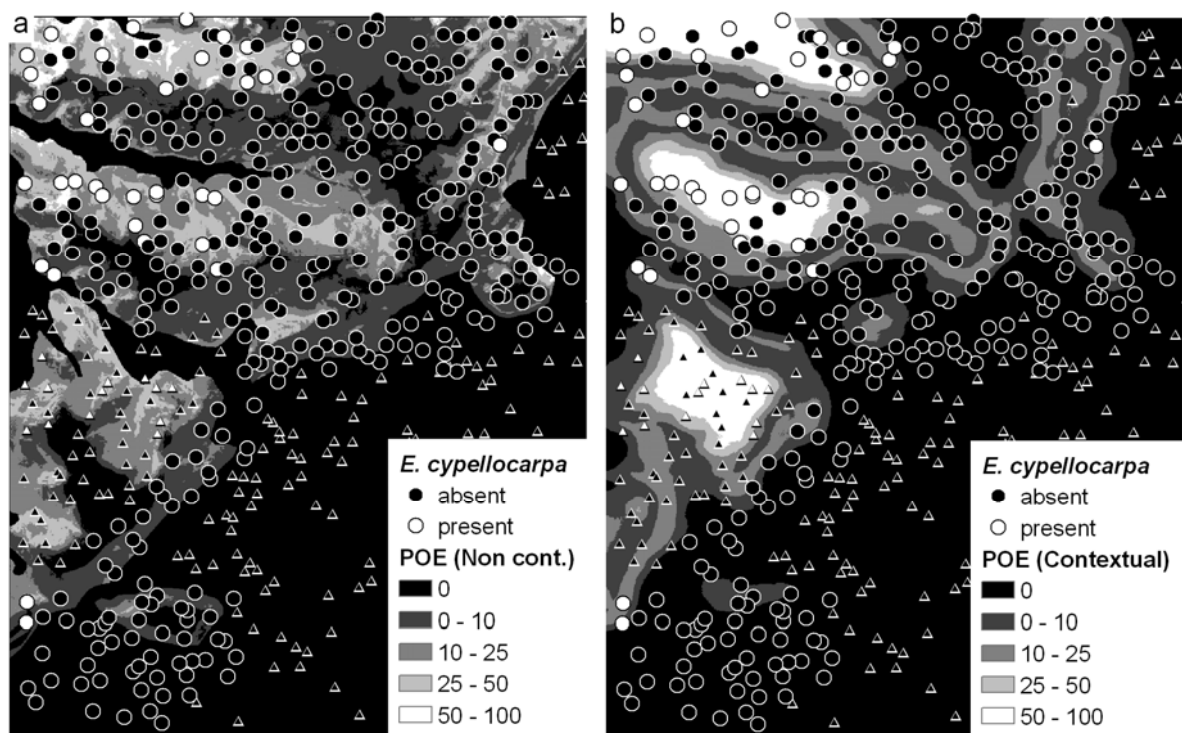


Figure 4.6: The probability of existence (POE) maps produced by the multiple-predictor models for *E. cypellocharpa* using predictors of elevation, shelter from southerly winds, geology and shelter from west to northwesterly winds. Circles represent the presence/absence data from the 360 sites that were used to train both models, while the triangles represent the additional sites that were used to test predictions in another area. The model produced without using contextual indices (Non cont., (a)) uses the same environmental variables as the contextual model (b).

The contextual model was substantially more confident (>50% POE) that the species would be found in the simulated unsurveyed area. The species was observed at 13 sites in that area, demonstrating that the prediction was accurate (Figure 4.6b). The model without contextual indices indicated there was a low to moderate (10-50%) chance of finding the species in that location, but this was a less confident, and less reliable prediction given that similar POE values were also predicted for other areas where the species was not observed.

The AUC in the models for *E. cypellocarpa* increased from 0.866 to 0.903 when using contextual indices, indicating better classification performance on the training data. The AUC calculated with data in the simulated unsurveyed area suggested a larger improvement in the contextual model—from 0.761 to 0.905. This confirms the better predictions in the simulated unsurveyed area.

The residuals of the contextual model were spatially autocorrelated ($P < 0.05$), but less so than the non-contextual model ($P < 0.01$). Therefore, contextual indices reduced, but did not eliminate, the amount of spatial autocorrelation in the residuals. The contextual model consistently, but not always, predicted a smaller MPA than the non-contextual model. The output was very sensitive to the threshold, with the contextual model varying from an area 32% larger to 59% smaller. The contextual model had lower MPA in 86% of cases, and was on average 17% smaller.

The contextual model for *E. cypellocarpa* also performed better in terms of the average spatial error of false presences and false absences. When the model output threshold was high, the false positives and false negatives were both within approximately 200 to 300 m (Figure 4.7) of a location where there was a recorded presence or absence respectively, and this was approximately equal to the distance

between survey points. Therefore, whilst they were erroneous, there were neighbouring sites where the prediction was valid.

As the threshold was lowered the number of true positives and false positives increased. For the contextual model the average distance of these false positive errors was consistently between 280 and 305 m when the threshold was set to achieve less than 55% true positives, but increased sharply as the threshold was lowered further (Figure 4.7a). The larger errors indicated that the false positives were consistently in locations further away from observed presences—raising doubts about the realism of the predictions. The non-contextual model showed a similar trend, but its false positive predictions started to appear in apparently unrealistic locations as soon as there were more than 30% true positives.

The average spatial error of false negatives did not change dramatically as the threshold changed, but the contextual model consistently had smaller errors than the non-contextual model. There was a slight trend towards smaller distances as the number of false negatives decreased (Figure 4.7b), but the results became sporadic when there were only a few errors. The trend may be more pronounced in common species, as absences would be rare.

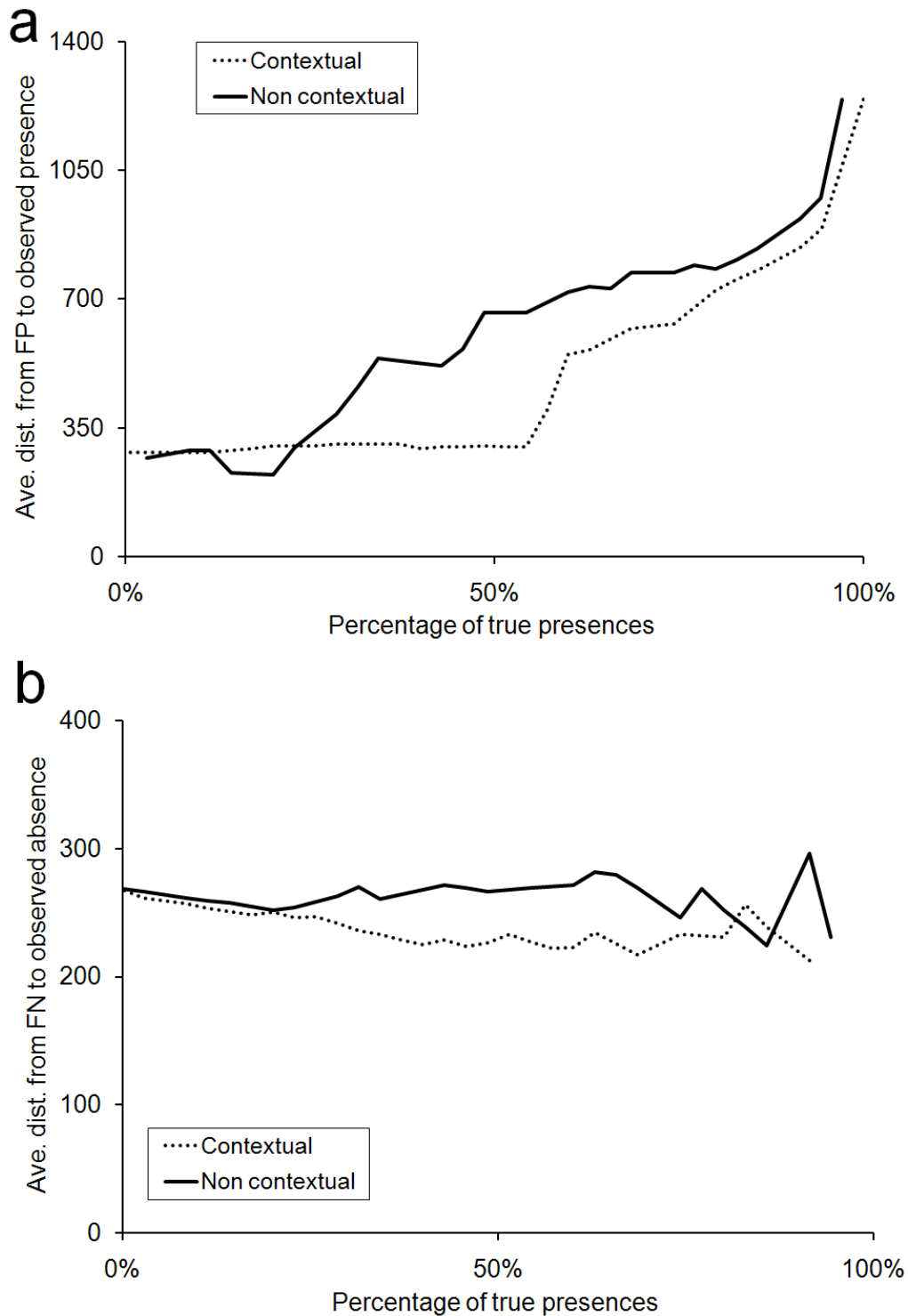


Figure 4.7: The estimated average spatial errors of (a) false positives (FP) and (b) false negatives (FN) as determined by averaging the distances (Ave. dist.) to the nearest presences and absences respectively. The two lines illustrate the results for the multiple-predictor contextual and non-contextual models for *E. cypellocarpa*.

4.4 Discussion

Using contextual indices to incorporate spatial factors into vegetation models produced statistically better results than purely niche-based models for both single and multiple-predictor models. The average deviance explained by single-predictor models increased from 13% to 20% using contextual indices, and average gains of 4.7% to 7.5% were observed in multiple-predictor models. The multiple-predictor contextual model for *E. cypellocharpa* performed better in terms of D^2 , AUC, Moran's I, average MPA, and the average spatial error of incorrect classifications.

The ecological interpretations of the two models for *E. cypellocharpa* are different, but it cannot be proven which conclusion is correct. The non-contextual model could not adequately explain the observed distribution, which implies that there was a missing predictor (Austin 2002), the species was not at equilibrium (Guisan and Zimmermann 2000), the predictors were not accurate enough, or the species was influenced by competition (Leathwick and Austin 2001), source-sink effects (Pulliam 2000) or dispersal limitation (Svenning and Skov 2004; Moore and Elmendorf 2006). On the other hand, the contextual model could explain the distribution in terms of an environmental envelope for each predictor, and suggested there needed to be more than 40% suitable habitat within 500 m. This could be because dispersal effects or other spatial processes were included in the model, but there is always the possibility that the improvements were due to an increasing correlation with more significant and missing predictors.

While the statistical improvement in performance of models containing contextual information is worth noting, the improvements in predictions are even more significant. After all, it is the predictive power of models that will ultimately

determine their usefulness (Rushton et al. 2004). The model for *E. cypellocarpa* that was based on contextual indices more confidently predicted where the species would be located by subsequent surveying, and its predictions were more accurate in terms of AUC. These predictions could not have been made with an autologistic model, as we were simulating a situation where there was no survey data in that vicinity.

The statistical improvement in the contextual model for *E. cypellocarpa* (28% to 39%) was larger than the average improvement that was reported for all models. Therefore, this result is not necessarily indicative of other species. Lilly pilly, for example, had a much lower optimal radius and there was less statistical improvement when using contextual indices. These factors combined to produce less obvious differences in the POE maps when comparing models with and without contextual indices.

4.4.1 What determines the optimal radius?

The results suggested that the success of contextual indices in vegetation models was more likely due to ecological factors than spatial errors or correlation between predictors. If spatial errors in GPS locations or plot size were the reason for the success, then the optimal radii should have been low—approximately equal to the errors. This was not the case, as the radii were typically 200–1000 m, while the errors were less than 50 m (and usually estimated to be less than 10 m). If the spatial errors in predictor layers were the reason for the success, then it was expected that the geology map would have a higher radius than other predictors due to its higher spatial errors. On the contrary, the geology predictor, and the other predictors that usually explained a high amount of deviance, had the lowest average radii. The larger radii of less significant predictors probably indicates an increasing correlation with more significant predictors, but this result is of little importance given that predictors that

explain little deviance are usually omitted. The large difference in radii between species was consistent with ecological differences, and is the most likely cause for the improvements in contextual models. The largest statistical improvements were when the optimal radius was large, and for species with a distribution that was clumped into large patches. Species that could be found in both large and small patches had lower radii, and smaller improvements.

It is already well known that species vary in their ability to exist in small patches of habitat (Hobbs and Yates 2003). Common species are typically observed in both large and small patches, but rare species are affected more by fragmentation and are likely to be found only in large patches (Honnay et al. 1999; Davies et al. 2000; Hill and Curran 2003; Debinski 2006). There is reduced seed rain in fragments (Hobbs and Yates 2003), but it is worth noting that the ability for species to persist in small patches is determined by its colonisation capabilities, not just dispersal. Colonisation ability is also influenced by factors such as fecundity, dormancy, seedling establishment characteristics, species interactions, and habitat quality (Fahrig and Merriam 1994; Levin et al. 2003; Levine and Murrell 2003; Guisan and Thuiller 2005). Dispersal may cause seedlings to be clumped, but such patterns can disappear as plants mature and habitat quality and competition have more influence (Levine and Murrell 2003). Therefore, although contextual indices do not capture an accurate short distance seed dispersal curve, there may not be a large benefit in doing so (e.g. by weighting cells in contextual indices according to distance from the centre) if other factors are more important in determining the colonisation ability. Indeed, I found that larger radii were associated with species with poorer dispersal ability, which is contrary to what would be expected if the radius reflected dispersal distance. This reinforces that view that contextual indices captured the combined effects of multiple

spatial processes— not just seed dispersal. I did, however, test one method for weighting based on radius (unpublished data), and found little difference in results. The extra complexity was not justified, and constant weighting was the more parsimonious solution. I cannot exclude the possibility that some methods of weighting could perform better.

I propose that species that can exist in both large and small patches will be good colonisers, have a low optimal radius for contextual indices, require accurate response curves to model their distribution, be more common, and have less potential for improvement from using contextual indices. The distribution of these species will be predominately determined by environmental niche factors. Conversely, poor colonisers would exist mainly in large patches, have a large optimal radius, be modelled less well using response curves, be less common, and obtain greater improvements from contextual indices. The distribution of these species is strongly influenced by fragmentation and source-sink effects. Some species may be distinctly niche or dispersal limited, but these are the extreme ends of a continuum (Gravel et al. 2006; Moore and Elmendorf 2006), and contextual indices offer a simple method to model species in between the two extremes as they consider both habitat quality and spatial factors.

4.4.2 Comparison with other models and predictors

Some of the species modelled in this study have been included in previous studies in an adjacent region, albeit at a coarser scale (e.g. Austin and Meyers 1996 and references therein). The models in the current study are not directly comparable with these models because of the differences in predictors and scale. Indeed, Pearson and Dawson (2003) propose that the environmental factors that affect the distribution of species are expected to be different given the difference in scale.

The predictors used in this study were the best that were available at the time the analysis was undertaken, but more accurate temperature estimates are now available (Ashcroft et al. 2008; Chapter 3). These predictors explain more deviance than those used in this study, and this reduces the benefit of using contextual indices (unpublished data). Indeed, it appears that the more deviance is explained by environmental factors, the less the improvement when contextual information is included (e.g. Figure 4.3). This makes intuitive sense in the context of niche and dispersal limited species, but it is also important to note that Bahn and McGill (2007) proposed that niche and spatial predictors were competing to explain the same portion of the variance rather than complementing each other. Combining niche and spatial factors using contextual indices is a promising method to investigate whether environmental factors only perform well because they are spatially autocorrelated, whether spatial autocorrelation terms are compensating for missing predictors, or whether there are truly a continuum of niche and dispersal limited species.

4.4.3 How should habitat quality be determined?

The contextual models produced significantly better results than non-contextual models, but they are almost certainly not optimal. There was some doubt as to whether response curves were the best method to determine habitat quality, as environmental envelopes performed better in a third of the models tested. New methods are needed to analyse data in both niche and geographic space without the need to transfer data between GIS and statistical packages used for modelling. This will increase the ability to answer ecological questions on the interactions between niche and spatial characteristics.

4.4.4 Further recommendations

Contextual indices were originally justified in this study because they could approximate the supply of seeds at a given site, although the results suggest that seed dispersal is not the only reason for the success of contextual indices (see above). Nevertheless, if results are influenced by seed production and dispersal, then there are a number of ways in which models could be further improved. Firstly, not all individuals may produce the same number of seeds. The conversion to habitat quality may produce better results if it is based on the number of seeds produced under given environmental conditions rather than simply the presence or absence of the species. Secondly, the output from the contextual model may be better used as a predictor rather than as a final model. That is, for *E. cypellocarpa* a better model may have included the four raw environmental factors as well as the output from the contextual model. This was not performed here, as it would mean the models did not have the same number of predictors and could not be as easily compared. Nevertheless, it would make more ecological sense, as two sites with the same supply of seeds may have different POE under different environmental conditions.

4.5 Supplementary material

4.5.1 Geology and vegetation communities in the study area

The Hawkesbury sandstone (TRh, Figure 4.8) is the most distinct unit in the study area. It forms the top of the escarpment, including Mt Keira and Mt Kembla (Figure 4.2), and supports vastly different vegetation communities to the rest of the study area. These communities include Escarpment Edge Silvertop Ash Forest (EESAF), Exposed Sandstone Scribbly Gum Woodland (ESSGW), Cliffline Coachwood Scrub (CCS), and Sandstone Gully Peppermint Forest (SGPF, all communities in shades of

grey in Figure 4.8). Upland swamps develop where the drainage on the Hawkesbury sandstone is poor. The geology for these upland swamps (Qs) is distinguished from the rest of the Hawkesbury sandstone. The upland swamps are often treeless, and support sedgelands, shrublands of species such as *Hakea* spp. (USSHC – Upland Swamps: Sedgeland-Heath Complex), or dense thickets of *Banksia* spp. or *Leptospermum* spp. (USBT – Upland Swamps: Banksia Thicket, both communities shown in shades of yellow in Figure 4.8).

Deep valleys occur on the Woronora Plateau (Figure 4.2), draining into water supply dams (shown in red in Figure 4.8). The slopes of the valleys consist of Bald Hill claystone (TRnz) and Bulgo sandstone (TRnbu), with Hawkesbury sandstone forming the ridges in between. The Hawkesbury sandstone forms cliffs of up to 40 m where it meets the Bald Hill claystone. Where the gullies are exposed to hot, dry NW winds, temperatures are warmer (see Ashcroft et al. 2008; Chapter 3), and tall open eucalypt forests are found (i.e. Tall Open Blackbutt Forest – TOBF, Tall Open Peppermint-Blue Gum Forest – TOPBGF, and Tall Open Gully Gum Forest – TOGGF, shades of orange in Figure 4.8). Rainforest species are uncommon in the understorey of these communities. In contrast, where there is shelter from the NW winds, moist eucalypt forests (shown in shades of green) and rainforests (shades of blue) occur. Communities include Moist Gully Gum Forest (MGGF), Moist Blue Gum-Blackbutt Forest (MBGBF) and Coachwood Warm Temperature Rainforest (CWTR), all of which contain many rainforest species in the understorey.

The escarpment slopes consists of numerous layers of sandstones, claystones, and coal seams from the Narrabeen Group and Illawarra Coal Measures. There are also areas of Quaternary alluvium and talus (Qa, Qt). The escarpment slopes support a variety of moist eucalypt forests and rainforests, including Escarpment Blackbutt

Forest (EBF), Escarpment Moist Blue Gum Forest (EMBGF), Moist Coastal White Box Forest (MCWBF), CWTR, and Illawarra Escarpment Subtropical Rainforest (IESR). The foothills support many of the same species, but also contain Moist Box-Red Gum Foothills Forest (MBRGFF) and Lowland Dry-Subtropical Rainforest (LDSR). Much of the coastal plain is Cleared (C), or Modified Lands (ML) due to the City of Wollongong. Disturbance related communities (shown in shades of purple) occur throughout the study area, but are especially common on the lower escarpment. These include Acacia Scrub (AS), Turpentine Regeneration (TR) and Weeds and Exotics (WAE). While these communities are disturbed, they are generally semi-natural communities (unmanaged land), and support many native species.

While the distribution of vegetation communities is known, there is less information available on individual species (but see Fuller 1995 for rough maps). It is unknown which factors are influencing the distribution of species and communities. For example, there is a distinct NE-SW trend in eucalypts on the escarpment slopes (from EBF to EMBGF to MCWBF) and it is unknown why this occurs. Each of the dominant species (Blackbutt, Blue Gum, and White Box) occurs at similar elevations and on similar geologies.

4.5.2 Survey design

A survey of presence and absence of 37 species was undertaken with the explicit purpose of species distribution modelling in mind. The target species (listed in Ashcroft et al. 2008; Chapter 3) consisted of common rainforest and eucalypt species, as well as a selection of disturbance related species. I was particularly mindful that I needed 50 presences of each species for modelling to be robust (Stockwell and Peterson 2002; Coudun and Gégout 2006), and also that models would be most useful if the dominant eucalypts were included. The documentation on the communities in

the study area (NPWS 2002) gave approximate frequencies of each species in each community. Based on these figures, I determined that a survey of 600 sites was necessary to achieve at least 50 presences of the common eucalypts. This approximation proved accurate and I achieved the target of 50 presences for most species.

