# Balancing the Costs and Benefits of Increasing Information in Ecological Models

Thesis submitted by

**Collin James Storlie**

BSc (Washington State University)

GDipResMeth (James Cook University)

For the degree of Doctor of Philosophy (Terrestrial Ecology)

in the College of Marine & Environmental Sciences,

James Cook University, Townsville, Australia

November 2014

## Acknowledgements

My supervisor, Jeremy VanDerWal is an absolute champion. He showed me how to be a better writer, how to be a better modeller, and even how to be a better person. Most importantly, when things got tough, Jeremy didn't give up on me.

Ben Phillips deserves a huge thank you for teaching me how to write concisely, answering my math related questions, and for listening when I had things on my mind.

Justin Welbergen provided me with key financial assistance and a keen eye for manuscripts.

The support of Stephen Williams in establishing this project and providing my scholarship was critical to my success. Everything I know about ecology, I know because Steve gave me a chance. He has been an excellent mentor and a great friend for nearly ten years, and for this I can't thank him enough.

The support of Yvette Williams was also critical to my completion. She listened to my problems and helped me to believe in myself and in what I could achieve. From the bottom of my heart, thank you.

Many friends helped over the years simply by being themselves: Arnaud Gourret, Scott Parsons, Joseph & Samantha Clarey, Rohan Wilson, Luke Shoo, Andres Merino-Viteri, Marios Aristophanous, Marcus McGrath, Jessica Cheok, Michael Ellison, Lorenzo Fattori, Jaime Thompson, Nicolas VonAlvenSleben, Kate Stookey, Kate Nairn, Sam Holland, April Reside, Brett Scheffers, Nadiah Roslan, Lauren Hodgson, and many more.

Lastly, my family, without your love none of this would be possible.

## Statement of Contribution of Others

Chapters 1 and 6 of this thesis were conceived and written by me (referred to as CS in this Statement), with editorial assistance from my supervisors, Associate Professor Jeremy J. VanDerWal and Dr. Ben L. Phillips.

Chapter 2 of this thesis has been published in collaboration with my supervisors, Associate Professor Jeremy J. VanDerWal (herein JJV), Professor Stephen E. Williams (herein SEW) and Dr. Ben L. Phillips (herein BP). CS and JJV conceived the ideas for the manuscript. CS, SEW and numerous others collected the data. I led the manuscript writing and completed the analyses with technical support and contributions from JJV and BP.

Chapter 3 of this thesis has been published in collaboration with my supervisors, JJV, SEW, BP and Dr. Justin Welbergen (herein JW). Data critical to this chapter, as well as revisions on the manuscript were also provided by Andres Merino-Viteri (herein AMV). JJV, BP and JW provided assistance with statistical analyses and manuscript revision. SEW provided data on vertebrate localities. AMV provided data on Critical Thermal Maxima for the study species and assistance with manuscript revision.

Chapter 4 of this thesis has been prepared for publication in collaboration with supervisors JJV, BP and SEW. I conceived the ideas for the manuscript. JJV and BP rendered assistance with statistical models, analysis, and manuscript revision. SEW provided data on vertebrate localities.

Chapter 5 of this thesis has been prepared for publication in collaboration with supervisors JJV and an external associate, Dr. Luke Shoo (herein LS). Additional assistance with revisions to the manuscript and statistical analyses were provided by LS and JJV.

I was responsible for the maintenance of climate dataloggers, as well as storing, cleaning, and aggregating data for use in the statistical downscale analysis of Chapter 2. Further empirical weather data were obtained from the Autralian Bureau of Meteorology. Surveys for the vertebrate occurrence and count data were also conducted

by myself; supplemental data was provided by Stephen Williams and the Centre for Tropical Biodiversity and Climate Change. Spatial layers of temperature and rainfall were sourced from the Australian Water Availability Project.

Financial support was provided by the National Environmental Research Program (NERP) *Tropical Ecosystems* Hub and *Environmental Decisions* Hub, the Marine and Tropical Sciences Research Facility (MTSRF), James Cook University and the Skyrail Foundation.

Computational and hardware support was provided by the High Performance Research Computing Unit at James Cook University, Townsville.

## Abstract

Ecology transitioned from observational studies to experimental studies and hypothesis testing, and is now transitioning back again; this reversal is largely due to technological advances in data collection, storage and computation that have enabled mining disparate sources of data to explore broad ecological relationships and theories. While mining these disparate sources of data has facilitated whole new fields of ecology and better understanding of ecological processes, there is a tendency to assume advanced analytics with complex data yields better results or better understanding; this is not always the case. As models and analysis become more complex so do the underlying assumptions, but the increased complexity may not be necessary. Herein, this thesis explores the value of mining disparate data sources, and of increasing model and data complexity, for exploring species-environment relationships (SERs) in ecology.