The survey locations were randomly chosen subject to a number of constraints that were imposed to ensure that a representative and complete range of communities and environmental conditions were sampled. First, the proportion of each community in the study area was used to determine the approximate number of samples that should be taken from each community shown in Figure 4.8 (apart from ML and C). Some rare communities are not shown. Some communities were very common (e.g. CWTR) and the number of samples in these communities was reduced. Other communities were less common, and it was necessary to over-sample these communities to ensure that the full range of environmental conditions could be sampled.

Once the number of samples for each community had been determined, a list of potential locations was placed in a random order. The highest ranked locations were selected provided the locations were spread geographically and environmentally, and I had permission to access the land. If a potential location was too close or too similar to another location then it was discarded and the next site in the list used. I was particularly careful to ensure that, where possible, each community was sampled on each geological unit on which it occurred, and locations spanned all large and most small patches of the community. For example, I determined that randomly selecting 20 sites of IESR usually omitted at least one of the large patches and one of the main geologies, and sites were often close together. The constraints I imposed ensured the

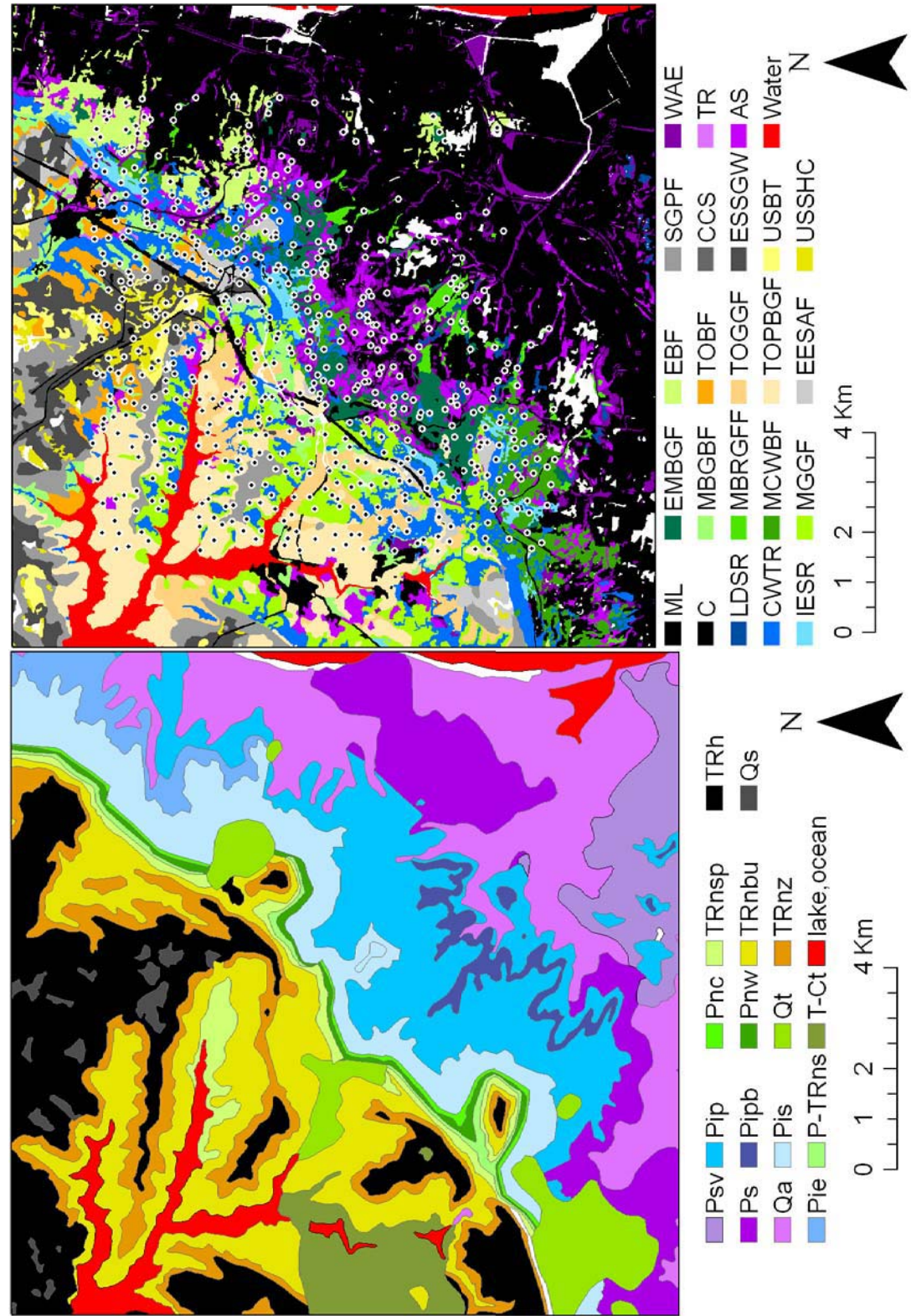
full range of conditions was sampled, which is more beneficial for modelling than ensuring the sample is random (Hirzel and Guisan 2002).

A GPS was used to locate selected sites. No bias was introduced by ensuring sites were qualitatively homogenous or pristine, and many sites were disturbed or contained transition zones between communities. The disturbed communities (WAE, AS, TR) were found to contain many of the target native species (each contained an average of 4–8) — more than USSHC, USBT, ESSGW, CCS, and SGPF. The rainforest and moist eucalypt forests had more target species, as 20 of the 37 species were rainforest species. The disturbed communities usually contained similar species to adjacent ‘natural’ communities – albeit in lower abundance. Therefore, the effect of disturbance is somewhat compensated for when using presence-absence instead of abundance. All sites were accessed where practical, however some dense thickets of prickly weeds, particularly *Caesalpinia decapetala* (cat’s claw, wait-a-while) were avoided. Where access could not be achieved from another direction, or a site was found to be managed rather than semi-natural, the site was replaced with the next suitable one from the random list (as above).

Overall, the survey contains no bias towards or away from roads (to ensure easy access or avoid edge effects), no subjectivity introduced by only surveying ‘homogeneous’ or ‘pristine’ sites, and all reasonable attempts were made to ensure a representative and complete range of vegetation and environmental conditions were sampled. While some sites were close together, this only occurred where the sites contained different vegetation communities, and usually on different geologies as well. The average walking distance between sites was approximately 300 m, and many transitions in vegetation were typically seen over this distance. This can be seen from the relatively small size of vegetation patches relative to the distance between

sites (Figure 4.8). Therefore, the spatial auto-correlation in the survey has been kept to a minimum. There are relatively few sites on the coastal plain, and these are typically near creeks or in hilly areas, but this bias reflects land clearing preferences rather than a bias in the survey design itself. As the vast majority of sites are on the escarpment slopes and Woronora plateau, the locations on the coastal plain do not affect the models for most species.

Figure 4.8 (next page): The geology (left) and vegetation communities (right) present in the study area. Geology map is provided by the NSW Department of Primary Industries. Vegetation map is from NPWS (2002). Black and white spots in the vegetation diagram illustrate the locations where the vegetation survey was conducted. See Section 4.5.1 for definitions and descriptions of each geology and community.



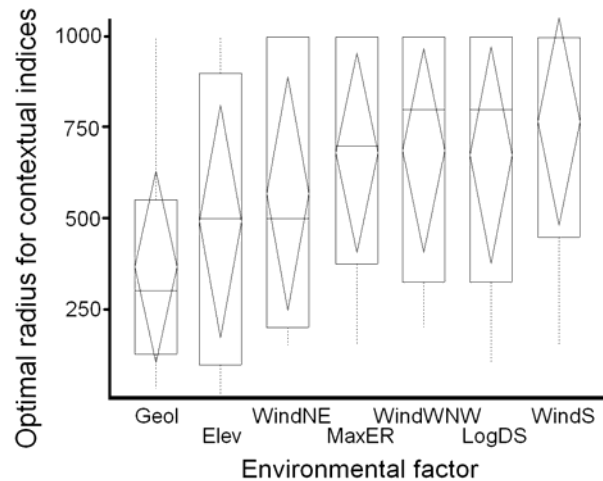


Figure 4.9: The optimal radius of the contextual indices for seven predictors, as determined using eight canopy tree species in single predictor models. The diamonds show the mean and standard deviation. The boxes show the median and quartiles. The seven predictors were geology (Geol), elevation (Elev), the average summer maximum temperature (MaxER), the log of distance to streams (LogDS), and exposure to northeasterly, southerly and west to northwesterly winds (WindNE, WindS, WindWNW).

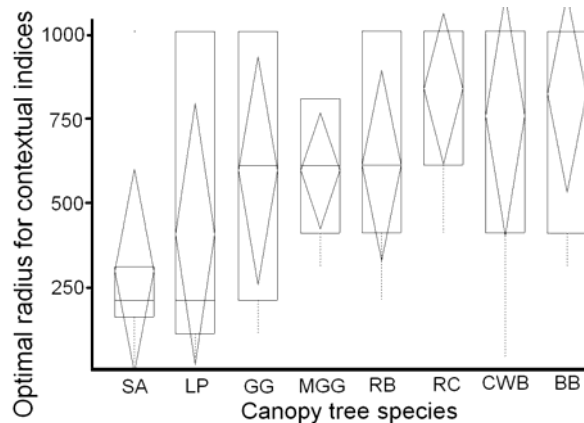


Figure 4.10: The optimal radius of the contextual indices for eight tree species in the Illawarra region, as determined using seven environmental factors in single predictor models. The diamonds show the mean and standard deviation. The boxes show the median and quartiles. The eight species were gully gum (GG), mountain grey gum (MGG), coastal white box (CWB), blackbutt (BB), silvertop ash (SA), red bloodwood (RB), red cedar (RC) and lilly pilly (LP).

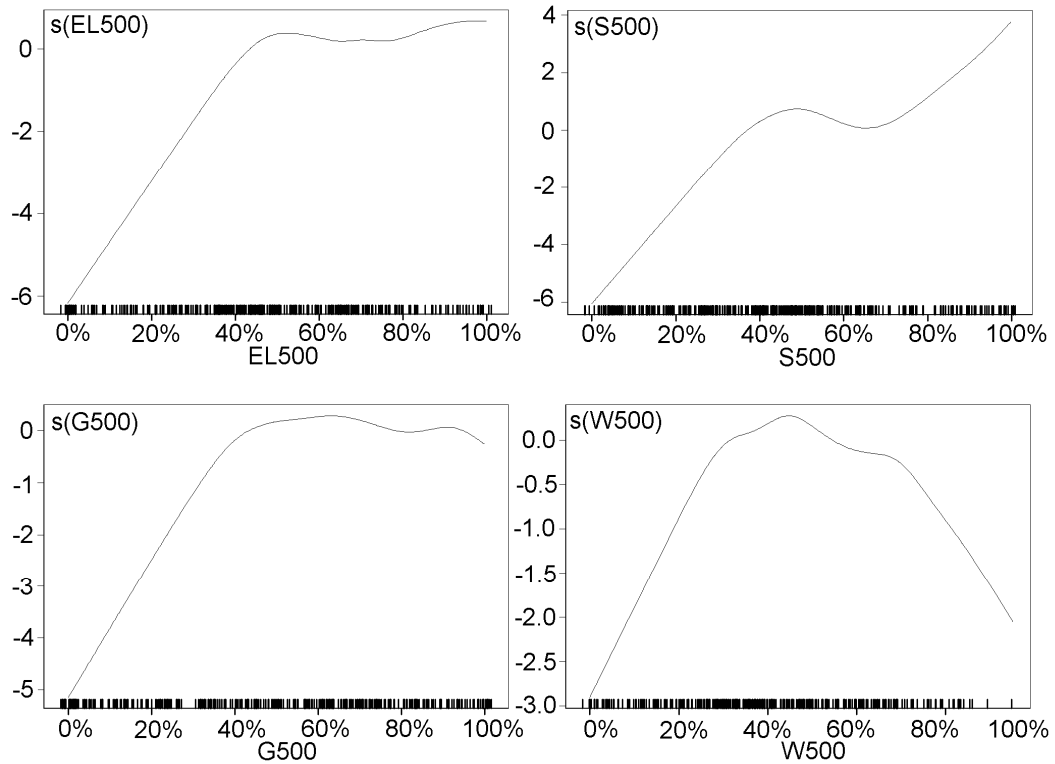


Figure 4.11: The response curves generated by a four-predictor GAM for *E. cypellocharpa* using contextual indices. Each predictor represents the average habitat quality within 500 m, where 100% habitat quality is defined as either: Elevation between 320 m and 420 m (EL500); Exposure to southerly winds of between 0.8 and 1.4 (S500); Geology of either 'TRnbu' or 'TRnz' (G500); or, Exposure to west to northwesterly winds of between 0.5 and 1.0 (W500).

Chapter 5: Climate change at the landscape scale: predicting fine-grained spatial heterogeneity in warming and potential refugia for vegetation³

5.1 Introduction

Current predictive modelling of the response of species to climate change produces highly variable results according to variations in the Global Climate Model (GCM, Beaumont et al. 2007) and statistical technique used (Araújo et al. 2005a). These differences raise questions regarding the usefulness of the models (Araújo et al. 2005a), and highlight a need to improve methodologies.

Much discussion on methodology has concerned the choice between mechanistic process-based models of plant demographics and bioclimatic models based on the correlation between climatic factors and the current distribution of species. Both approaches have advantages and disadvantages (Loehle and LeBlanc 1996; Pearson and Dawson 2003, 2004), yet there are a number of problems common to both that are associated with the climate data used. First, climatic data is typically either from coarse-grained (e.g. 50 km) GCMs, or from an idealised scenario of a fixed and uniform increase in temperature (see Beaumont et al. 2007 for a comprehensive list of examples). Both cases are unable to distinguish fine-scale heterogeneity in climate change, and this may introduce a bias in predictions (Loehle and LeBlanc 1996; Araújo and Rahbek 2006). Temperature increases will vary across microclimates (Beaumont and Hughes 2002), and species respond to spatially heterogeneous regional climates

³ This chapter has published as: Ashcroft MB, Chisholm LA, French KO (2009) Climate change at the landscape scale: predicting fine-grained spatial heterogeneity in warming and potential refugia for vegetation. *Global Change Biology*, 15, 656–667.

rather than global averages (Walther et al. 2002). Local scale effects of climate change have been poorly explored (del Barrio et al. 2006), and further investigation is needed to identify refugia from apparently unfavourable conditions at coarser scales (Pearson 2006). These refugia are a known problem with coarse-scale models, and may mean that predictions of extinctions are exaggerated (Thuiller et al. 2005; Anciães and Peterson 2006; Pearson 2006; Botkin et al. 2007).

The second issue is that seasonal temperatures are often reduced to a limited and predefined set of variables prior to modelling species distributions—most commonly mean annual temperature, winter minimum, and either summer maximum or an estimate of growing degree-days (e.g. Araújo et al. 2005a; Thuiller et al. 2005; Anciães and Peterson 2006; Beaumont et al. 2007). While these variables have been selected based on the general physiological response of species, this *a priori* selection of predictors will lead to erroneous predictions for species that are limited by temperatures during other seasons. Some seasons have warmed more than others, and minimum temperatures have increased by more than maximum and average temperatures (Loehle and LeBlanc 1996). Seasonal fine-tuning of climatic variables has been shown to improve bird models (Heikkinen et al. 2006), and this may be true for plants as well.

The third issue concerns the accuracy of temperature predictions. Errors in temperature estimates are usually stated as being less than 5% (e.g. Beaumont and Hughes 2002), but the absolute magnitude of these errors can still be in the order of 1 °C when temperatures are approximately 20 °C. In addition, there can be variations of up to 33.8 °C within one 50 km cell (mean 1.8 °C, Hijmans et al. 2005), and errors of this magnitude are significant for many species. For example, Hughes et al. (1996) found that 41% of eucalypts in Australia had a mean annual temperature range of less than 2 °C, and 25% less than 1 °C.

Errors in temperature estimates are partially due to the assumption that temperature can be interpolated based only on elevation. While elevation is the dominant factor controlling the distribution of most seasonal temperatures, coastal influences and exposure to wind are more important in some seasons—especially the extreme temperatures that are limiting for many species (Ashcroft et al. 2008; Chapter 3). Incorporating these other factors into temperature predictions changes the spatial distribution of seasonal temperature estimates and dramatically affects model predictions.

This chapter addressed these issues with the temperature data used in climate change models by estimating the fine-scale spatial heterogeneity in warming that has occurred between 1972 and 2007 in an approximately 10 km by 10 km area. A better understanding of these past changes in temperature could assist scientists to improve predictions of future changes. I used the estimates of past warming to investigate whether some locations were potential refugia because they had warmed by less than others. The reduced warming in refugia could act to prevent extinctions, or at least slow the rate at which climate change affects species. I estimated the average amount of warming that 37 plant species had experienced to determine if any species was at more risk because of the bias in the topographic and geographic locations they occupied.

5.2 Materials and methods

5.2.1 Overview of approach

The approach I adopted combined 35 years of data from Bureau of Meteorology (BoM) weather stations with one year of personal observations of soil-surface temperatures. BoM weather stations provided a good record of historical climate change, but were not sufficient to make fine-grained predictions of warming as there were not enough

stations, they were too sparsely distributed, and they did not cover the full range of microclimates in the study area. By recording soil temperatures at 40 sites for one year, and determining the relationship between the air temperatures at the BoM stations and the soil temperatures at each site, I could estimate the fine-grained spatial distribution of temperatures for the whole period of interest (1972–2007). This was based on the assumption that the soil-air temperature relationships were temporally stable. The relationships between BoM air temperatures and site soil temperatures were determined separately for each weather pattern (e.g. wind speed and direction, humidity) as these factors can affect the spatial distribution of soil temperatures (Ashcroft et al. 2008; Chapter 3).

5.2.2 Temporal changes in weather patterns

The study was conducted on approximately 12000 ha of the Illawarra Escarpment and Woronora Plateau (34.4 °S, 150.9 °E), approximately 80 km south of Sydney, Australia (Figure 5.1). The study area was selected because it contains a complex mosaic of vegetation (NPWS 2002), and the patterns cannot be easily explained using common predictors such as elevation and geology. I suspected that elevation may have been a poor surrogate for temperature in this area, and have subsequently shown that this is the case for the extreme temperatures (winter minimums, summer maximums) that have a strong influence on the distribution of species (Ashcroft et al. 2008; Chapter 3).

Long-term weather data from the Bureau of Meteorology (www.bom.gov.au) was only available for one weather station within the study area (Wollongong University). Therefore, I also obtained data from the five nearest long-term stations in a variety of directions (Figure 5.1). Data for the period of March 1972 to February 2007 was obtained for all six stations, although the Nowra and Point Perpendicular data was actually a combination of two stations for different periods.

Hierarchical cluster analysis (using JMP, Ward method) was performed to group days with similar weather patterns. There were 48 variables used, eight for each of the six weather stations. The eight variables were northerly wind component at 9am and 3pm, easterly wind component at 9am and 3pm, humidity at 9am and 3pm, and the relative minimum and maximum temperatures. The northerly and easterly wind components were calculated as $\text{wind speed} \times \sin/\cos(\text{wind direction})$. The relative minimum and maximum temperatures were the difference between each station and the average of all stations. The relative temperatures were used so that uniform spatial warming within and between years would not cause the weather pattern to change. Clusters would only be affected if there were changes in wind speed, wind direction or humidity, or some stations had larger temporal differences in temperature than others. A small number of observations were missing from each weather station, but these could be estimated using linear regression and the data from other stations. Any errors introduced by this process are considered negligible due to the small amount of missing data relative to the 48 variables considered for each day ($8 \text{ variables} \times 6 \text{ stations}$), and the high correlations that I observed between data at different stations.

Cluster analysis was used to group all individual days from March 1972 to February 2007 into one of eight groups—each representing a different weather pattern. The number of groups was selected as a trade-off between more groups (less data in each group to establish relationships between air and soil temperatures) and less groups (more variable weather conditions within each group). I could just have validly used more or less than 8 groups, and the implications of this trade-off are included in the discussion.

The number of days of each weather pattern in each calendar year (1973–2006) was regressed against years to determine if there was an increase or decrease in the

frequency of different weather patterns. I also calculated the percentage of days in each calendar month that belonged to each weather pattern to determine if there was a seasonal trend. I assessed selected statistics for each weather pattern by calculating the mean (and standard deviation) of daily variables using all days in all years that were part of that pattern. Daily variables were northerly and easterly wind component at 3pm and 9am, humidity at 3pm and 9am, and minimum and maximum temperatures. All variables were the average of the respective values at the six weather stations. Note that relative temperatures were used to produce clusters (see above), but weather patterns were assessed using actual temperatures.

5.2.3 Relationships between air and soil temperatures

The weather station data used above was only available at 6 locations, and was therefore insufficient to determine the fine-grained spatial distribution of warming. To counter this problem, I obtained soil temperatures at 40 locations (Figure 5.1) for the period September 2005 to August 2006, and determined the relationship between soil and air temperatures for this period. These relationships were used to predict minimum and maximum soil temperatures at all 40 sites for each day that BoM data was available (1972-2007, as above).

Soil temperatures were recorded using DS1921G iButton temperature loggers (Dallas Semiconductor/MAXIM), which were placed 1 cm below the surface and recorded hourly temperatures (Ashcroft et al. 2008; Chapter 3). Linear regression was used to relate the daily minimum and maximum air temperatures (the average of the six BoM stations) with the respective minimum and maximum soil temperatures from the iButtons for the period September 2005 to August 2006. I used the average of six stations, rather than simply the one station that was recorded within the study area, because I wanted to relate the soil temperatures to the average air temperature and

weather pattern of the region. Individual weather stations vary in factors such as distance to coast and exposure to key wind directions, and therefore have greater potential to be biased. Using the regional temperature and weather pattern means the method has greater potential to downscale coarse-grained GCM data.

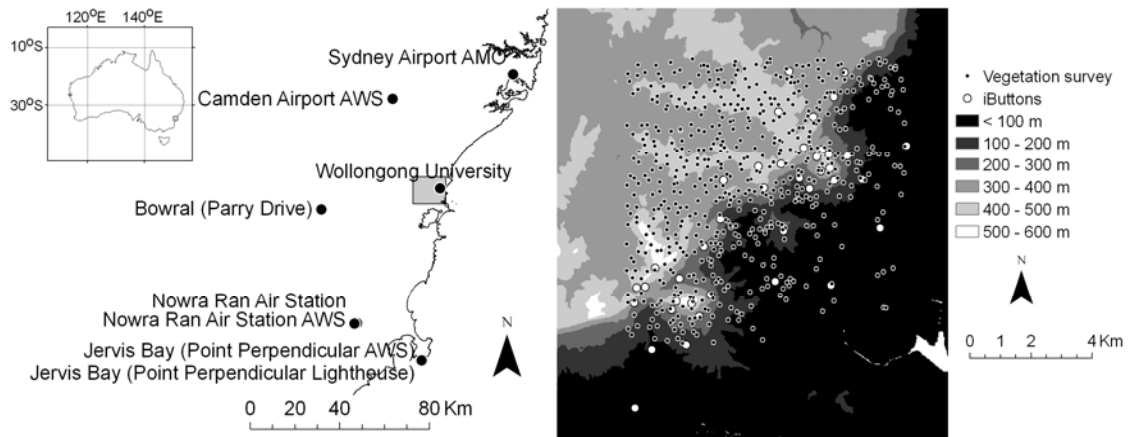


Figure 5.1: The topography of the study area (right) as an elevation between 0 and 573 m. Superimposed are 40 locations where temperatures were recorded with iButtons, and 600 sites where a vegetation survey was conducted. Daily Bureau of Meteorology observations were obtained from six nearby locations (left).

Analysis was performed separately for each of the eight weather patterns and 40 sites (320 regressions for maximum and minimum temperature), as air temperature may have a different effect on soil temperature at different sites according to wind direction or humidity. In addition, different sites are affected differently according to the exposure to those wind directions (Ashcroft et al. 2008; Chapter 3). The established relationships between air and soil temperatures were used to estimate the soil temperatures at each of the 40 iButton sites for each day from March 1972 to February 2007.

I investigated a number of factors that had the potential to affect the relationships between soil and air temperatures. First, the average slopes of the air-soil relationships

for the eight weather patterns were regressed against their average humidity, average minimum and maximum temperatures, and average northerly and easterly wind components. Second, the average slopes of the air-soil relationships for the 40 sites were regressed against their elevation, distance to coast, distance to streams, and exposure to different directions. The ‘exposure’ predictors were topographically derived estimates of exposure to wind based on the angle to the horizon in a given direction (see Ashcroft et al. 2008; Chapter 3 for the accuracy and source of all predictors). Exposure has a number of advantages over aspect, and has been shown to be important for the spatial and temporal variations in temperature (Ashcroft et al. 2008; Chapter 3).

5.2.4 Estimating spatial variations in temperature change from 1972 to 2007

The average seasonal minimum and maximum temperatures for each site in each of the 35 years (from March 1972–February 1973 to March 2006–February 2007) were calculated by averaging the respective daily temperatures. Seasons conformed to those of the southern hemisphere. That is, summer (December–February), autumn (March–May), winter (June–August) and spring (September–November). For each site, the 35 years of seasonal average minimum and maximum temperatures were regressed against years to determine the trend in temperatures. The amount of warming was estimated as the slope of the regression $\times 34$ (the difference in years between the start and end dates). The four respective seasonal estimates of warming at each site were averaged to estimate the annual warming in minimum and maximum temperatures. In total, there were 10 estimates of warming for each of the 40 sites—minimum and maximum temperatures for four seasons plus an annual average. The averages across the 40 sites were used to estimate the amount of warming in each season that had occurred between 1972/3 and 2006/2007.

The factors influencing the spatial distribution of warming were examined by regressing the warming at the 40 sites (10 separate regressions—one for each of the seasonal temperatures above) against elevation, distance from coast, distance from streams and exposure. These predictors are known to influence the distribution of temperatures in this study area (Ashcroft et al. 2008; Chapter 3). The selected direction of the exposure predictor was the one that maximised the r^2 of the regression.

The relative contribution that each environmental factor made to the amount of warming was compared with other factors by multiplying the coefficient of each predictor with its range. This estimated its overall ‘influence’ in degrees Celsius, and allowed predictors with different units and ranges to be directly compared (Ashcroft et al. 2008; Chapter 3).

5.2.5 Estimating the impact of climate change on plants

The established relationships, between the amount of warming at a site and the respective environmental variables, were used to generate ‘warming surfaces’ for the whole landscape in a GIS. These warming surfaces were then used to estimate the amount of warming at each of 600 sites where a vegetation survey was conducted (Figure 5.1).

The study area contains a complex mosaic of moist and dry rainforests, moist eucalypt forests, tall open eucalypt forests, upland swamps and woodlands (NPWS 2002). Each site (20 m by 20 m) was surveyed for presence or absence of 37 species that were common in these communities (NPWS 2002), of which the majority were trees (Table 5.1). Previous results have shown that the distributions of these species are explained well by models with landscape scale soil temperatures (Ashcroft et al. 2008; Chapter 3).

For each species, I averaged the amount of warming in the 10 seasonal temperature gradients using only the sites where that species was found. This produced 10 estimates of warming for each species—each representing the average amount of warming for that species in terms of that seasonal temperature. I determined the potential bias in coarse-grained climate models by determining the difference in warming that different species experienced on each gradient. If there were no bias in coarse grained models, then all species should experience the same amount of warming. Differences in the amount of warming could occur if species were biased in the topographic and geographic locations they occupied within the landscape.

For each of the 10 seasonal temperatures, I ranked all species using a linear scale from 0% (experienced the least amount of warming) to 100% (experienced the most amount of warming). I calculated the mean and standard deviation for each species across the 10 seasonal temperatures to determine if they consistently experienced a similar amount of warming relative to the other species. This was used to determine whether the bias in warming for each species was predictor specific, or whether the bias was consistent across all temperature predictors.

Table 5.1: The 37 species from the Illawarra Escarpment that were modelled as part of this study.

Abbreviation	Common name	Scientific name
TVH	Two-veined hickory	<i>Acacia binervata</i>
GW	Green wattle	<i>A. mearnsii</i>
LP [†]	Lilly pilly	<i>Acmena smithii</i>
BS	Black she-oak	<i>Allocasuarina littoralis</i>
ROP [†]	Red olive plum	<i>Cassine australis</i>
CW [†]	Coachwood	<i>Ceratopetalum apetalum</i>
HC [†]	Hairy clerodendrum	<i>Clerodendrum tomentosum</i>
RB	Red bloodwood	<i>Corymbia gummifera</i>
NC [†]	Native cascarilla	<i>Croton verreauxii</i>
JW [†]	Jackwood	<i>Cryptocarya glaucescens</i>
MG [†]	Murrogun	<i>C. microneura</i>
PT [†]	Prickly tree fern	<i>Cyathea leichhardtiana</i>
GST [†]	Giant stinging tree	<i>Dendrocnide excelsa</i>
SF [†]	Sassafras	<i>Doryphora sassafras</i>
MGG	Mountain grey gum	<i>Eucalyptus cypellocarpa</i>
YS	Yellow stringybark	<i>E. muellerana</i>
BB	Blackbutt	<i>E. pilularis</i>
SPM	Sydney peppermint	<i>E. piperita</i>
CWB	Coast white box	<i>E. quadrangulata</i>
SG	Scribbly gum	<i>E. racemosa</i> X <i>haemastoma</i>
BGH	Blue gum hybrid	<i>E. saligna</i> X <i>botryoides</i>
SA	Silvertop ash	<i>E. sieberi</i>
GG	Gully gum	<i>E. smithii</i>
FRG	Forest red gum	<i>E. tereticornis</i>
BWR [†]	Bolwarra	<i>Eupomatia laurina</i>
CSF [†]	Creek sandpaper fig	<i>Ficus coronata</i>
LT [†]	Lantana	<i>Lantana camara</i>
CTP [†]	Cabbage tree palm	<i>Livistona australis</i>
HLD [†]	Hairy-leaved doughwood	<i>Melicope micrococca</i>
VMO [†]	Veined mock-olive	<i>Notelaea venosa</i>
SP [†]	Sweet pittosporum	<i>Pittosporum undulatum</i>
FW [†]	Featherwood	<i>Polyosma cunninghamii</i>
WB [†]	Whalebone tree	<i>Streblus brunonianus</i>
TT	Turpentine	<i>Syncarpia glomulifera</i>
SR [†]	Scentless rosewood	<i>Synoum glandulosum</i>
BP [†]	Brush pepperwood	<i>Tasmannia insipida</i>
RC [†]	Red cedar	<i>Toona ciliata</i>

[†] introduced species

[‡] rainforest species

5.3 Results

5.3.1 Temporal changes in weather patterns

The eight weather patterns that were identified using cluster analysis each favoured different seasons during the year. Winter was dominated by group 7 conditions, which were characterised by low temperatures, intermediate humidity, and light west to northwest winds (Figure 5.2). Group 7 days decreased in frequency from approximately 64 to 46 days per year over the period of interest. Group 8 and group 4 were also common in early and late winter respectively. Group 4 days were characterised by strong westerly winds, low temperatures, and low humidity. They increased from approximately 33 to 38 days per year. Group 8 days were characterised by light NE to NW winds, low temperatures, but relatively high humidity. They decreased in frequency from 35 to 18 days per year, the largest percentage drop of any of the weather patterns (Figure 5.2f). All of the three groups that were common in winter had similar average temperatures, but the higher humidity groups (7, 8) decreased in frequency, and were replaced by group 4 (lower humidity, stronger westerly winds) and an increasing frequency of spring conditions (see group 2 below). This suggested a change towards drier and/or shorter winters.

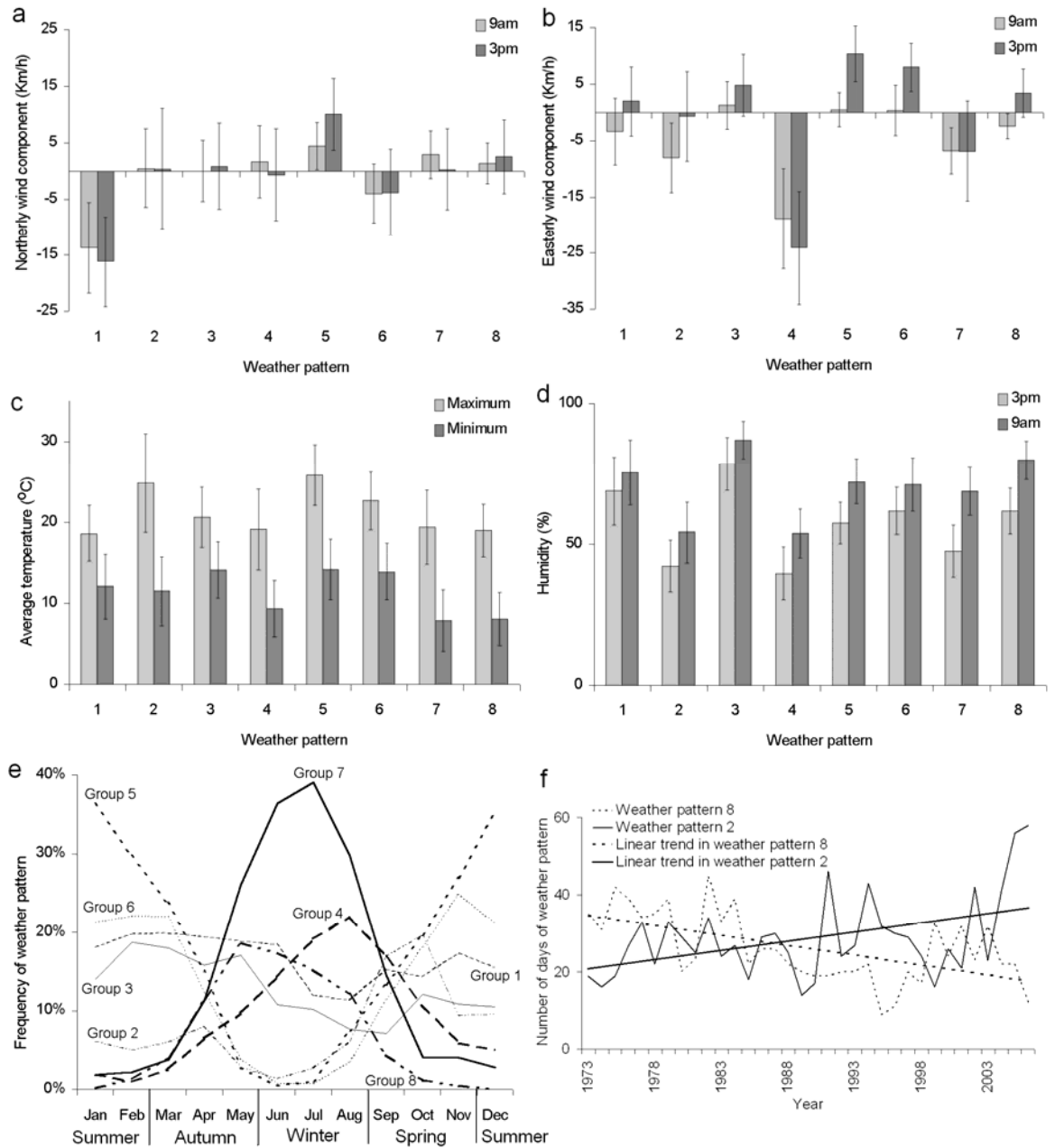


Figure 5.2: Eight weather patterns were identified for the study area. Each weather pattern is represented as the average northerly (a) and easterly (b) components of the wind across the six BoM weather stations (wind speed \times sine/cosine of wind direction), average minimum and maximum temperatures (c), the average humidity at 9am and 3pm (d). Panel (e) illustrates the seasonal change in weather patterns using the percentage of days in each month that belonged to each of the eight weather patterns (average between 1973 and 2006). Panel (f) shows the inter-annual change in the annual number of days in weather patterns 2 and 8 over that period.

The most common conditions in summer were group 5 and group 6. Both were characterised by high temperatures and moderate humidity, but group 5 had northeast winds, while group 6 had southeast winds. Both increased in frequency, with group 5 increasing from 60 to 67 days per year, and group 6 from 48 to 50. This suggested an increase in the frequency and/or length of summer conditions.

Spring and autumn were transition periods where both winter and summer conditions were observed. However, spring also contained the peak of group 2 days. Group 2 was characterised by high temperatures and low humidity under the influence of westerly winds. These conditions are desiccating for moist rainforest plants, and pose bushfire hazards (Fuller 1995). They increased in frequency from 20 to 36 days per year, which was the biggest increase in frequency of all the weather patterns (Figure 5.2f).

The final two groups (1 and 3) were observed over the whole year, but were more common in summer than winter. Both had moderate temperatures, low diurnal ranges, and high humidity, suggesting they occurred during rainy periods. Group 1 was characterised by strong southerly winds and increased from 55 to 66 days per year. Group 3 was characterised by light easterly winds, and decreased from 50 to 44 days per year.

5.3.2 The relationships between soil and air temperatures

The 640 correlations (40 sites, 8 weather patterns, minimum and maximum temperatures) between soil and air temperatures were strong (mean $r^2 = 0.83$, s.d. = 0.10), however the slope of the regressions varied dramatically from 0.30 to 1.35 (mean = 0.71, s.d. = 0.15). This illustrates that, on average, a 1 °C increase in average air temperatures across the six weather stations corresponded with a 0.71 °C rise in soil temperatures on the Illawarra Escarpment, but there were noticeable variations.

There were large differences in the average slope of the air-soil temperature relationships between the eight different weather patterns (mean = 0.71, s.d. = 0.12, n = 8), which were best explained in terms of humidity ($r^2 = 0.54$, $P < 0.05$, Figure 5.3). The weather patterns with low average humidity (groups 2, 4, 7) were less sensitive to changes in air temperature than those with high humidity.

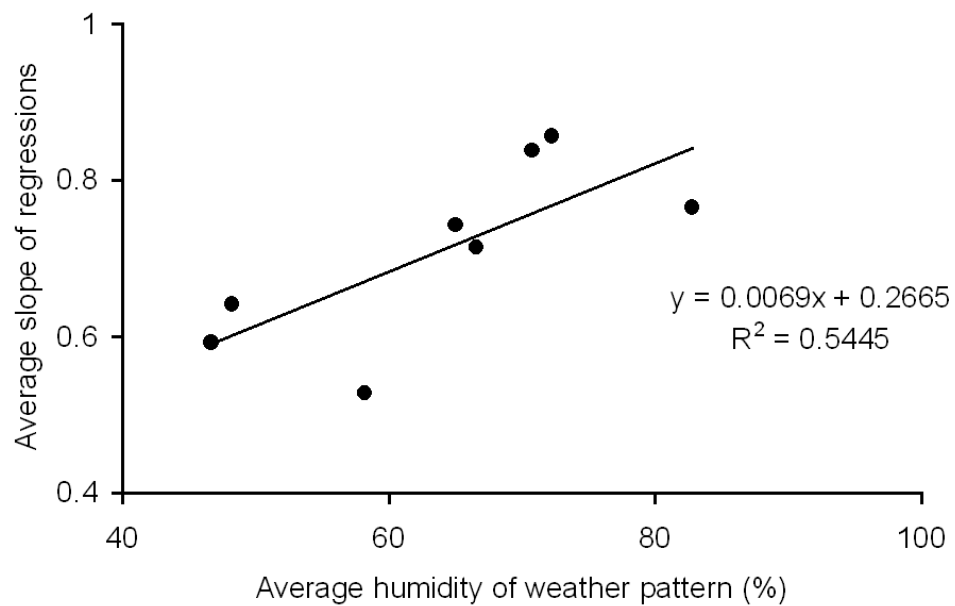


Figure 5.3: The relationship between average slope of the air-soil temperature regressions and average humidity, where each point corresponds with one of the eight weather patterns used in this study.

There were smaller differences in average slope of the air-soil relationships between the 40 different sites (mean = 0.71, s.d. = 0.07, n = 40), but these were significantly correlated with exposure to the WNW ($r^2 = 0.14$, $P < 0.05$). Exposed sites had higher regression slopes, indicating that a 1 °C change in air temperature had more affect on ‘west facing’ slopes than ‘east facing’ slopes.

5.3.3 Estimating the spatial distribution of warming from 1972 to 2006

The amount of warming (averaged over the 40 sites) that was estimated to have occurred over the last 34 years was higher in winter and spring than summer and autumn, and was higher for minimum temperatures than maximum temperatures (Figure 5.4). The 10 estimates of warming (minimum and maximum temperatures for the four seasons and annual period) at the 40 sites were explained using multiple regressions against environmental factors (mean $r^2 = 0.39$, s.d. = 0.07).

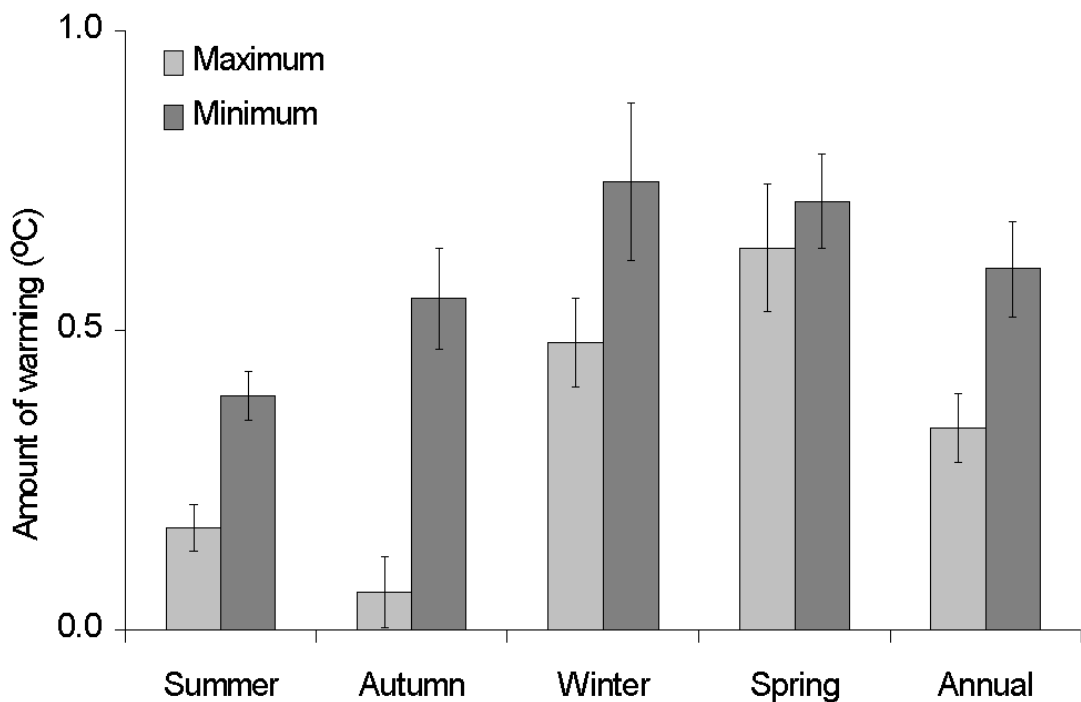


Figure 5.4: The mean (and standard deviation) amount of warming (1972-2007) estimated at 40 sites where soil temperatures were recorded.