The spatially explicit models underlying exploration of the SERs typically rely on linking attributes of a species to coarsely interpolated and temporally aggregated information such as 'climate'. Climate is typically a 30 or 50 year average (as opposed to 'weather', which is more temporally discrete, e.g. daily maximum temperature), and spatially explicit estimates of climate and weather are typically interpolated between known locations based on latitude, longitude and elevation. Such climate and weather estimates represent spatial data which are naïve to the importance of key factors (e.g. topography and vegetation) that structure thermal regimes at fine scales. Further, such climate and weather surfaces may be biased in a non-random fashion as a result of estimating the environment at fine scale without reference to certain biotic and abiotic factors. Hence, gridded climate and weather data are often poor predictors of the true environmental conditions to which species are exposed – but how much does this matter in exploring spatiotemporal patterns in species distribution and abundance, and how these SERs may change?

Species tend to experience the environment at very local scales of time and space, thus a major flaw in spatially explicit ecological studies may be temporally aggregated, inaccurate, or spatially biased environmental data. SERs based on spatial environmental layers with any of the above problems will be biased, potentially leading

to false inference of how the species interacts with the abiotic environment, and how this in-turn structures the species distribution. Herein, I focus on improving the accuracy of spatial weather and climate layers and proceed to quantify the value and need to improve these estimates using several algorithms of increasing complexity to estimate SERs. I demonstrate increased concordance of model outcomes with ecological niche theory when using accurate spatial data and increased utility for exploring new relationships.

Underlying the entire thesis is a statistical downscaling of broad-scale weather layers for the Wet Tropics Bioregion of north east Queensland. I statistically downscale 30 years of existing spatial weather estimates against empirical weather data and spatial layers of topography and vegetation to produce highly accurate spatial layers of daily weather. The downscaled weather layers are more accurate with respect to empirically measured temperature, particularly for maximum temperature, when compared to current best-practice weather layers. Current best-practice climate layers are least accurate in heavily forested upland regions, frequently over-predicting empirical mean maximum temperature by as much as 7°C. This thesis examines the value of the extra effort, complexity and assumptions required to produce these data with respect to SERs.

Correlative Species Distribution Models (SDMs) combined with spatial layers of climate and species' localities represent a frequently utilised and rapid method for imputing relationships between a species and its environment, as well as generating spatial estimates of species distributions. However, an SDM is only as accurate as the inputs upon which it is based – garbage in, garbage out. Using current best-practice climate data and my improved climate data, I proceed to demonstrate the effect of inaccurately quantified spatial data on SDM outcomes for a group of seven rainforest skinks. Generally, the distributions of the focal species are not visibly different (at a coarse scale) but the predictions generated using the improved climate layers are more fragmented and contain less core distributional area.

To assess a species' vulnerability to climate change, we commonly use mapped environmental data that are coarsely resolved in time and space. Coarsely-resolved temperature data are typically inaccurate at predicting temperatures in microhabitats used by an organism and may also exhibit spatial bias in topographically complex areas. As a result, simple correlations between where a species occurs and mapped

environmental data may predict thermal regimes at a site that exceed species' known thermal limits. In this study, I use statistical downscaling to account for environmental and behavioural factors to develop high-resolution estimates of daily maximum temperatures for the preferred diurnal shelter of a group of rainforest frogs (Family: Microhylidae). I then demonstrate that this statistical downscaling provides temperature estimates that consistently place focal species within their fundamental thermal niche, whereas coarsely resolved layers do not. These results highlight the need for incorporation of fine-scale weather data into species vulnerability analyses, and demonstrate that statistical downscaling approaches are valuable for yielding biologically relevant estimates of thermal regimes.

Methods to predict spatially explicit patterns of species abundance are numerous in form. The most accurate techniques account for variable detection rates, so that we can separate detection from our estimate of abundance. While elegant, these detection models require large presence-absence datasets, derived from repeated surveys across temporal and geographic gradients. In many cases, however, the data are simply not available for these statistical approaches. In these cases, detection-invariant models, which do not require repeated survey effort, represent an alternative. Importantly, if detection rates are unaffected by the predictor variables, then these detection-invariant approaches may yield just as useful a measure of abundance as the more data-intensive models. Thus, by avoiding the use of predictor variables that likely affect detectability, some of the pitfalls of detection-invariant methods can be avoided. To test this, I model the abundance patterns of a group of rainforest skinks using two techniques: occupancy modelling, which accounts for variable detection rate, and a commonly-used presence-only approach (MaxEnt) which does not. I verify the veracity of model outputs against a large dataset of surveys for skink abundance at 200+ sites over 10 years of time. I find that variable detection models and detection invariant models correlate well with carrying capacity across a number of sites, although variable detection models consistently predict abundance with greater accuracy. This result indicates that detection-invariant models, such as MaxEnt, are not as good as variable detection models but in the absence of repeat survey data, they can come close to the accuracy of a variable detection model. As such, they are still useful for the majority of cases when we require rapid assessment of species abundance patterns in the absence of more robust datasets.