Distance from coast was, on average, the most influential predictor of warming, with inland sites warming by more than coastal sites in all ten models. The magnitude of the effect varied from 0.06 °C to 0.30 °C, with highest influence on winter minimums. Distance from streams was the least influential predictor, but there was more warming away from streams in all ten models. The effect ranged from 0.01 °C to 0.11 °C. Elevation was the most influential predictor of warming in minimum temperatures, but was less influential than exposure and distance to coast for maximum temperatures. All ten models suggested there was more warming at lower elevations, and the difference ranged from 0.02 °C to 0.37 °C. Exposure was the most influential predictor for maximum temperatures, but was less significant than elevation and distance from coast for minimum temperatures. The effect ranged from 0.05 °C to 0.20 °C. Annual minimums, and autumn and winter temperatures were best explained using an exposure direction of S to SE, with exposed sites warming by less than sheltered sites. Annual maximums and spring and summer temperatures were best explained using exposure to the W to NW or N to NE, with exposed sites warming by more than sheltered sites. These directions were consistent with more warming at sites exposed to the warm-dry NW winds, and less warming at sites exposed to cold SE sea breezes (Fuller 1995; Ashcroft et al. 2008; Chapter 3).

The different contributions of environmental factors in different seasons meant that the surfaces for warming displayed different spatial patterns (Figure 5.5). For example, summer maximums only displayed a small amount of warming (< 0.35°C), and were heavily influenced by distance to coast. In contrast, spring maximums displayed a high level of warming (0.4–1.0°C), with exposure to the WNW the dominant factor. This was consistent with the increasing frequency of the group 2 weather pattern (westerly winds

in spring) noted above. Finally, winter minimums had the most warming (0.4–1.4°C), and were influenced by both distance to coast and elevation.

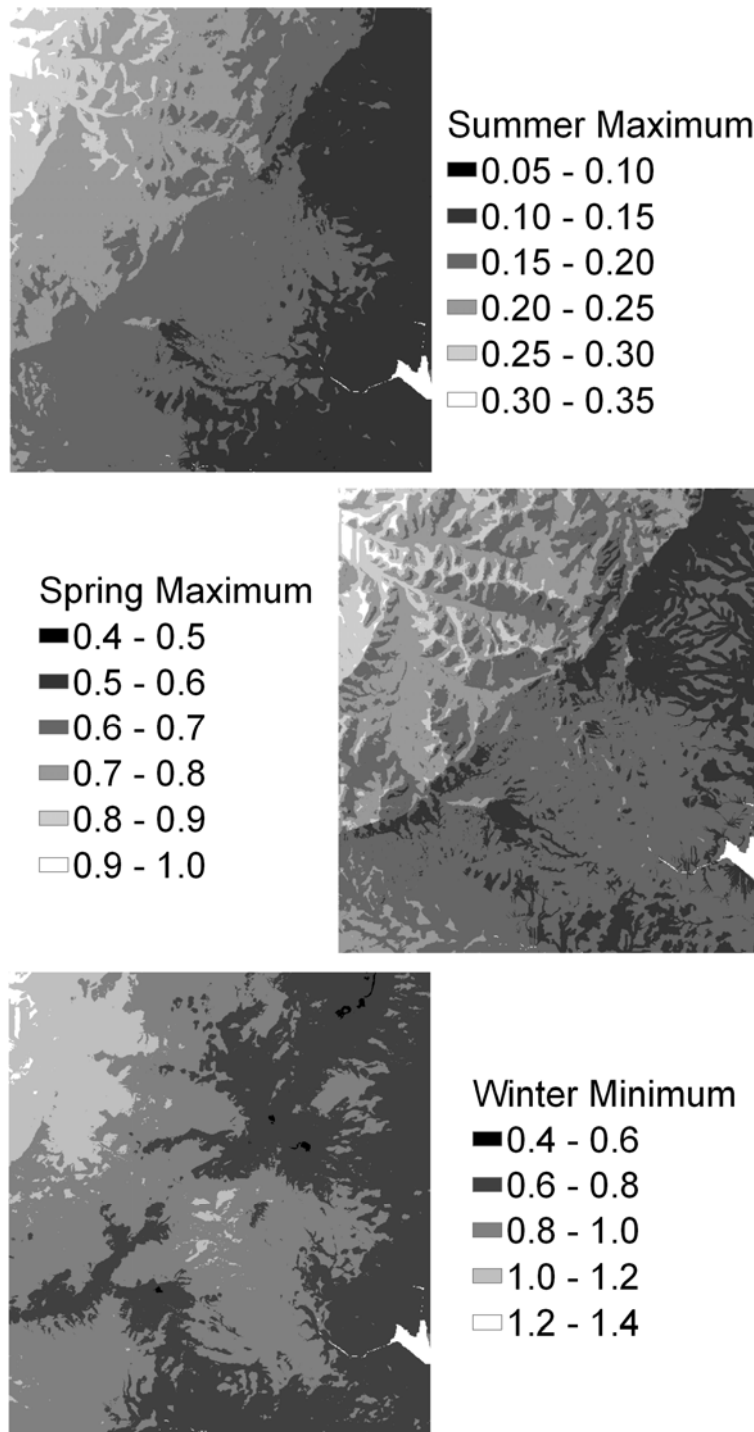


Figure 5.5: The spatial distribution of warming (1972-2007) in the study area (Figure 5.1) as estimated by regressing the amount of warming at the 40 sites where temperatures were recorded against environmental factors. All surfaces are in degrees Celsius.

5.3.4 Effect of warming on vegetation

The difference in warming between species on the same temperature gradient varied from 15% to 197% (Table 5.2). For the less extreme temperature gradients (winter maximums and spring and summer minimums), there was less than 16% difference in warming between all 37 species. For the more extreme temperatures (winter minimums and summer, autumn and annual maximums), there was more than 30% difference between species (Table 5.2).

Table 5.2: The amount of warming in ten seasonal temperature variables was estimated at 600 sites where a vegetation survey was conducted. I calculated the average amount of warming for each of 37 species by averaging the warming at the sites at which those species were observed. The middle columns represent the mean warming for those species with the lowest and highest averages respectively. The final column indicates the percentage difference between these two values.

Seasonal temperature	Lowest warming (°C)	Highest warming (°C)	Percentage difference
Summer maximum	0.15	0.23	47%
Autumn maximum	0.05	0.16	197%
Winter maximum	0.47	0.55	16%
Spring maximum	0.60	0.74	24%
Annual maximum	0.32	0.42	31%
Summer minimum	0.38	0.44	15%
Autumn minimum	0.54	0.69	28%
Winter minimum	0.74	0.96	30%
Spring minimum	0.71	0.82	15%
Annual minimum	0.59	0.73	23%

Most species had a similar amount of warming relative to other species over all ten temperature gradients (Figure 5.6). Species that were only found at inland sites, and typically on drier slopes exposed to the west and northwest, experienced a relatively high amount of warming on all ten gradients. These species included *Eucalyptus*

cypellocarpa (MGG), *E. piperita* (SPM), *Acacia binervata* (TVH) and *A. mearnsii* (GW). In contrast, species that experienced a relatively low amount of warming on all ten gradients included moist and dry rainforest species (e.g. *Acmena smithii* (LP), *Doryphora sassafras* (SF), *Toona ciliata* (RC) and *Cassine australis* (ROP)) as well as species that were predominately restricted to the sheltered slopes of the escarpment (e.g. *Syncarpia glomulifera* (TT) and *Eucalyptus pilularis* (BB)). The species that did vary in relative warming (high standard deviation in Figure 5.6) were typically those that were common on the Hawkesbury sandstone peaks (e.g. *Eucalyptus sieberi* (SA) and *Corymbia gummifera* (RB)) and dry rainforest species from the foothills and coastal plain (e.g. *Croton verreauxii* (NC) and *Melicope micrococca* (HLD)). The former two species (SA and RB) were below the 30th percentile in terms of relative warming for winter minimum, but above the 70th percentile in terms of summer maximum. The latter two species (NC and HLD) were below the 16th percentile in terms of summer maximum, but above the 80th percentile in terms of winter minimum.

5.4 Discussion

5.4.1 The importance of weather patterns for climate change predictions

The results of this study provide further evidence that the spatial distributions of landscape scale soil temperatures are heavily dependent on weather patterns such as the speed and direction of prevailing winds (Ashcroft et al. 2008; Chapter 3). This is an important finding with respect to climate change studies, because any change in weather patterns could dramatically change the spatial distribution of temperatures, and cause large differences in the temperature changes that different locations experience.

Locations where there is less warming could act as refugia, and prevent extinctions that

are typically predicted by coarse-scale models, or at least reduce the rate at which climate change affects different species.

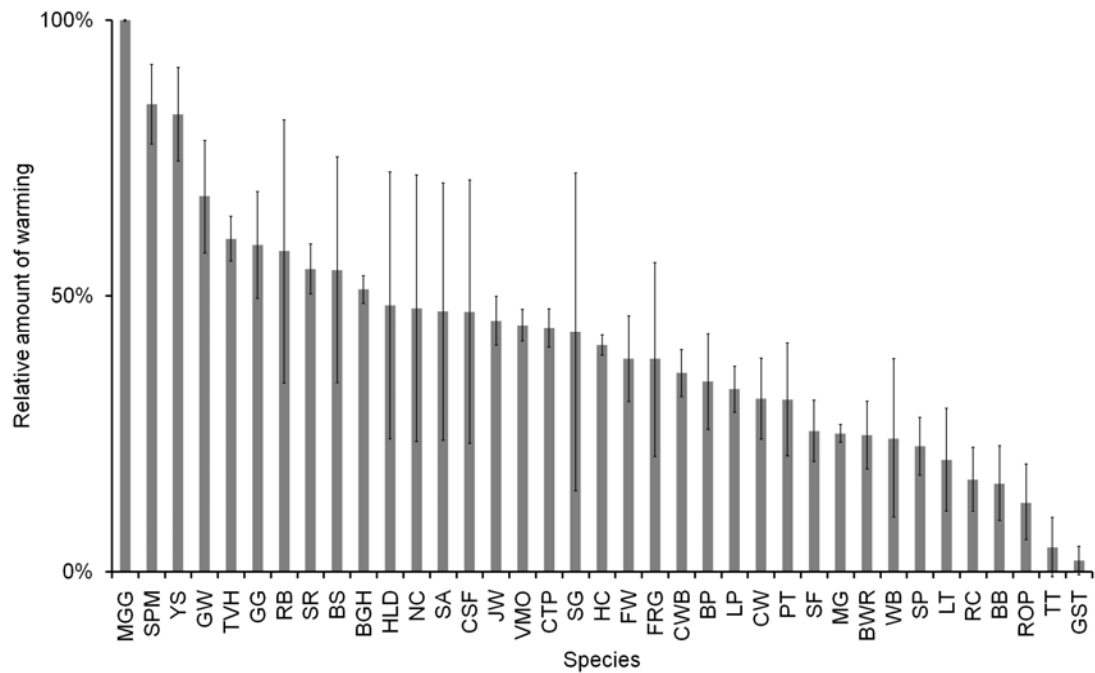


Figure 5.6: The average amount of warming (1972-2006) was estimated for 37 species (Table 5.1) using ten seasonal temperature gradients (Table 5.2). The relative warming for each species on each gradient was calculated on a linear scale from 0% (least warming of all species) to 100% (most warming of all species). This graph illustrates the mean (standard deviation) relative warming for each species over the ten gradients.

Less warming has occurred at sites that are nearer the coast, closer to streams, at higher elevations, exposed to cold S to SE winds, or sheltered from warm, dry, W to NW or N to NNE winds. Species are biased in the topographic and geographic positions they occupy, and therefore different species have experienced different amounts of warming over the last 34 years. For example, moist rainforest species are typically found in locations that are sheltered from the warm, dry W to NW winds. Therefore,

these species have experienced less warming than species such as *Eucalyptus cypellocarpa*, *E. piperita* and *E. saligna* *X* *botryoides*, which are often found at inland sites exposed to these winds. In this context, moist rainforest species are already restricted to refugia within the landscape.

The topographic and geographic biases in species distributions are not unique to this study area. For example, rainforest species are restricted to sheltered gullies at other locations along the east coast of Australia (e.g. Van Niel and Austin 2007), and eucalypts have consistent biases in the topographic positions they favour (e.g. Austin and Meyers 1996). Therefore, although this study has only determined the species-specific bias in warming in one study area, it is possible that these biases result in consistent over or underestimates across the whole range. If this were the case, then it would represent a consistent bias in each cell of coarse-grained SDMs.

The results of this study highlight the need to improve the accuracy of methods that are used to downscale coarse-grained temperature surfaces. Currently, coarse-scale bioclimatic models are downscaled using elevation as a surrogate for temperature (Trivedi et al. 2008), or SDMs consider fine-scale heterogeneity by including the elevational range of each cell as an extra predictor (Luoto and Heikkinen 2008). The accuracy of these approaches could be improved by considering the regional weather pattern (wind speed, direction, humidity) and the topographic exposure to key wind directions. To successfully implement this over large geographic regions, it will be necessary to develop general rules as to how regional weather patterns relate to fine scale temperature distributions. This will require applying methods similar to those in this chapter to numerous other study areas, and determining whether generally applicable relationships can be established.

Further research is also needed to confirm how the fine-grained spatial variability of climate change, and the species-specific biases, affects the results of Species Distribution Models. However, it is already known that climate change predictions vary substantially according to variations in the GCM (Beaumont et al. 2007) and statistical technique (Araújo et al. 2005a), and the results of SDMs vary when fine-grained spatial variability in temperature is considered (Ashcroft et al. 2008; Chapter 3). Therefore, it is likely that new methods to downscale temperature data would dramatically change climate change predictions from SDMs.

5.4.2 Selecting the correct temperature predictors in models

While spatial variations in temperature change caused small differences (up to 0.22 °C, Table 5.2) in the average warming each species experienced on the same temperature gradient, these differences were small when compared to the difference between different temperature gradients (0.68 °C–0.81 °C). For example, minimum temperatures increased almost twice as much as maximum temperatures (except in spring), and winter and spring temperatures increased approximately twice as much as those in summer and autumn. Therefore, an important area for climate change studies is determining which seasonal temperatures are limiting the distribution of each species. Modelling species with the wrong seasonal temperature estimate will dramatically alter estimates of extinction risk.

Determining the seasonal temperatures that are limiting each species is not a simple task. Many seasonal temperatures are highly correlated (especially if they are all derived using only elevation and location), and there may be little difference in model performance using temperature estimates from different seasons. The wrong predictors can easily be selected, and this can drastically alter predictions if they are in different seasons than the true limiting factors. One of the advantages of deriving fine-grained

temperature estimates using distance to coast, distance to streams, exposure to key wind directions and elevation, is that it reduces the correlation between alternative temperature predictors. This reduces the probability that the wrong predictor will be selected, although it does not eliminate it completely when multiple seasons have a similar spatial distribution of temperature.

5.4.3 The relationship between soil and air temperatures

An unexpected and interesting result from this study was that the relationship between soil and air temperatures varied according to the humidity of the weather pattern and the exposure of the site to the WNW. Humidity was low when winds were from the west (groups 2, 4, 7), and a 1 °C difference in air temperature made an average difference of 0.53–0.64 °C to soil temperatures under these conditions. In contrast, for the other weather patterns, a 1 °C difference in air temperature made an average difference of 0.71–0.86 °C to soil temperatures.

The reason that humidity affects the relationship between soil and air temperatures is not clear, but there are at least two possibilities. Firstly, the high specific heat of water may affect the transfer of heat between soil and air. That is, it may be more efficient to transfer heat to the soil when the air is humid and the soil is dry, than when the air is dry and the soil is moist. Secondly, this result may reflect a bias in the locations of the iButtons relative to the broader study area covered by the weather stations. The study area where the iButtons were placed is near the coast, and many of the sites are sheltered from the westerly winds by the escarpment. These possibilities require further investigation.

Understanding the interactions between soil and air temperatures is important, as both may be important for determining the response to climate change. More data is available from the BoM on air temperatures, but soil temperatures are more spatially

heterogeneous and may be better able to explain the patchy nature of vegetation (Ashcroft et al. 2008; Chapter 3). The results of this study suggest that a change in humidity may affect soil temperatures, and therefore the distribution of species, even if there is no change in mean air temperatures.

5.4.4 Assumptions and limitations

The primary assumption of this study is that the relationships between soil and air temperatures are temporally stable. This assumption needs to be tested to ensure that there has been no bias in the relationships introduced by climate change. The primary limitation is that the study is restricted to one 10 km by 10 km study area, and further research is needed to determine if the results are indicative of other areas.

I selected eight weather patterns to conduct my analysis. This ensured that I had sufficient data to determine the air-soil temperature relationships, but meant I only had eight points (Figure 5.3) when determining which factors affected the slope of the regressions. Increasing the period (one year) over which soil temperatures were recorded would provide extra data, and this would allow more weather patterns to be analysed without compromising the strength of relationships between soil and air temperatures.

5.5 Conclusions

Spatial variations in temperature are influenced by the prevailing weather pattern (wind direction, wind speed, humidity), and geographic and topographic factors such as distance to coast and exposure to winds. Climate change has altered the frequency of different weather patterns and this has led to fine-grained spatial differences in the amount of warming. As species are biased in the topographic and geographic locations they occupy, these spatial variations in warming mean that some species are at more

risk of extinction than others, and these differences can not be detected by coarse-grained models. Determining which seasonal temperatures affect each species' distribution and improving the accuracy of temperature distributions will improve the accuracy with which models can predict the response of species to climate change.

Chapter 6: An evaluation of the environmental factors in species distribution models for vegetation on the Illawarra Escarpment

6.1 Introduction

Species Distribution Models (SDMs; Guisan and Zimmermann 2000; Rushton et al. 2004; Guisan and Thuiller 2005) explain or predict the distribution of species using environmental factors such as temperature, moisture availability, topographic position, and geology. These models are developed based only on correlations, yet the relationships are implicitly assumed to be causal when they are used to predict the response of species to climate change (e.g. Bakkenes et al. 2002; Pearson and Dawson 2003; Thomas et al. 2004), applied to other regions (e.g. Randin et al. 2006), or used to estimate the potential spread of invasive species (e.g. Peterson 2003). Correlations alone are insufficient to justify these applications of SDMs, and a paradigm shift is necessary to improve the methods that are used to build and evaluate models (Araújo and Rahbek 2006).

The most common process for producing SDMs involves randomly dividing presence/absence data on the distribution of species into separate calibration and validation data sets (Araújo et al. 2005b). The calibration data set is used to choose which environmental factors are included in the selected model, and determine the response function for each gradient. The selected model is then evaluated by calculating statistics such as the Area Under the receiver operating characteristic Curve (AUC; Swets 1988; Pearce and Ferrier 2000b) on the separate validation data set.

One issue with this model building process is that only the selected model is validated with the independent dataset. Corroboration assessment (Faith 2003; Faith et al. 2004) suggests that this provides weak inference to accept the selected model, as there is no information on how it compares with the alternative models. The alternatives may perform similarly, and be equally valid alternative models (hypotheses) for the distribution of species.

The dangers of accepting any model with a strong correlation is further illustrated by recent results showing that random spatially structured predictors offer statistically strong explanations for species distributions (Bahn and McGill 2007; Currie 2007). As species distributions and environmental factors are often spatially autocorrelated, SDMs are expected to have strong correlations even when there is no causal link between the two. Therefore, demonstrating that the selected model has good explanatory ability is insufficient to justify its use in environmental management.

To address these problems, it is necessary to consider the theoretical basis of model production and hypothesis testing. In many ecological studies, a strong correlation between two variables can be used to reject the null hypothesis that there is no relationship between them. However, model building in a multivariate context is more about weight of evidence than null hypothesis testing (Stephens et al. 2007). There may be many models that produce similar results, especially when predictors are highly correlated, and this makes it difficult to determine the model with the most support. Many commonly used modelling methods can produce erroneous results, or identify spurious correlations, and it may be more ecologically beneficial to provide evidence on causal factors rather than identify the ‘best model’ (MacNally 2002).

MacNally (2002) detailed a hierarchical partitioning approach whereby the effect of an environmental factor was estimated by averaging its effect in all candidate models in

which it occurred. Hierarchical partitioning offers a method to increase confidence in the predictors that are selected in models, however 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. Hierarchical partitioning may help resolve ambiguities due to correlated predictors, but cannot help avoid spurious correlations between a species distribution and an environmental factor.

In this chapter I introduce an alternative method that is designed to examine the strength of evidence that an environmental factor affects species distributions. There are two key differences from other model building processes. First, all candidate models are validated using the independent dataset. This provides better protection against overfitting. For example, if a candidate model is overfitting the training data, then it should have a lower performance on the independent validation data set. If only the selected model is validated, then the process will be capable of detecting overfitting, but it will not prevent an overfitted model being selected in the first place. However, if all candidate models are validated using the independent dataset, then any overfitting is detected during the model building process, and can be used to avoid selecting overfitted models.

The second difference 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 is used in each of the models. Hence, the method avoids the

problem associated with hierarchical partitioning, which is based on multiple models but with the same data in each candidate model.

The method is applied in this chapter to provide a more robust examination of three environmental factors that appear 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; Chapter 3) but have not been comprehensively evaluated in a multivariate context. Here, the three predictors are examined to determine if they make a unique contribution to the performance of SDMs, or whether other predictors can combine to produce alternative, and potentially equally plausible hypotheses for the distribution of species.

Furthermore, the methods are used to test whether it is important to consider spatial predictors in SDMs. The distribution of a species is not just determined by environmental niche factors, with dispersal, source-sink effects, mass-effects, fragmentation and other spatial processes also contributing to observed patterns (Pulliam 2000). In Chapter 4 I demonstrated that contextual indices (Ferrier et al. 2002; Wintle et al. 2005) could capture spatial processes without the problems of autologistic regression, and could improve the performance of SDMs over purely niche-based models. Here, I provide a more comprehensive evaluation to determine whether contextual indices provide information that is not captured by the raw environmental factors.

Finally, I evaluate whether species distributions are better correlated with past or present temperatures. The spatial distribution of warming has not been uniform between 1972 and 2006 (Ashcroft 2009; Chapter 5), and species that have not adjusted their distribution may be better correlated with past temperatures.

6.2 Methods

6.2.1 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; Figure 6.1).

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 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 (see Section 4.5.1 for more details on geology in the study area).

6.2.2 Environmental predictors

A categorical geology layer was obtained courtesy of the NSW Department of Primary Industries and contained spatial errors of up to 150 m in the locations of boundaries (see Section 3.6.1 for more information on predictor accuracy). I 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 (Ashcroft et al. 2008; Chapter 3). The surfaces were developed by recording temperatures using DS1921G iButtons at 40 sites for a period of 21 months from December 2004 to August 2006. The data was originally collected and analysed in three-week periods (Ashcroft et al. 2008; Chapter 3), however all surfaces for each season were later averaged to produce the seasonal temperature surfaces (referred to hereafter as 2005 temperature surfaces).

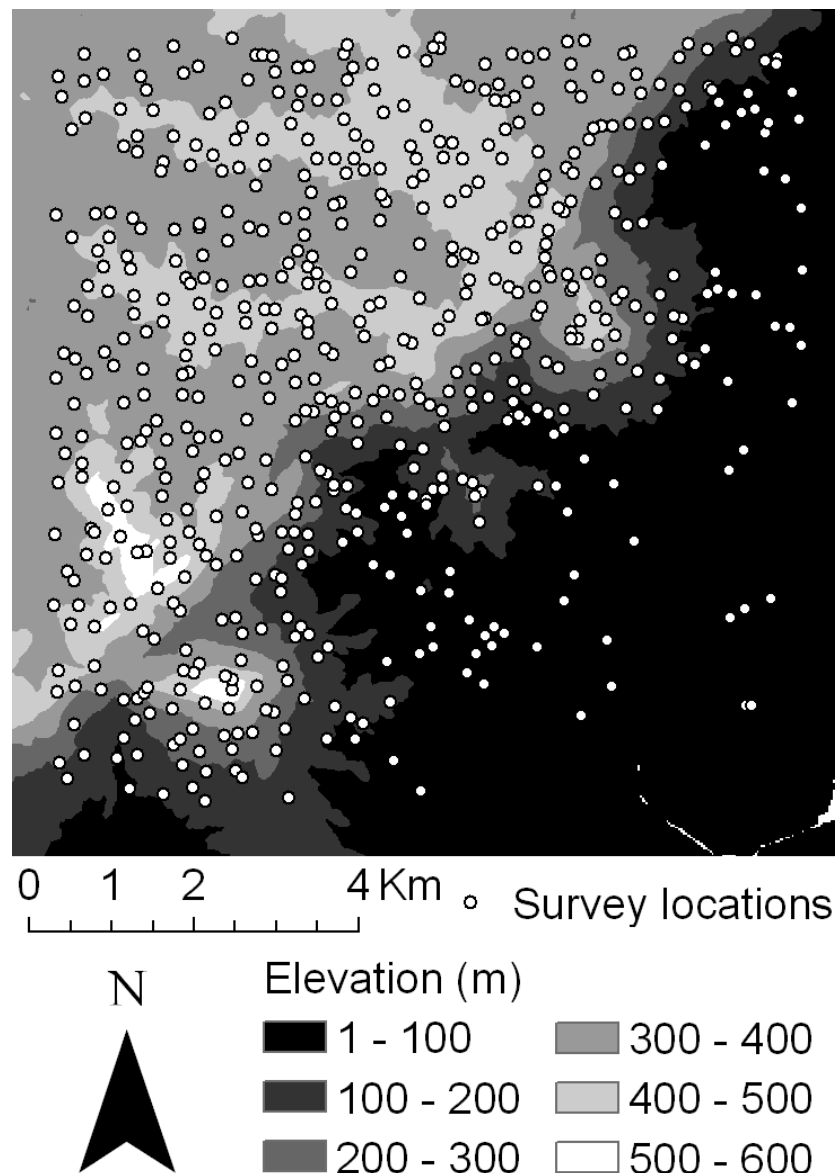


Figure 6.1: The topography of the Illawarra Escarpment, 80 km south of Sydney, Australia (34.4 °S, 150.9 °E). A vegetation survey was conducted at each of the 600 sites indicated.

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 (Ashcroft et al. 2009; Chapter 5). The results suggested that 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 (Ashcroft et al. 2009; Chapter 5). This spatial and seasonal heterogeneity meant that the distribution of 1972 temperatures was different to that of the 2005 temperature surfaces. Therefore, I subtracted the amount of warming from the 2005 surfaces to create the 1972 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.

6.2.3 Vegetation data

A vegetation survey of 600 sites (Figure 6.1) 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). See Section 4.5.2 for more details on survey design.

Table 6.1: The 37 species from the Illawarra Escarpment that were modelled as part of this study.

‘Please see print copy for image’



All sites were 20 m by 20 m and were surveyed for presence or absence of 37 common species (Table 6.1). Species were selected based on the vegetation communities present in the region (NPWS 2002), with the sample size of 600 selected so that I would obtain approximately 50 presences for most of the dominant eucalypt species. This proved effective, and only 4 of the 37 species had less than 40 presences.

6.2.4 Model production

All models were produced using Maxent version 3.2.1 (Phillips et al. 2006; 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) as 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 validated using the AUC in the remaining 30%.

The benchmark model was compared against a separate model for each time period—each containing geology and the 10 temperatures surfaces from either 1972 or 2005. The AUC results on the validation data were compared using a 2-sided paired Student t-test.

A spatial predictor was developed for each species using contextual indices. For each species, the neighbourhood average of the output from the benchmark model was calculated using a radius of 500 m. This created a smooth surface of the model output and provided a general indication of which areas have more favourable habitat. I then produced a contextual model using the contextual predictor as well as the original 21 predictors in the benchmark model. As above, the differences in AUC results were compared using a paired t-test to determine whether the contextual (spatial) models were significantly better than the benchmark (purely niche-based) models.

I 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 are 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, I determined whether more parsimonious models could perform as well as the models that considered all 21 predictors. I tested two models with only three predictors—geology, winter minimum and either summer maximum or summer minimum. Summer minimum is well correlated with elevation, while winter minimum and summer maximum had very different distributional patterns (Ashcroft et al. 2008; Chapter 3). All four predictors performed well when evaluated on their ‘alone’ contributions (Lehmann et al. 2003).

6.3 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 (Figure 6.2). Three species were more variable in AUC than the 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’.

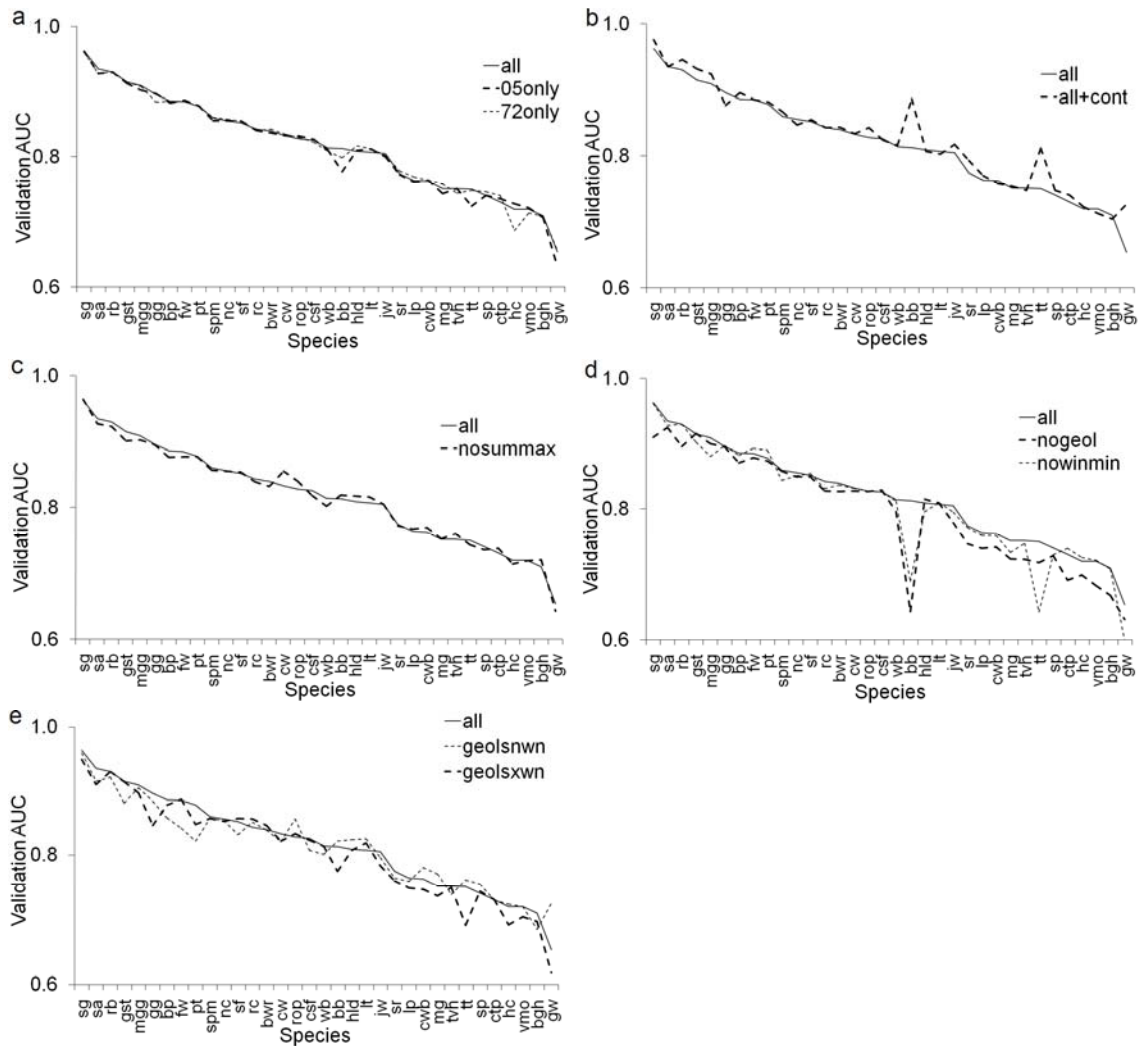


Figure 6.2: The results of Maxent models for 37 species (Table 6.1) as evaluated using the Area Under the receiver operating characteristics Curve (AUC) on a random validation dataset consisting of 30% of the surveyed sites. The ‘all’ model used geology and 20 temperature surfaces as predictors. The ‘05only’ and ‘72only’ models used geology and the 10 temperature surfaces from either 2005 or 1972 respectively. The ‘all+cont’ model used the predictors from the ‘all’ model, but added a contextual index as an additional predictor. 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.

There were no significant ($P < 0.05$) differences between the benchmark model and those based only on the 1972 or 2005 predictors (Figure 6.2a). 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 ($P = 0.058$).

The contextual indices significantly improved models when evaluated with either all species ($P = 0.012$) or without the three outliers ($P = 0.026$). The improvements did not appear related to the AUC of the benchmark models. Models with either good ($AUC > 0.9$) or fair ($AUC < 0.8$) performance were improved, although 9 of the 37 contextual models had marginally poorer performance than the benchmark models (Figure 6.2b). Contextual indices increased the average AUC of the 37 models from 0.818 to 0.827.

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 (Figure 6.2c). 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 (Figure 6.2d). 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 (Figure 6.2e). 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 I considered, and it was difficult to come up with any explanation that performed as well as the restricted species.

Two of the three species that were sensitive to the predictors used (BB, TT) were restricted to the northern portion of the study area. Indeed, there is a north-south trend in vegetation communities in the region, with Escarpment Blackbutt forest dominating in the north, Moist Coastal White Box Forest in the south, and Escarpment Moist Blue-Gum Forest in between (NPWS 2002). The benchmark models that only used niche

factors could not explain why BB and TT were restricted to the north of the study area, but the contextual model could (Figure 6.3).

As summer maximum temperature performed well in single factor models (Ashcroft et al. 2008; Chapter 3), it was somewhat surprising that excluding summer maximum did not have a detrimental effect on models (see above). Therefore, I investigated what the potential alternative explanations were. I 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 (Figure 6.4).

6.4 Discussion

6.4.1 Factors affecting the distribution of vegetation on the Illawarra Escarpment

The methods introduced in this chapter were designed to provide evidence as to which environmental factors determine the distribution of vegetation on the Illawarra Escarpment. There was a significant decrease in model performance when either winter minimum temperatures or geology were excluded from models, thus providing strong support for these predictors. If the correlations were spurious and only due to spatial structure, then other spatially structured predictors should have been able to combine to produce similar results. Instead, the results suggest that winter minimum and geology contain unique information, at least within the predictors examined. There was also strong support for spatial predictors, as models with contextual indices performed significantly better than purely niche-based models. The results provide further evidence that contextual indices produce effects similar to what would be expected by

source-sink or fragmentation effects (Chapter 4), and increase the explanatory power of SDMs (Figure 6.3).

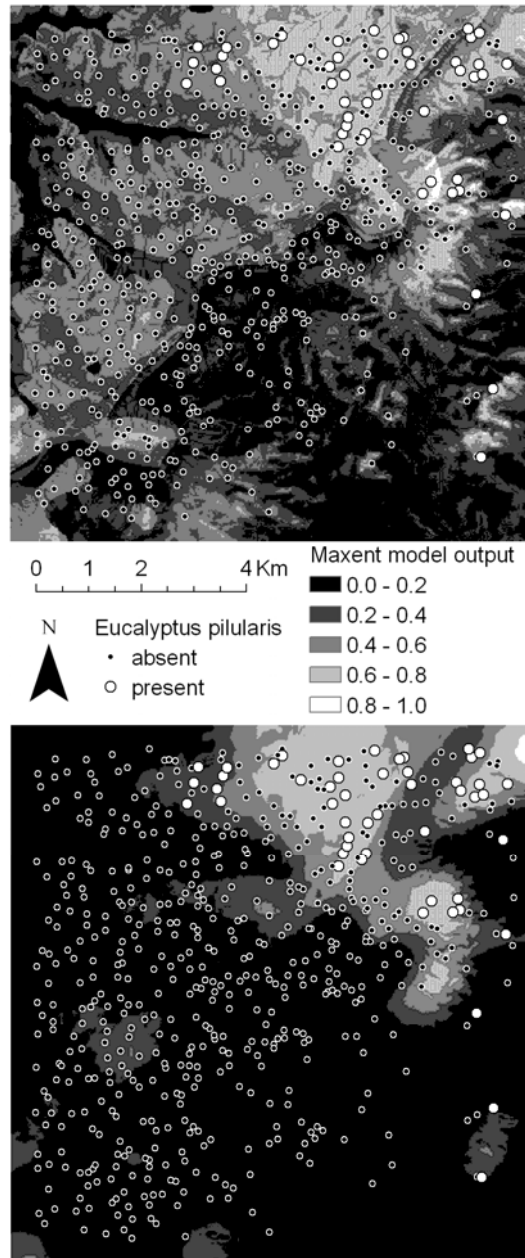


Figure 6.3: The non-contextual (top) and contextual (bottom) Maxent models for *Eucalyptus pilularis* (Blackbutt). The non-contextual model is based only on environmental niche factors (temperature and geology), while the contextual model also includes a spatial predictor based on the average model output within 500 m. The presence or absence of *E. pilularis* illustrates that the species was only observed in the north and east of the area, but the non-contextual model predicts moderate quality habitat in the south and west.

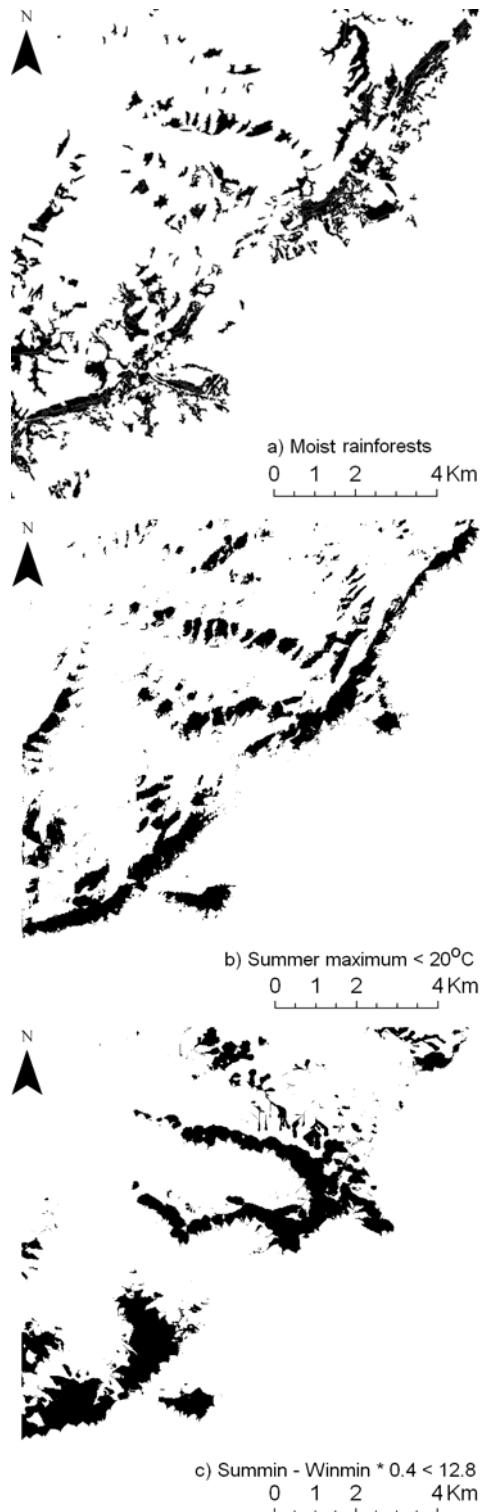


Figure 6.4: Panel (a) indicates the location of moist rainforests on the Illawarra Escarpment (Coachwood Warm Temperate Rainforest or Illawarra Escarpment Subtropical Rainforest; NPWS 2002). Panel (b) indicates the locations where summer maximum temperatures are less than 20 °C, and panel (c) indicates where summer minimum temperatures – 0.4 * winter minimum temperatures is less than 12.8.

There was no significant difference in the performance between models that contained only 1972 or 2005 temperature predictors. There was no evidence that any species, or group of species, was better modelled by one set of predictors or the other. These results are most likely because both sets of predictors have similar spatial structure, and hence explanatory ability, even though the distributions of the climate surfaces are different (Ashcroft et al. 2009; Chapter 5).

Somewhat surprisingly, there was less support for summer maximum temperature as a predictor. While summer maximums are strongly correlated with the distributions of many species (Ashcroft et al. 2008; Chapter 3), other predictors could combine to compensate for their 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 validate and refine models (Engler et al. 2004; Guisan et al. 2006b) or by manipulation experiments conducted to test the effect of different environmental factors.

6.4.2 Evidence on causal and spurious predictors

It is crucial that models include causal predictors rather than spurious correlations based on spatially structured predictors. For example, winter minimums have increased by approximately 1 °C between 1972 and 2005, while summer maximums have only increased by approximately 0.2 °C (Ashcroft et al. 2009; Chapter 5). If a spurious predictor is included in SDMs, then this will reduce predictive accuracy (Pearce and Ferrier 2000a) and dramatically alter predictions of species extinction risk. Currently, many ecological studies are over-reliant on statistical methods and parsimony, to the point that they defy known causal relationships (Clark and Agarwal 2007). While statistical methods are undoubtedly important, statistically valid models may lead to

negative conservation outcomes if they are not grounded in ecological reality (Austin 2002; Burger and Page 2007).

In this chapter I provided evidence 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. Another commonly used alternative is to make an *a priori* assumption on which predictors 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 (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 this 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.

6.4.3 Implications for model validation

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 2000b). The results of this 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. 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. 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. 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 underestimating uncertainty in the 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). The uncertainty can be dealt with by averaging the predictions from multiple alternative models (Araújo and New 2006), however this merely hides the uncertainty (Beaumont et al. 2007).

One limitation of the methods used in this chapter is that the validation was not performed with a truly independent dataset. Dividing the data into separate training and evaluation datasets 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 validation 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). Similarly, models that perform well under current climatic conditions do not necessarily perform well under other climatic conditions (Araújo and Rahbek 2006). There is often no truly independent dataset that has the same species and environmental conditions as the calibration data set (Austin 2007), and care needs to be taken to ensure that validation with an apparently independent dataset is not used to provide an overly optimistic view of model performance (Araújo et al. 2005b).

Chapter 7: Discussion: The benefits and challenges of Species Distribution Models at the landscape scale

7.1 Introduction

Species distributions can be modelled at a variety of spatial scales. Theory suggests that best results will be achieved when the spatial scale of the environmental predictors corresponds with the scale at which the important ecological processes operate, and also matches the resolution of the biological survey data (Graf et al. 2005; Guisan and Thuiller 2005). However, determining the optimal scale is also confounded by the accuracy of data used (Scott et al. 2002). If environmental data is of poor quality at any particular scale, then the corresponding species distribution models will not be optimal, and will underestimate the importance of those predictors at that scale. Therefore, studies that seek to determine the importance of environmental factors at different spatial scales need to consider the accuracy of predictors.

I conducted a series of landscape scale studies on the Illawarra Escarpment, 80 km south of Sydney, Australia. These studies were unusual in that they did not rely on elevation as a surrogate for temperature, but used empirical data to create more accurate landscape scale temperature surfaces (Ashcroft 2006; Ashcroft et al. 2008; Chapters 2 and 3). In addition, they did not rely on coarse-grained climate change models or assumptions of uniform warming, but estimated fine-scale variations in warming and potential refugia for vegetation (Ashcroft et al. 2009; Chapter 5). In this chapter I review the effects of these higher accuracy temperature surfaces on Species Distribution Models (SDMs), and discuss the lessons for the future of landscape scale modelling as a whole.

7.2 The challenges of landscape scale models

Landscape scale models are crucially important for environmental planning, as this is the scale at which many management decisions are made (Lookingbill and Urban 2003). Landscape scale SDMs can provide different insights than coarse-scale models, and yet there are unique challenges that must be overcome to obtain these benefits.

The first challenge that must be overcome is the spatial accuracy of biological data. Many studies use records from herbaria and museums, and these often have poor positional accuracy (Guisan et al. 2007). Fine-scale studies require data with higher spatial accuracy, resulting in smaller sample sizes (Engler et al. 2004). However, Engler et al. (2004) found that the higher data accuracy of fine-scale models could compensate for lower sample sizes, and lead to better results than coarse-scale models.

If there is insufficient data for fine-scale models, then existing records will need to be supplemented with field surveys. Indeed, landscape scale studies often collect their own field data (e.g. Dirnböck et al. 2003; Randin et al. 2006; Trivedi et al. 2008), while continental scale studies often use data from existing databases. Therefore, the challenge of data accuracy at the landscape scale can be overcome if sufficient resources can be obtained to collect appropriate data. This task will be easier if the benefits of landscape scale studies are known, and these will be discussed later in this chapter.

The second challenge of landscape scale studies that must be overcome is the accuracy of the environmental factors that are used in the models. Many landscape scale models use simplified climatic surfaces that are largely determined by elevation (e.g. Dirnböck et al. 2003; Trivedi et al. 2008). These may underestimate the potential contribution of climate at this scale. Indeed, I found large improvements in model performance when temperature surfaces based on empirical data were used instead of

elevation on the Illawarra Escarpment (Ashcroft et al. 2008; Chapter 3). Summer maximum temperatures were poorly correlated with elevation ($r^2 < 0.1$), and were predominately determined by exposure to hot, dry northwesterly winds. Similarly, winter minimums were correlated more strongly with distance from coast than elevation.

It has been known for sometime that elevation is not a sufficient surrogate for temperature at the landscape scale (Lookingbill and Urban 2003), and yet there has been little work done to determine how more accurate temperature surfaces affect SDMs. There is great potential for improved spatial datasets of environmental factors to lead to a new generation of gradient analysis at the landscape scale (Lookingbill and Urban 2005). The resources required to generate these surfaces is much less than that to gather the necessary biotic data. For example, I spent approximately 75 days surveying the vegetation at 600 sites in my study area, but only needed 18 days to gather sufficient data on temperature. The cost of the temperature sensors (DS1921G iButtons in this case) was not high, and so approximately 30% more resources were required to supplement the biotic survey with higher accuracy temperature surfaces. This extra expense yielded models based on more direct and accurate environmental predictors, which more than justified the additional cost. Other iButtons are available that also measure humidity and these would provide further information on fine-scale variability in growing conditions.

The third challenge for landscape scale models is to obtain data over a broad range of environmental conditions. The environmental niche of a species can only be properly defined if the conditions extend past the environmental limits (Austin 2007), and therefore a broad range of environmental conditions is needed to increase the applicability of landscape scale SDMs.

This challenge is more difficult to overcome, as landscapes are expected to have a narrower range of environmental conditions than the continents that they reside within. In addition, even if landscapes do cover a broad range of conditions, they may not cover all combinations of predictors. This problem was evident in my study on the Illawarra Escarpment. For example, *Eucalyptus sieberi* is common on the Hawkesbury sandstone in the Sydney basin. In my study area it only occurred at elevations of approximately 300 m to 600 m, however I know it exists at other elevations elsewhere. A model based on elevation performed well for this species—even when validated with an ‘independent’ dataset from the same study area—however it would not be broadly applicable. It is difficult to determine how many landscape scale models suffer this same problem.

The third challenge can potentially be overcome by sampling multiple landscapes, however this may be prohibitively expensive given the cost of obtaining appropriate biotic and environmental data (discussed earlier), and may be beyond the immediate scope and needs of the research. Where landscape scale models do not cover the full range of environmental conditions, this will impact their transferability—that is, the applicability of the models in areas other than the one where they were created. There is already some evidence that this is the case. Continental scale models have been transferred with some success (e.g. Peterson 2003; Peterson et al. 2007), while landscape scale models have performed more poorly (Randin et al. 2006). This highlights an important limitation of landscape scale models—they can not be transferred to new areas without verifying that the range of conditions in the new areas are similar to those in the area where the models were developed.

These limitations and challenges that must be overcome to produce landscape scale SDMs must be offset against the potential benefits. I will now review four potential benefits of landscape scale SDMs that were apparent from my study.

7.3 Benefit 1: reduction of pseudoreplication

The first benefit of landscape scale models is a reduction in pseudoreplication. For example, consider the simple relationship between *Eucalyptus cypellocarpa* and summer maximum temperature. The species only occurs over a limited temperature range at the continental scale, however this temperature range only occurs in southeast Australia (Figure 7.1). Even though this is not a manipulative experiment, it is clear that there is pseudoreplication due to the lack of interspersed ‘treatments’ (Hurlbert 1984). *E. cypellocarpa* could be limited to southeast Australia by summer maximum temperature, but it may also be due to neutral processes (Hubbell 2001), or by any number of other spatially structured environmental factors or biological processes. A continental-scale model produced using Maxent performs very well statistically (AUC of 0.992 when 70% of data used to produce model and 30% to validate), but it is inappropriate to infer that summer maximum is important for this species because there are other possible explanations (Faith 2003; McGill 2003; Faith et al. 2004).

This problem has recently gained attention by comparing how SDMs perform when compared with models produced using randomised spatially structured predictors (Bahn and McGill 2007; Currie 2007). These studies showed that almost any spatially structured predictor has the power to explain species distributions, and this casts doubt over many continental-scale models.

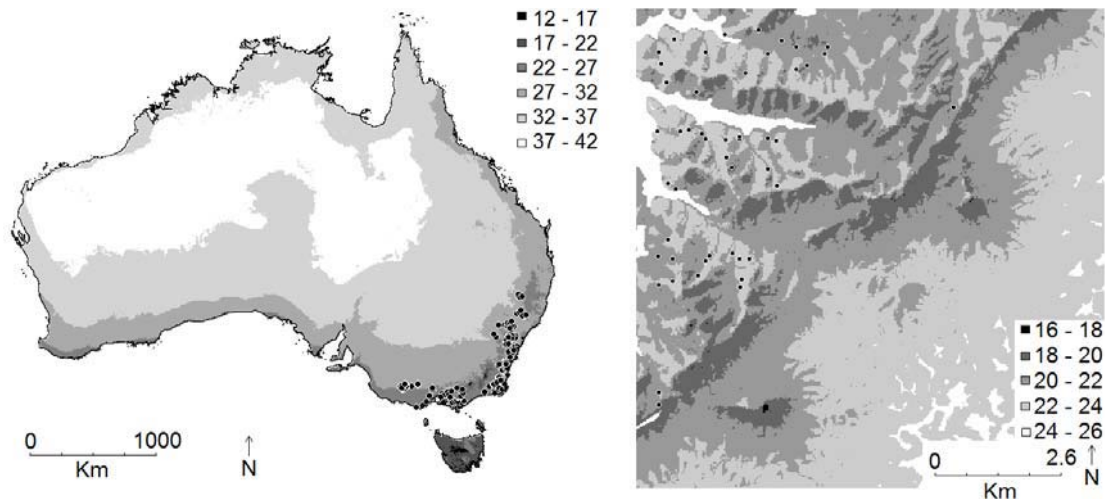


Figure 7.1: A comparison between the distribution of *Eucalyptus cypellocarpa* (black dots) and summer maximum temperature (°C) at the continental (left) and landscape (right) scales. Continental scale data was obtained from the Global Biodiversity Information Facility (<http://data.gbif.org>) and ANUCLIM (Houlder et al. 2000). Landscape scale data was obtained from my study on the Illawarra Escarpment (Ashcroft et al. 2008; Chapter 3).

Landscape scale models cannot solve this problem completely, but they can reduce the effect. For example, *Eucalyptus cypellocarpa* still favours a limited range of summer maximum temperatures at the landscape scale (Figure 7.1), but these are now distributed across the whole study area and there are interspersed patches of unfavourable temperature (treatment). This reduces pseudoreplication, and reduces the probability that a randomised spatially structured variable would be able to explain the distribution. This is not to say that the landscape scale will work best for all species, however, and modellers are encouraged to select a scale that maximises interspersion and minimises pseudoreplication.

7.4 Benefit 2: less predictor colinearity

Many climatic predictors are created using elevation sensitive interpolations (e.g. Houlder et al. 2000), and this leads to high colinearity between different seasonal

temperature predictors. This is a problem for model building, and can lead to the erroneous selection of environmental factors and distorted response curves (MacNally 2002; Graham 2003; Lehmann et al. 2003; Chapter 6).

The uncertainty introduced by correlated predictors is especially undesirable in a climate change context. For example, temperatures are typically increasing much more in winter and spring than summer and autumn, and minimum temperatures are increasing almost twice as much as maximum temperatures (Hughes 2000; Walther et al. 2002; Ashcroft et al. 2009; Chapter 5). If the wrong seasonal temperature predictors are included in models then this will dramatically over- or under-estimate the predicted effect of climate change, and hence introduce error into the estimated extinction risk.

To mitigate the potential for the wrong predictors to be selected in SDMs, colinearity in temperature predictors should be reduced by establishing relationships with a variety of environmental factors—not just elevation. I did this on the Illawarra Escarpment, and found that the correlation between summer maximum temperatures and winter minimums temperatures was actually very low ($r^2 = 0.07$; Ashcroft et al. 2008; Chapter 3). As noted above, this was because temperatures were affected by different factors during different seasons, and the spatial distribution of seasonal temperatures displayed vastly different patterns (Ashcroft et al. 2008; Chapter 3). Some seasonal temperature predictors were still highly correlated with elevation and each other, but the methods did reduce the problem to some extent. It is especially noteworthy that it was the extreme temperatures that had the poorest correlation with elevation (summer maximum and winter minimum), and these are the factors that are frequently selected in models because they are physiologically limiting for many species (Ashcroft et al. 2008; Chapter 3).

Using a variety of factors to explain the spatial distribution of temperature is not necessarily a method that can only be applied at the landscape scale. Mountainous landscapes are more likely to have complex temperature mosaics than coarse-scale study areas (e.g. Figure 7.1), but factors other than elevation will also affect the distribution of regional and continental temperatures. To discuss this in more detail, it is necessary to consider the benefits and implications of standardised weather stations.

Temperatures are affected by a number of factors. Standardisation of climatic recording equipment (e.g. Stevenson screens) is performed so that the temperatures in different locations can be easily compared and interpreted. That is, standardisation ensures that differences are due to coarse-scale positional factors (elevation, distance to coast, latitude) and not due to localised variations in vegetation, topography, moisture, human structures, or micro-climatic factors.

Standardised temperature recordings are designed to be representative of the surrounding landscape, however they are really only representative of areas that match the standard. That is, 1.25 to 2 m above ground, no vegetation, flat topography, etc. Therefore, standardised climate data has been referred to as the ‘human climate’, and is a poor indicator of the ‘habitat climate’ that is of interest when studying species such as ground dwelling animals and trees that germinate in forests (Wolfe 1945; Geiger 1971). While it has long been known that human climate is not ideal for ecological studies, climate surfaces based on standardised data are still commonly used, as little else is available.

It is possible to create climate surfaces that consider variations in topography and other factors (Lookingbill and Urban 2003; Ashcroft et al. 2008; Chapter 3), however this cannot be done with standard weather station data. Weather station data can be used to detect temperature trends associated with elevation, distance to coast (continentality)

and latitude (e.g. McVicar et al. 2007), but cannot predict the effects of vegetation, topography, and other factors that are standardised, because there is no information on how temperatures are affected as these factors vary. Therefore, creating surfaces that are better approximations of habitat climate requires a purpose built network of ‘climate stations’ with less standardisation.

Climate stations with less standardisation will be less comparable with other locations, because differences could be due to any number of environmental differences. It will not be possible to take any one station from a region and imply it is representative of that region, nor compare two stations to make inferences about which region is hotter. Recorded temperatures may change over short distances due to changes in vegetation, moisture or topography, however this is not a problem if the data is analysed to determine the effect of each of these factors. More weather stations are needed so that the effect of each factor can be reliably determined, and the less standardisation is performed, the larger the sample size will need to be.

There can actually be a certain amount of ‘error’ introduced by non-standardised weather stations, as the temperature at any location is no longer of primary concern. Instead, the data is used to determine how temperatures are affected as each environmental factor varies, as this is used to estimate the temperature at each location in the study area. For example, standardised weather stations may have a stronger trend between temperature and elevation, and even though reducing standardisation will weaken this trend, it will still be present (Figure 7.2). The advantage of this extra ‘noise’ is that the effect of other factors can also be determined.

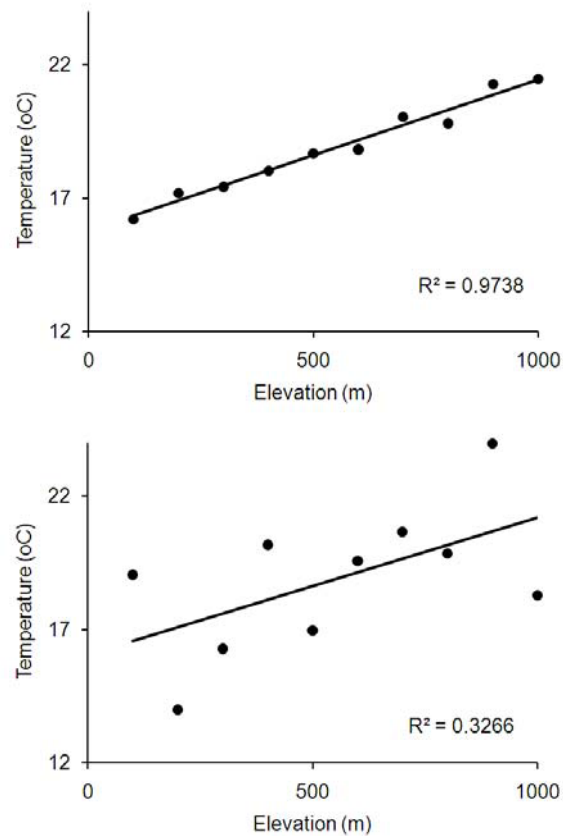


Figure 7.2: The graphs illustrate simulated results from standardised (top) and unstandardised (bottom) weather stations. Standardised weather stations have limited noise introduced by factors such as vegetation and topography, and show a strong trend towards factors such as elevation. Unstandardised weather stations are noisier, because temperatures are influenced by a variety of factors. This means that it is more difficult to compare different locations, but it is now possible to work out how multiple factors affect temperatures (see text for more information).

In summary, standardised weather stations give a better indication of the general climate in a region, and are better for comparing the ‘human climate’ in different regions. Reducing standardisation makes comparing different weather stations more problematic, but allows scientists to determine how different factors affect the distribution of temperatures, and produce temperature surfaces that are a better representation of ‘habitat climate’. As the relative effect of each factor varies

seasonally, this reduces colinearity between seasonal temperature predictors. Producing habitat climate surfaces may have been prohibitively expensive in the past, but is now possible due to the ready availability of cheap, automatic temperature sensors.

7.5 Benefit 3: identification of refugia from climate change

Current climate change models may dramatically overestimate extinction risk because they are at too coarse a scale to identify refugia where species can persist despite apparently unfavourable regional conditions (Thuiller et al. 2005; Anciães and Peterson 2006; Pearson 2006; Botkin et al. 2007). This is classically illustrated by the long-term survival of the Wollemi Pine (*Wollemia nobilis*), an ancient species that has been able to survive in deep rainforest gorges 150 km northwest of Sydney, presumably because the topography offers shelter from regular bushfires and harsh climates (Offord et al. 1999).

My work on the Illawarra Escarpment identified that rainforests were also confined to sheltered refugia within the landscape. Moist rainforests only occur where summer maximum temperatures are less than 20–20.5 °C (Figure 7.3; Ashcroft et al. 2008; Chapter 3), and these locations have not warmed by as much as the more exposed locations (Ashcroft et al. 2009; Chapter 5). Crucially, the identification of these refugia relied heavily on factors other than elevation, as their locations were predominately determined by shelter from the hot, dry northwesterly winds that are associated with extremely hot temperatures and bushfire hazards. It is not surprising, therefore, that landscape scale studies that rely on elevation to produce temperature surfaces cannot identify refugia (Trivedi et al. 2008).

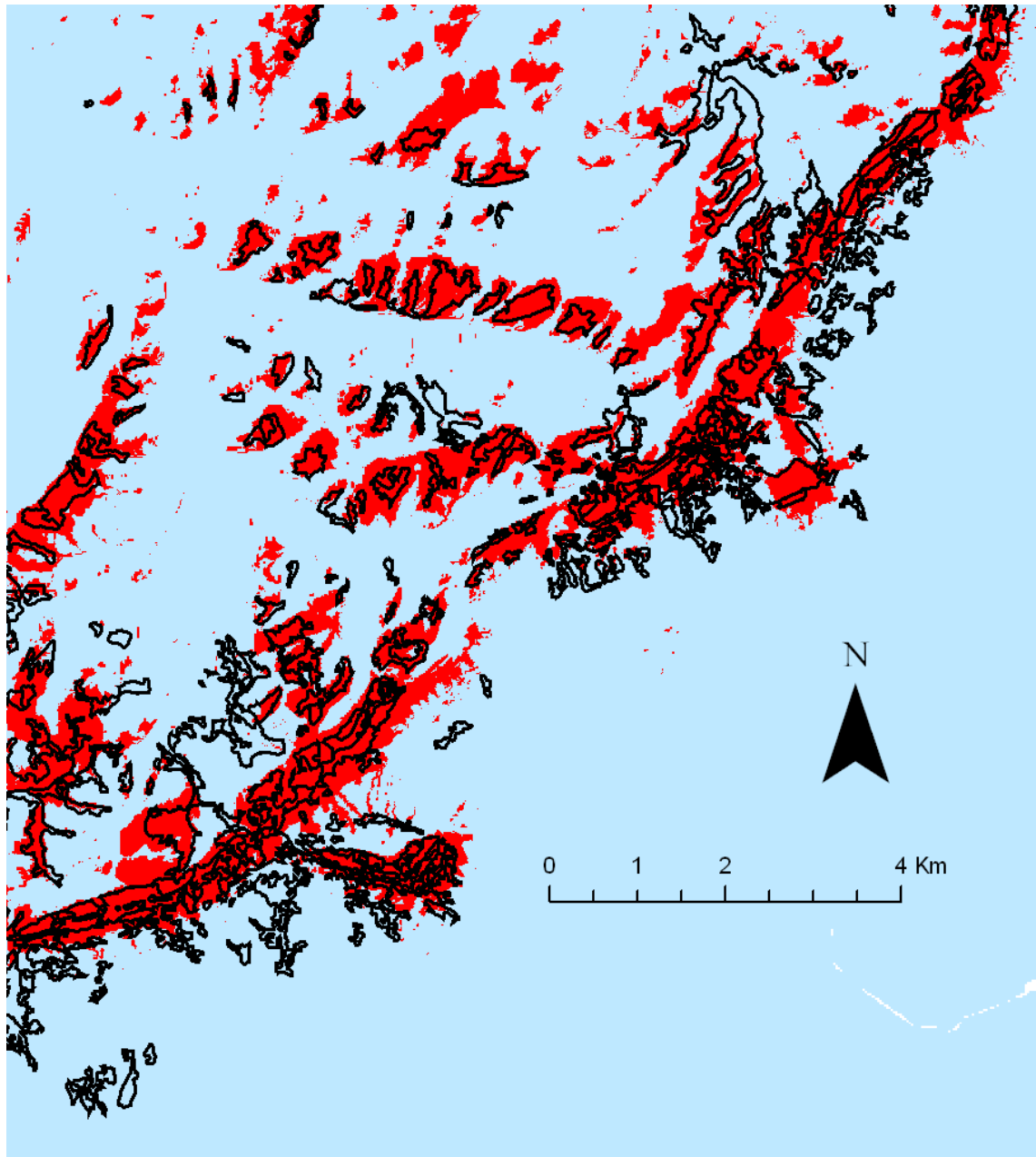


Figure 7.3: The locations where summer maximum temperatures are less than 20.5 °C are shown in red, while the locations of moist rainforest communities are shown as black polygons. Temperature data is from Ashcroft et al. (2008; Chapter 3), while vegetation data is from NPWS (2002).

Temperature surfaces based only on elevation predict that temperatures in the deep rainforest gorges are warmer than the surrounding area, because they are at lower elevation. They do not identify any difference between the gorges and locations at

similar elevations in more exposed positions. The data I gathered clearly illustrated that the rainforest gullies were cooler than the surrounding areas, and were very different from exposed locations at similar elevations.

The identification of these rainforest refugia offers an alternative to the simplified generalisation that species will shift poleward and upward in response to climate change (Hughes 2000). Moist rainforest communities are already restricted to small patches where the summer maximum temperature is lower than the surrounding region (Figure 7.3), and models suggest they cannot move in any direction in response to climate change. As temperatures warm, these rainforest patches will only be able to contract. Fortunately, however, summer maximum temperatures have not increased as much as other seasonal temperatures (Hughes 2000; Walther et al. 2002; Ashcroft et al. 2009; Chapter 5), and temperatures would have to increase substantially more to cause a complete loss of all favourable rainforest habitat.

Many other species on the Illawarra Escarpment are best modelled using winter minimum temperatures, which are predominately determined by distance to the coast (Ashcroft et al. 2008; Chapters 3 and 6). Similarly to above, these species would not be expected to move upward or poleward either. Upward shifts are generally unlikely, because the species that currently occupy the nutrient-rich soils on the escarpment slopes are unlikely to survive on the nutrient-poor Hawkesbury sandstone that is found at higher elevations. Southward shifts are possible, but winter minimum temperatures decrease more rapidly towards inland areas, and therefore it is probably easier to move inland than southward. The eucalypts that are currently only found in the cooler areas of the Escarpment, such as *Eucalyptus cypellocarpa*, *E. smithii*, and *E. piperita*, are three examples of species which may struggle to persist at the easternmost limits of their distributions.

Landscape scale climate change predictions are important for conservation planning, as this is the scale at which management decisions are often made (Lookingbill and Urban 2003). It is problematic to associate landscape scale species distributions with continental scale climate predictors, and there is a need to ensure that both climatic data and biotic data are at the same scale (Walther et al. 2005). The fact that coarse scale generalisations of poleward and upward shifts to climate change did not hold on the Illawarra Escarpment illustrates the difference that landscape scale studies can make.

7.6 Benefit 4: matching scale and process

Species respond to different environmental conditions at different spatial scales, so it is impossible to determine an optimal scale at which modelling should be performed (Graf et al. 2005). Multi-scale studies often identify different environmental factors as being important at different spatial scales (e.g. Lindenmayer 2000; Lindenmayer et al. 2000; del Barrio et al. 2006), and this has lead to some general rules of thumb. For example, it has been suggested that climate may be the dominant factor in continental scale models, while land-use, soil variables and biotic interactions become more important at finer scales (Pearson and Dawson 2003). These rules of thumb require further examination to determine whether they truly reflect the scale at which these processes operate, or whether they reflect the scale at which the predictors are most accurately recorded. For example, my results provide strong evidence that climate is important in fine scale models (Ashcroft et al. 2008; Chapter 3), and it would be premature to discount this using fine-scale climate surfaces based on elevation or simple downscaling methods. Similarly, the importance of soil variables at coarse scales may be improved if better soil maps were made available at that scale.

Landscape scale models have an advantage in that they can model processes that cannot be detected using coarse-grained analysis. For example, it is more robust to study plant competition at high spatial resolution models, where it is known that species coexist in small areas (Leathwick and Austin 2001; Austin 2007). It is more problematic to infer species are interacting in coarse-scale models, as species that coexist within large grains (up to 50 km by 50 km) may actually be geographically separate or in different habitats.

Similarly, it has been shown that species respond to conditions in the surrounding areas. Contextual indices, such as the amount of rainforest within 500 m, have been used to model mobile fauna species (Ferrier et al. 2002; Wintle et al. 2005) as well as sessile flora (Chapter 4). Contextual indices offer a means to capture the home range of organisms, as well as source-sink, mass effects and fragmentation effects (Chapter 4). These processes cannot be studied if the grain-size is larger than the distance over which neighbouring locations have an effect.

Another advantage of landscape scale models is that the temperature of each grain is more likely to represent the temperature that the species' actually experiences. There can be up to 33.8 °C variation in temperature within one 50 km by 50 km grain (mean 1.8 °C; Hijmans et al. 2005), and each grain usually contains the average for that area. A species that can only tolerate summer maximum temperatures of 20–25 °C may be found in locations with temperatures well outside this range in coarse-scale models. The optimal scale for models depends on their purpose (Austin 2007), and landscape scale models are better suited to studying the species responses curves along environmental gradients (Lookingbill and Urban 2005).

7.7 Conclusions

There is no one correct scale to model species distributions, and landscape scale models have unique challenges and benefits. Landscape scale models require biotic and environmental data that is of higher accuracy, and will not be widely applicable unless they are produced in a landscape that has a broad range of environmental conditions. The benefits of landscape scale models include a reduction in pseudoreplication and predictor colinearity, improved ability to detect refugia from climate change, and the ability to detect processes that are obscured in coarser-scale models.

The underlying theme that is common across a number of these challenges and benefits is predictor accuracy. Error in environmental predictors is usually ignored in SDMs (Austin 2007), but can have a large impact on the results (Van Niel et al. 2004; Van Niel and Austin 2007). More attention needs to be paid to predictor accuracy (Dormann et al. 2008), and should be considered before predictors are selected in species distribution models (Van Niel and Austin 2007).

Environmental predictor data is not scale independent, and this hinders the development of SDMs that may otherwise be scale independent (Pearson et al. 2002). However, there are opportunities to better link landscape scale models with coarser scale models. One opportunity is to take advantage of fine-scale environmental data, even when biotic data is inaccurate and can only be used at a coarse-scale. Luoto and Heikkinen (2008) did this by incorporating the within-cell difference in elevation as a surrogate for the within-cell variability in climate noted earlier (Hijmans et al. 2005). Another option is to replace coarse-scale climate predictors with predictors based on contextual indices. For example, suppose a landscape scale models finds that a species favours locations where a large proportion of the surrounding landscape has a summer maximum temperature between 20 and 25 °C. Instead of producing a coarse-grained

model based directly on summer maximum temperature, fine-grained temperature data could be used to estimate the proportion of each coarse-grained cell that is within this temperature range. Effectively, the coarse-scale model would then be based on the amount of favourable habitat within each cell instead of the average conditions, and there would be stronger links between fine and coarse scale models.

The other advantages of landscape scale models could also be incorporated into coarse scale models. If landscape scale models have less pseudoreplication and predictor collinearity, then they could be useful to select which predictors should be included in coarse scale models. If the locations of refugia are easier to identify with fine-scale models, then it may be possible to include these locations as predictors in coarse-scale models. Environmental planning and management will undoubtedly benefit if models at different scales are linked to obtain a more complete understanding of species ecology. Landscape scale models are undoubtedly useful for environmental planning and management, but they are not the complete solution.

References

- Anciães M, Peterson AT (2006) Climate change effects on neotropical manakin diversity based on ecological niche modelling. *The Condor*, 108, 778–791.
- Araújo MB, Williams PH (2000) Selecting areas for species persistence using occurrence data. *Biological Conservation*, 96, 331–345.
- Araújo MB, Whittaker RJ, Ladle RJ, Erhard M (2005a) Reducing uncertainty in projections of extinction risk from climate change. *Global Ecology and Biogeography*, 14, 529–538.
- Araújo MB, Pearson RG, Thuiller W, Erhard M (2005b) Validation of species-climate impact models under climate change. *Global Change Biology*, 11, 1504–1513.
- Araújo MB, Guisan A (2006) Five (or so) challenges for species distribution modelling. *Journal of Biogeography*, 33, 1677–1688.
- Araújo MB, New M (2006) Ensemble forecasting of species distributions. *Trends in Ecology and Evolution*, 22, 42–47.
- Araújo MB, Rahbek C (2006) How does climate change affect biodiversity? *Science*, 313, 1396–1397.
- Araújo MB, Luoto M (2007) The importance of biotic interactions for modelling species distributions under climate change. *Global Ecology and Biogeography*, 16, 743–753.
- Arii K, Parrott L (2006) Examining the colonization process of exotic species varying in competitive abilities using a cellular automaton model. *Ecological Modelling*, 199, 219–228.
- Arundel ST (2005) Using spatial models to establish climatic limiters of plant species' distributions. *Ecological Modelling*, 182, 159–181.

- Ashcroft MB (2006) A method for improving landscape scale temperature predictions and the implications for vegetation modelling. *Ecological Modelling*, 197, 394–404.
- Ashcroft MB, Chisholm LA, French KO (2008) The effect of exposure on landscape scale soil surface temperatures and species distribution models. *Landscape Ecology*, 23, 211–225.
- Ashcroft MB, Chisholm LA, French KO (2009) Climate change at the landscape scale: predicting fine-grained spatial heterogeneity in warming and potential refugia for vegetation. *Global Change Biology*, 15, 656–667.
- Austin MP, Nicholls AO, Doherty MD, Meyers JA (1994) Determining species response functions to an environmental gradient by means of a β -function. *Journal of Vegetation Science*, 5, 215–228.
- Austin MP, Meyers JA (1996) Current approaches to modelling the environmental niche of eucalypts: implication for management of forest biodiversity. *Forest Ecology and Management*, 85, 95–106.
- Austin MP, Pausas JG, Noble IR (1997) Modelling environmental and temporal niches of eucalypts. In: JE Williams, JCZ Woinarski (Editors), *Eucalypt Ecology*. Cambridge University press, Cambridge, UK, pp. 129–150.
- Austin MP (2002) Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological Modelling*, 157, 101–118.
- Austin MP, Belbin L, Meyers JA, Doherty MD, Luoto M (2006) Evaluation of statistical models used for predicting plant species distributions: Role of artificial data and theory. *Ecological Modelling*, 199, 197–216.
- Austin M (2007) Species distribution models and ecological theory: a critical assessment and some possible new approaches. *Ecological Modelling*, 200, 1–19.

- Bahn V, McGill BJ (2007) Can niche-based distribution models outperform spatial interpolation? *Global Ecology and Biogeography*, 16, 733–742.
- Bakkenes JR, Alkemade RM, Ihle F, Leemans R, Latour JB (2002) Assessing effects of forecasted climate change on the diversity and distribution of European higher plants for 2050. *Global Change Biology*, 8, 390–407.
- Balster H, Braun PW, Köhler W (1998) Cellular automata models for vegetation dynamics. *Ecological Modelling*, 107, 113–125.
- Balster H (2000) Markov chain models for vegetation dynamics. *Ecological Modelling*, 126, 139–154.
- Beadle NCW (1954) Soil phosphate and the delimitation of plant communities in eastern Australia. *Ecology*, 35, 370–375.
- Beadle NCW (1966) Soil phosphate and its role in molding segments of the Australian flora and vegetation, with special reference to xeromorphy and sclerophylly. *Ecology*, 47, 992–1007.
- Beaumont LJ, Hughes L (2002) Potential changes in the distributions of latitudinally restricted Australian butterfly species in response to climate change. *Global Change Biology*, 8, 954–971.
- Beaumont LJ, Pitman AJ, Poulsen M, Hughes L (2007) Where will species go? Incorporating new advances in climate modelling into projections of species distributions. *Global Change Biology*, 13, 1368–1385.
- Bell DT, Williams JE (1997) Eucalypt ecophysiology. In: JE Williams, JCZ Woinarski (Editors), *Eucalypt Ecology*. Cambridge University press, Cambridge, UK, pp. 168–196.

- Bigler C, Bugmann H (2004) Assessing the performance of theoretical and empirical tree mortality models using tree-ring series of Norway spruce. *Ecological Modelling*, 174, 225–239.
- Bleher B, Oberrath R, Böhning-Gaese K (2002) Seed dispersal, breeding system, tree density and the spatial pattern of trees – a simulation approach. *Basic and Applied Ecology*, 3, 115–123.
- Bond-Lamberty B, Wang C, Gower ST (2005) Spatiotemporal measurement and modeling of stand-level boreal forest soil temperatures. *Agricultural and Forest Meteorology*, 131, 27–40.
- Botkin DB, Saxe H, Araújo MB, Betts R, Bradshaw RHW, Cedhagen T, Chesson P, Dawson TP, Etterson JR, Faith DP, Ferrier S, Guisan A, Hanson AS, Hilbert DW, Loehle C, Margules C, New M, Sobel MJ, Stockwell DRB (2007) Forecasting the effects of global warming on biodiversity. *Bioscience*, 57, 227–236.
- Burger AE, Page RE (2007) The need for biological realism in habitat modeling: a reinterpretation of Zharikov et al. (2006). *Landscape Ecology*, 22, 1273–1281.
- Bywater J (1985) Edaphic and Other Environmental Conditions Associated with Rainforest Development in the Illawarra District. M.Sc. Thesis. Department of Geography, The University of Wollongong, Wollongong.
- Campbell GS, Norman JM (1998) *An Introduction to Environmental Biophysics*. Springer-Verlag, New York.
- Canham CD, Uriarte M (2006) Analysis of neighborhood dynamics of forest ecosystems using likelihood methods and modeling. *Ecological Applications*, 16, 62–73.

- Chuanyan Z, Zhongren N, Guodong C (2005) Methods for modelling of temporal and spatial distribution of air temperature at landscape scale in the southern Qilian mountains, China. *Ecological Modelling*, 189, 209–220.
- Childress WM, Crisafulli CM, Rykiel Jr EJ (1998) Comparison of Markovian matrix models of a primary successional plant community. *Ecological Modelling*, 107, 93–102.
- Clark JS, Agarwal PK (2007) Rejoinder to Clark et al. (2007): response to Chesson and Rees. *Ecology Letters*, 10, 661–662.
- Coudun CC, Gégout J-C (2006) The derivation of species response curves with Gaussian logistic regression is sensitive to sampling intensity and curve characteristics. *Ecological Modelling*, 199, 164–175.
- Coudun C, Gégout J-C, Piedallu C, Rameau J-C (2006) Soil nutritional factors improve models of plant species distribution: an illustration with *Acer campestre* (L.) in France. *Journal of Biogeography*, 33, 1750–1763.
- Crosetto M, Tarantola S, Saltelli A (2000) Sensitivity and uncertainty analysis in spatial modelling based on GIS. *Agriculture, Ecosystems and Environment*, 81, 71–79.
- Currie DJ (2007) Disentangling the roles of environment and space in ecology. *Journal of Biogeography*, 34, 2009–2011.
- Dale MB, Dale PER, Li C, Biswas G (2002) Assessing the impacts of small perturbations using a model-based approach. *Ecological Modelling*, 156, 185–199.
- Davies KF, Margules CR, Lawrence JF (2000) Which traits of species predict population declines in experimental forest fragments? *Ecology*, 81, 1450–1461.
- Debinski DM (2006) Forest fragmentation and matrix effects: the matrix does matter. *Journal of Biogeography*, 33, 1791–1792.

- de Frutos Á, Olea PP, Vera R (2007) Analyzing and modelling spatial distribution of summering lesser kestrel: the role of spatial autocorrelation. *Ecological Modelling*, 200, 33–44.
- del Barrio G, Harrison PA, Berry PM, Butt N, Sanjuan ME, Pearson RG, Dawson T (2006) Integrating multiple modelling approaches to predict the potential impacts of climate change on species' distributions in contrasting regions: comparison and implications for policy. *Environmental Science and Policy*, 9, 129–147.
- Dirnböck T, Dullinger S, Grabherr G (2003) A regional impact assessment of climate and land-use change on alpine vegetation. *Journal of Biogeography*, 30, 401–417.
- Dirnböck T, Dullinger S (2004) Habitat distribution models, spatial autocorrelation, functional traits and dispersal capacity of alpine plant species. *Journal of Vegetation Science*, 15, 77–84.
- Dormann CF (2007) Assessing the validity of autologistic regression. *Ecological Modelling*, 207, 234–242.
- Dormann CF, Purschke O, García Márquez JR, Lautenbach S, Schröder B (2008) Components of uncertainty in species distribution analysis: a case study of the great grey shrike. *Ecology*, 89, 3371–3386.
- Dullinger S, Dirnböck T, Grabherr G (2004) Modelling climate change-driven treeline shifts: relative effects of temperature increase, dispersal and invasibility. *Journal of Ecology*, 92, 241–252.
- Dymond CC, Johnson EA (2002) Mapping vegetation spatial patterns from modeled water, temperature and solar radiation gradients. *ISPRS Journal of Photogrammetry and Remote Sensing*, 57, 69–85.
- Elith J, Graham CH, Anderson RP, Dudik M, Ferrier S, Guisan A, Hijmans RJ, Huettmann F, Leathwick JR, Lehmann A, Li J, Lohmann LG, Loiselle BA,

- Manion G, Moritz C, Nakamura M, Nakazawa Y, Overton JMcC, Peterson AT, Phillips SJ, Richardson K, Scachetti-Pereira R, Schapire RE, Soberón J, Williams S, Wisz MS, Zimmermann NE (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. *Journal of Applied Ecology*, 41, 263–274.
- Ersline J (1984) The Distributional Ecology of Rainforest in the Illawarra in Relation to Fire. B.Sc.(Hons) Thesis. Department of Biology, The University of Wollongong, Wollongong.
- Faith DP (2003) Environmental diversity (ED) as surrogate information for species-level biodiversity. *Ecography*, 26, 374–379.
- Faith DP, Ferrier S, Walker PA (2004) The ED strategy: how species-level surrogates indicate general biodiversity patterns through an 'environmental diversity' perspective. *Journal of Biogeography*, 31, 1207–1217.
- Fahrig L, Merriam G (1994) Conservation of fragmented populations. *Conservation Biology*, 8, 50–59.
- Ferrier S, Watson G, Pearce J, Drielsma M (2002) Extended statistical approaches to modelling spatial pattern in biodiversity in northeast New South Wales. I. Species-level modelling. *Biodiversity and Conservation*, 11, 2275–2307.
- 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. *Diversity and Distributions*, 13, 252–264.

- Fielding AH, Bell JF (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, 24, 38–49.
- Finardi S, Tinarelli G, Faggian P, Brusasca G (1998) Evaluation of different wind field modeling techniques for wind energy applications over complex topography. *Journal of Wind Engineering and Industrial Aerodynamics*, 74–76, 283–294.
- Florence RG (2004) *Ecology and Silviculture of Eucalypt Forests*. CSIRO Publishing, Victoria, Australia.
- Foody GM (2004) Spatial nonstationarity and scale-dependency in the relationship between species richness and environmental determinants for the sub-Saharan endemic avifauna. *Global Ecology and Biogeography*, 13, 315–320.
- Frank D, Esper J (2005) Characterization and climate response patterns of a high-elevation, multi-species tree-ring network in the European Alps. *Dendrochronologia*, 22, 107–121.
- Fuller L (1995) *Wollongong's Native Trees*. Kingsclear Books, Alexandria.
- Geiger R (1971) *The Climate Near the Ground* (revised edition). Harvard University Press, Cambridge, Massachusetts.
- Graf RF, Bollmann K, Suter W, Bugmann H (2005) The importance of spatial scale in habitat models: capercaillie in the Swiss Alps. *Landscape Ecology*, 20, 703–717.
- Graham MH (2003) Confronting multicollinearity in ecological multiple regression. *Ecology*, 84, 2809–2815.
- Gravel D, Canham CD, Beaudet M, Messier C (2006) Reconciling niche and neutrality: the continuum hypothesis. *Ecology Letters*, 9, 399–409.
- Guisan A, Weiss SB, Weiss AD (1999) GLM versus CCA spatial modeling of plant species distribution. *Plant Ecology*, 143, 107–122.

- Guisan A, Zimmermann NE (2000) Predictive habitat distribution models in ecology. *Ecological Modelling*, 135, 147–186.
- Guisan A, Edwards Jr TC, Hastie T (2002) Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological Modelling*, 157, 89–100.
- Guisan A, Thuiller W (2005) Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, 8, 993–1009.
- Guisan A, Lehmann A, Ferrier S, Austin M, Overton JMcC, Aspinall R, Hastie T (2006a) Making better biogeographical predictions of species' distributions. *Journal of Applied Ecology*, 43, 386–392.
- Guisan A, Broennimann O, Engler R, Vust M, Yoccoz NG, Lehmann A, Zimmermann NE (2006b) Using niche-based models to improve the sampling of rare species. *Conservation Biology*, 20, 501–511.
- Guisan A, Graham CH, Elith J, Huettmann F, NCEAS Species Distribution Modelling Group (2007) Sensitivity of predictive species distribution models to change in grain size. *Diversity and Distributions*, 13, 332–340.
- Hampe A (2004) Bioclimatic envelope models: what they detect and what they hide. *Global Ecology and Biogeography*, 13, 469–471.
- Hastie TJ, Tibshirani RJ (1990) *Generalized Additive Models*. Chapman and Hall, London.
- He HS, Mladenoff DJ, Crow TR (1999) Linking an ecosystem model and a landscape model to study forest species response to climate warming. *Ecological Modelling*, 114, 213–233.

- He HS, Mladenoff DJ, Gustafson EJ (2002) Study of landscape change under forest harvesting and climate warming-induced fire disturbance. *Forest Ecology and Management*, 155, 257–270.
- Heikkinen RK, Luoto M, Virkkala R (2006) Does seasonal fine-tuning of climatic variables improve the performance of bioclimatic envelope models for migratory birds? *Diversity and Distributions*, 12, 502–510.
- Hengl T, Gruber S, Shrestha DP (2004) Reduction of errors in digital terrain parameters used in soil-landscape modelling. *International Journal of Applied Earth Observation and Geoinformation*, 5, 97–112.
- Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25, 1965–1978.
- Hill JL, Curran PJ (2003) Area, shape and isolation of tropical forest fragments: effects of tree species diversity and implications for conservation. *Journal of Biogeography*, 30, 1391–1403.
- Hirzel A, Guisan A (2002) Which is the optimal strategy for habitat suitability modelling. *Ecological Modelling*, 157, 331–341.
- Hobbs RJ, Yates CJ (2003) Turner review No. 7. Impacts of ecosystem fragmentation on plant populations: generalising the idiosyncratic. *Australian Journal of Botany*, 51, 471–488.
- Honnay O, Hermy M, Coppin P (1999) Impact of habitat quality on forest plant species colonization. *Forest Ecology and Management*, 115, 157–170.
- Hörsch B (2003) Modelling the spatial distribution of montane and subalpine forests in the central Alps using digital elevation models. *Ecological Modelling*, 168, 267–282.

- Houlder DJ, Hutchinson MF, Nix HA, McMahon JP (2000). ANUCLIM User Guide, Version 5.1. Centre for Resource and Environmental Studies, Australian National University, Canberra.
- Hubbell SP (2001) The Unified Neutral Theory of Biodiversity and Biogeography. Princeton University Press.
- Hughes L, Cawsey EM, Westoby M (1996) Climatic range sizes of Eucalyptus species in relation to future climate change. *Global Ecology and Biogeography Letters*, 5, 23–29.
- Hughes L (2000) Biological consequences of global warming: is the signal already apparent? *Trends in Ecology and Evolution*, 15, 56–61.
- Hurlbert SH (1984). Pseudoreplication and the design of ecological field experiments. *Ecological Monographs*, 54, 187–211.
- Jiménez-Valverde A, Lobo JM, Hortal J (2008) Not as good as they seem: the importance of concepts in species distribution modelling. *Diversity and Distributions*, 14, 885–890.
- Kearney M, Porter WP (2004) Mapping the fundamental niche: physiology, climate and the distribution of a nocturnal lizard. *Ecology*, 85, 3119–3131.
- Kearney M (2006) Habitat, environment and niche: what are we modelling? *Oikos*, 115, 186–191.
- Keitt TH, Bjørnstad ON, Dixon PM, Citron-Pousty S (2002) Accounting for spatial pattern when modeling organism-environment interactions. *Ecography*, 25, 616–625.
- Kramer MG, Hansen AJ, Taper ML, Kissinger EJ (2001) Abiotic controls on long-term windthrow disturbance and temperate rain forest dynamics in southeast Alaska. *Ecology*, 82, 2749–2768.

- Lassueur T, Joost S, Randin CF (2006) Very high resolution digital elevation models: Do they improve models of plant species distribution? *Ecological Modelling*, 198, 139–153.
- Leathwick JR, Whitehead D, McLeod M (1996) Predicting changes in the composition of New Zealand's indigenous forests in response to global warming: a modelling approach. *Environmental Software*, 11, 81–90.
- Leathwick JR (1998) Are New Zealand's *Nothofagus* species in equilibrium with their environment? *Journal of Vegetation Science*, 9, 719–732.
- Leathwick JR, Austin MP (2001) Competitive interactions between tree species in New Zealand's old-growth indigenous forests. *Ecology*, 82, 2560–2573.
- Leathwick JR, Whitehead D (2001) Soil and atmospheric water deficits and the distribution of New Zealand's indigenous tree species. *Functional Ecology*, 15, 233–242.
- Leathwick JR, Rowe D, Richardson J, Elith J, Hastie T (2005) Using multivariate adaptive regression splines to predict the distributions of New Zealand's freshwater diadromous fish. *Freshwater Biology*, 50, 2034–2052.
- Legendre P (1993) Spatial autocorrelation: trouble or new paradigm? *Ecology*, 74, 1659–1673.
- Lehmann A, Overton JM, Leathwick JR (2003) GRASP: generalized regression analysis and spatial prediction. *Ecological Modelling*, 160, 165–183.
- Lemenih M, Gidyelew T, Teketay D (2004) Effects of canopy cover and understorey environment of tree plantations on richness, density and size of colonizing woody species in southern Ethiopia. *Forest Ecology and Management*, 194, 1–10.

- Levin SA, Muller-Landau HC, Nathan R, Chave J (2003) The ecology and evolution of seed dispersal: A theoretical perspective. *Annual Review of Ecology, Evolution, and Systematics*, 34, 575–604.
- Levine JM, Murrell DJ (2003) The community-level consequences of seed dispersal patterns. *Annual Review of Ecology, Evolution, and Systematics*, 34, 549–574.
- Lexer MJ, Hönniger K (2004) Effects of error in model input: experiments with a forest patch model. *Ecological Modelling*, 173, 159–176.
- Lindenmayer DB, Mackey BG, Mullen IC, McCarthy MA, Gill AM, Cunningham RB, Donnelly CF (1999) Factors affecting stand structure in forests – are there climatic and topographic determinants? *Forest Ecology Management*, 123, 55–63.
- Lindenmayer DB (2000) Factors at multiple scales affecting distribution patterns and their implications for animal conservation – Leadbeater’s Possum as a case study. *Biodiversity and Conservation*, 9, 15–35.
- Lindenmayer DB, Mackey BG, Cunningham RB, Donnelly CF, Mullen IC, McCarthy MA, Gill AM (2000) Factors affecting the presence of the cool temperate rain forest tree myrtle beech (*Nothofagus cunninghamii*) in southern Australia: integrating climatic, terrain and disturbance predictors of distribution patterns. *Journal of Biogeography*, 27, 1001–1009.
- Liu J, Ashton PS (1998) FORMOSAIC: an individual-based spatially explicit model for simulating forest dynamics in landscape mosaics. *Ecological Modelling*, 106, 177–200.
- Lobo JM, Jiménez-Valverde A, Real R (2008) AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, 17, 145–151.

- Loehle C, LeBlanc D (1996) Model-based assessments of climate change effects on forests: a critical review. *Ecological Modelling*, 90, 1–31.
- Lookingbill TR, Urban DL (2003) Spatial estimation of air temperature differences for landscape-scale studies in montane environments. *Agricultural and Forest Meteorology*, 114, 141–151.
- Lookingbill T, Urban D (2004) An empirical approach towards improved spatial estimates of soil moisture for vegetation analysis. *Landscape Ecology*, 19, 417–433.
- Lookingbill TR, Urban DL (2005) Gradient analysis, the next generation: towards more plant-relevant explanatory variables. *Canadian Journal of Forest Research*, 35, 1744–1753.
- Luoto M, Heikkinen RK (2008) Disregarding topographic heterogeneity biases species turnover assessments based on bioclimatic models. *Global Change Biology*, 14, 483–494.
- MacArthur RH, Wilson EO (1967) *The Theory of Island Biogeography*. Princeton University Press, New Jersey.
- MacNally R (2002) Multiple regression and inference in ecology and conservation biology: further comments on identifying important predictor variables. *Biodiversity and Conservation*, 11, 1397–1401.
- Margules CR, Pressey RL (2000) Systematic conservation planning. *Nature*, 405, 243–53.
- McGill B (2003) Strong and weak tests of macroecological theory. *Oikos*, 102, 679–685
- McVicar TR, Van Niel TG, Li L, Hutchinson MF, Mu X, Liu Z (2007) Spatially distributing monthly reference evapotranspiration and pan evaporation considering topographic influences. *Journal of Hydrology*, 338, 196–220.

- Miller J, Franklin J, Aspinall R (2007) Incorporating spatial dependence in predictive vegetation models. *Ecological Modelling*, 202, 225–242.
- Mills KG (1986) The Illawarra Rainforests: An Historical, Floristic, and Environmental Study of their Distribution and Ecology. Ph.D. Thesis. Department of Geography, The University of Wollongong, Wollongong.
- Moore ID, Norton TW, Williams JE (1993) Modelling environmental heterogeneity in forested landscapes. *Journal of Hydrology*, 150, 717–747.
- Moore KA, Elmendorf SC (2006) Propagule vs. niche limitation: untangling the mechanisms behind plant species' distributions. *Ecology Letters*, 9, 797–804.
- Mummery D, Battaglia M (2002) Data input quality and resolution effects on regional and local scale *Eucalyptus globulus* productivity predictions in north-east Tasmania. *Ecological Modelling*, 156, 13–25.
- NPWS (2002) Native Vegetation of the Illawarra Escarpment and Coastal Plain. NSW National Parks and Wildlife Service, Sydney.
- NPWS (2003) Illawarra Escarpment State Conservation Area Issues Paper: in preparation for a draft plan of management. NSW National Parks and Wildlife Service, Sydney.
- Offord CA, Porter CL, Meagher PF, Errington G (1999) Sexual reproduction and early plant growth of the Wollemi Pine (*Wollemia nobilis*), a rare and threatened Australian conifer. *Annals of Botany*, 84, 1–9.
- Olden JD, Poff NL, Bledsoe BP (2006) Incorporating ecological knowledge into ecoinformatics: An example of modeling hierarchically structured aquatic communities with neural networks. *Ecological Informatics*, 1, 33–42.

- Passioura JA, Ash JE (1993) Phenotypic, genetic, and ecological variation in the *Eucalyptus saligna*-*E. botryoides* complex. *Australian Journal of Botany*, 41, 393–412.
- Paul KI, Polglase PJ, Smethurst PJ, O’Connell AM, Carlyle CJ, Khanna PK (2004) Soil temperature under forests: a simple model for predicting soil temperature under a range of forest types. *Agricultural and Forest Meteorology*, 121, 167–182.
- Pearce J, Ferrier S (2000a) An evaluation of alternative algorithms for fitting species distribution models using logistic regression. *Ecological Modelling*, 128, 127–147.
- Pearce J, Ferrier S (2000b) Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological Modelling*, 133, 225–245.
- Pearson RG, Dawson TP, Berry PM, Harrison PA (2002) SPECIES: A spatial evaluation of climate impact on the envelope of species. *Ecological Modelling*, 154, 289–300.
- Pearson RG, Dawson TP (2003) Predicting the impacts of climate change on the distribution of species: are bioclimatic envelope models useful? *Global Ecology and Biogeography*, 12, 361–371.
- Pearson RG, Dawson TP (2004) Bioclimatic envelope models: what they detect and what they hide — response to Hampe (2004). *Global Ecology and Biogeography*, 13, 471–473.
- Pearson RG (2006) Climate change and the migration capacity of species. *Trends in Ecology and Evolution*, 21, 111–113.
- Peng YY, Dang Q (2003) Effects of soil temperature on biomass production and allocation in seedlings of four boreal tree species. *Forest Ecology and Management*, 180, 1–9.

- Peterson AT (2003) Predicting the geography of species' invasions via ecological niche modelling. *The Quarterly Review of Biology*, 78, 419–433.
- Peterson AT, Papeş M, Eaton M (2007) Transferability and model evaluation in ecological niche modeling: a comparison of GARP and Maxent. *Ecography*, 30, 550–560.
- Pfeffer K, Pebesma EJ, Burrough PA (2003) Mapping alpine vegetation using vegetation observations and topographic attributes. *Landscape Ecology*, 18, 759–776.
- Phillips SJ, Dudík M, Schapire RE (2004) A maximum entropy approach to Species Distribution Modelling. *Proceedings of the 21st International Conference on Machine Learning*.
- Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190, 231–259.
- Phillips SJ, Dudík M (2008) Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography*, 31, 161–175.
- Pierce Jr KB, Lookingbill T, Urban D (2005) A simple method for estimating potential relative radiation (PRR) for landscape-scale vegetation analysis. *Landscape Ecology*, 20, 137–147.
- Porté A, Huard F, Dreyfus P (2004) Microclimate beneath pine plantation, semi-mature pine plantation and mixed broadleaved-pine forest. *Agricultural and Forest Meteorology*, 126, 175–182.
- Price DT, McKenney DW, Nalder IA, Hutchinson MF, Kesteven JL (2000) A comparison of two statistical methods for spatial interpolation of Canadian monthly mean climate data. *Agricultural and Forest Meteorology*, 101, 81–94.

- Pulliam HR (1988) Sources, sinks and population regulation. *American Naturalist*, 132, 652–661.
- Pulliam HR (2000) On the relationship between niche and distribution. *Ecology Letters*, 3, 349–361.
- Randin CF, Dirnböck T, Dullinger S, Zimmermann NE, Zappa M, Guisan A (2006) Are niche-based species distribution models transferable in space? *Journal of Biogeography*, 33, 1689–1703.
- Raupach MR, Finnigan JJ (1997) The influence of topography on meteorological variables and surface-atmosphere interactions. *Journal of Hydrology*, 190, 182–213.
- Ridolfi L, D’Odorico P, Porporato A, Rodriguez-Iturbe I (2003) Stochastic soil moisture dynamics along a hillslope. *Journal of Hydrology*, 272, 264–275.
- Ritter E, Dalsgaard L, Einhorn KS (2005) Light, temperature and soil moisture regimes following gap formation in a semi-natural beech-dominated forest in Denmark. *Forest Ecology and Management*, 206, 15–33.
- Roberts DW (1996) Landscape vegetation modelling with vital attributes and fuzzy systems theory. *Ecological Modelling*, 90, 175–184.
- Rocha Corrêa LD, Fett-Neto AG (2004) Effects of temperature on adventitious root development in microcuttings of *Eucalyptus saligna* Smith and *Eucalyptus globulus* Labill. *Journal of Thermal Biology*, 29, 315–324.
- Ruel J-C, Pin D, Cooper K (2001) Windthrow in riparian buffer strips: effect of wind exposure, thinning and strip width. *Forest Ecology and Management*, 143, 105–113.
- Rushton SP, Ormerod SJ, Kerby G (2004) New paradigms for modelling species distributions? *Journal of Applied Ecology*, 41, 193–200.

- Ruxton GD, Saravia LA (1998) The need for biological realism in the updating of cellular automata models. *Ecological Modelling*, 107, 105–112.
- Schumacher S, Bugmann H, Mladenoff DJ (2004) Improving the formulation of tree growth and succession in a spatially explicit landscape model. *Ecological Modelling*, 180, 175–194.
- Schumacher S, Reineking B, Sibold J, Bugmann H (2006) Modeling the impact of climate and vegetation on fire regimes in mountain landscapes. *Landscape Ecology*, 21, 539–554.
- Scott JM, Heglund PJ, Morrison ML, Haufler JB, Raphael MG, Wall WA, Samson FB (2002) *Predicting species occurrences: issues of accuracy and scale*. Island Press, Washington.
- Segurado P, Araújo MB, Kunin WE (2006) Consequences of spatial autocorrelation for niche-based models. *Journal of Applied Ecology*, 43, 433–444.
- Soares-Filho BS, Cerqueira GC, Pennachin CL (2002) DINAMICA – a stochastic cellular automata model designed to simulate the landscape dynamics in an Amazonian colonization frontier. *Ecological Modelling*, 154, 217–235.
- Soberón J, Peterson AT (2005) Interpretation of models of fundamental ecological niches and species' distributional areas. *Biodiversity Informatics*, 2, 1–10.
- Stephens PA, Buskirk SW, del Rio CM (2007) Inference in ecology and evolution. *Trends in Ecology and Evolution*, 22, 192–197.
- Stockwell DRB, Peterson AT (2002) Effects of sample size on accuracy of species distribution models. *Ecological Modelling*, 148, 1–13.
- Svenning J-C, Skov F (2004) Limited filling of the potential range in European tree species. *Ecology Letters*, 7, 565–573.

- Swets JA (1988) Measuring the accuracy of diagnostic systems. *Science*, 240, 1285–1293.
- Thomas CD, Cameron A, Green RE, Bakkenes M, Beaumont LJ, Collingham YC, Erasmus BFN, de Siqueira MF, Grainger A, Hannah L, Hughes L, Huntley B, van Jaarsveld AS, Midgley GF, Miles L, Ortega-Huerta MA, Peterson AT, Phillips OL, Williams SE (2004) Extinction risk from climate change. *Nature*, 427, 145–148.
- Thuiller W (2004) Patterns and uncertainties of species' range shifts under climate change. *Global Change Biology*, 10, 2020–2027.
- Thuiller W, Araújo MB, Pearson RG, Whittaker RJ, Brotons L, Lavorel S (2004a) Uncertainty in predictions of extinction risk. *Nature*, 430, 6995.
- Thuiller W, Araújo MB, Lavorel S (2004b) Do we need land-cover data to model species distributions in Europe? *Journal of Biogeography*, 31, 353–361.
- Thuiller W, Lavorel S, Araújo MB, Sykes MT, Prentice IC (2005) Climate change threats to plant diversity in Europe. *Proceedings of the National Academy of Sciences of the United States of America*, 102, 8245–8250.
- Trivedi MR, Berry PM, Morecroft MD, Dawson TP (2008) Spatial scale affects bioclimatic model projections of climate change impacts on mountain plants. *Global Change Biology*, 14, 1089–1103.
- Uchida T, Ohya Y (1999) Numerical simulation of atmospheric flow over complex terrain. *Journal of Wind Engineering and Industrial Aerodynamics*, 81, 283–293.
- USDA Forest Service (2007) Solar Analyst – An ArcView® GIS extension for modelling solar radiation. <http://www.fs.fed.us/informs/tools.php#solar>. Cited 10th June 2007.

- Van Niel KP, Laffan SW, Lees BG (2004) Effect of error in the DEM on environmental variables for predictive vegetation modelling. *Journal of Vegetation Science*, 15, 747–756.
- Van Niel KP, Austin MP (2007) Predictive vegetation modeling for conservation: impact of error propagation from digital elevation data. *Ecological Applications*, 17, 266–280.
- Walther G-R, Post E, Convey P, Menzel A, Parmesan C, Beebee TJC, Fromentin J-M, Hoegh-Guldberg O, Bairlein F (2002) Ecological responses to recent climate change. *Nature*, 416, 389–395.
- Walther G-R, Berger S, Sykes MT (2005) An ecological ‘footprint’ of climate change. *Proceedings of the Royal Society of London Series B*, 272, 1427–1432.
- Wang H, Hall CAS, Scatena FN, Fetcher N, Wu W (2003) Modeling the spatial and temporal variability in climate and primary productivity across the Luquillo Mountains, Puerto Rico. *Forest Ecology and Management*, 179, 69–94.
- Wardell-Johnson GW, Williams JE, Hill KD, Cumming R (1997). Evolutionary biogeography and contemporary distribution of eucalypts. In: JE Williams, JCZ Woinarski (Editors), *Eucalypt Ecology*. Cambridge University press, Cambridge, UK, pp. 92–128.
- Weisberg PJ, Bugmann H (2003) Forest dynamics and ungulate herbivory: from leaf to landscape. *Forest Ecology and Management*, 181, 1–12.
- Weiss A, Hays CJ (2005) Calculating daily mean air temperatures by different methods: implications from a non-linear algorithm. *Agricultural and Forest Meteorology*, 128, 57–65.
- Wiens JA (2002) Predicting species occurrences: progress, problems, and prospects. In: JM Scott, PJ Heglund, ML Morrison, JB Haufler, MG Raphael, WA Wall, FB

- Samson (Editors), Predicting Species Occurrences: Issues of Accuracy and Scale. Island Press, Washington, pp. 739–749.
- Wintle BA, Elith J, Potts JM (2005) Fauna habitat modelling and mapping: a review and case study in the lower Hunter Central Coast region of NSW. *Austral Ecology*, 30, 719–738.
- Wintle BA, Bardos DC (2006) Modeling species-habitat relationships with spatially autocorrelated observation data. *Ecological Applications*, 16, 1945–1958.
- Wise S (2000) Assessing the quality for hydrological applications of digital elevation models derived from contours. *Hydrological Processes*, 14, 1909–1929.
- Wolfe JN (1945) The use of weather bureau data in ecological studies. *The Ohio Journal of Science*, 45, 1–12.
- Wu J, Hobbs R (2002) Key issues and research priorities in landscape ecology: An idiosyncratic synthesis. *Landscape Ecology*, 17, 355–365.
- Wu J (2006) Landscape ecology, cross-disciplinarity, and sustainability science. *Landscape Ecology*, 21, 1–4.
- Wunder J, Bigler C, Reineking B, Fahse L, Bugmann H (2006) Optimisation of tree mortality models based on growth patterns. *Ecological Modelling*, 197, 196–206.
- Zaniewski AE, Lehmann A, Overton JMcC (2002) Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. *Ecological Modelling*, 157, 261–280.