Spatial layers of the weather have applications beyond SDMs and in this section I leverage information from the statistical downscaling of weather maps to demonstrate the effects of vegetation clearance on thermal regimes. The impacts of deforestation are typically measured in terms of habitat: hectares lost, altered habitat fragmentation or connectivity. However, altered habitat extent is just one component of change stemming from vegetation clearance. Climatic conditions too are regulated by vegetation and so are liable to change as well. Vegetation buffers habitats from extreme climate and weather conditions, which are predicted to increase in frequency under global warming scenarios. Despite this, we know surprisingly little about the indirect legacy of deforestation on accelerating the loss of extant climates (and dependent species) projected to 'disappear' under climate change. Here I describe the legacy of deforestation on climatic availability in the Australian Wet Tropics by integrating spatial information on vegetation and weather to quantify 30 years of weather patterns under two alternative scenarios of vegetation extent: prior to European Settlement (ca. 1750) and current (1976-2005). I find that deforestation has on average increased region-wide maximum temperatures by 0.67°C with larger increases in localised areas subjected to more extensive deforestation (0.86-0.90°C). I also show that these modest climate shifts can be underpinned by dramatic reductions in the available area of particular thermal regimes including important cool environments projected to become increasingly scarce under climate change. Moreover, I demonstrate that thermal environments are more fragmented and less connected as a result of deforestation. Finally, I consider the potential for targeted reinstatement of vegetation to reduce range losses and buy time for adaptation to further climate change.

As data sources describing the environment and species localities proliferate, we are left asking what value these data lend to ecological analyses. Observational studies and statistical methods have developed to accommodate ever larger datasets, often assuming that more data will produce better results. The results of this thesis demonstrate that simpler models, with less restrictive datasets and assumptions can utilise large pools of data to form accurate predictions. However, the utility of data sources still needs to be address before they are applied. My research shows that inaccurate or spatially biased environmental data can lead to false inference of SERs, altered patterns of predicted spatial distribution, and a lack of concordance with ecological theory. However, in the process of tailoring these spatial data to suit a variety of ecological analyses, I have

further improved our understanding of the interplay between vegetation and the environment. Overall, these results indicate spatially biased climate and weather layers can be corrected with statistical downscaling techniques which explicitly consider abiotic and biotic factors that influence local processes. Downscaled layers meet both statistical (predicting empirical temperatures) and biological (concordance with species thermal limits) criterions of accuracy. Further, downscaling allows for an explicit understanding of how vegetation influences exposure, and the role of forest clearing in shifting thermal regimes.

## Contents

## List of Tables

## List of Figures

## List of Acronyms Used in this Thesis

**AU** .....Atherton uplands

**AWAP**.....Australian Water Availability Project

**AWT** .....Australian Wet Tropics

**BRT**.....Boosted regression tree

**CCL** .....Cairns-Cardwell lowlands

**cSDM** .....Correlative species distribution models

**FPC** .....Foliage projected cover

**GBIF** .....Global Biodiversity Information Facility

**mSDM**.....Mechanistic species distribution models

**MMTWM**.....Mean maximum temperature of the warmest month

**OLS**.....Ordinary least-squares (regression)

**RMSE** .....Root mean squared error

**SD**.....Species distribution

**SDM**.....Species distribution model

**SER**.....Species-environmental response

# Chapter 1: General Introduction

## The Importance and Roles of Biological Diversity and Ecosystems

Biological organisms and the complex ecological systems they form are recognised for possessing cultural, economic, service, and intrinsic values. Humanity has long recognised the cultural significance of many species and ecosystems, but it is humanity's growing knowledge of ecological systems that has shone light on the key services they provide. The properties of ecosystems and the services which they provide seem idiosyncratic, however they depend greatly upon the composition of its constituent species. Changes in species composition resulting from extinction, invasion, harvesting, or other mechanisms will alter ecosystems in ways which we may not be able to predict or understand. Ultimately, dynamic ecosystem patterns may flow-on to affect nutrient cycles, water cycles, soil deposition, erosion, and climate to name only a few key examples. These changes then affect our ability to extract key resources and services from the ecosystems. Ecological systems are entirely responsible for the abundance of breathable air, potable water, and food resources upon which humans rely, yet many ecosystems and species are threatened by numerous processes. As human populations continue to expand into the 22$^{nd}$ century it is vital that we identify threats to species and ecosystems, and develop methods to minimise threats and maintain sustainable ecosystems which continue to provide the vital resources which humanity needs to thrive.

## Threats to Biological Diversity

In particular, global climate change has been identified as a severe threat to biological systems across levels of organisation and spatiotemporal scales (IPCC 2014). Anthropogenic climate change has already been implicated in shifting patterns of distribution and phenology for many taxonomically distinct species (Parmesan & Yohe 2003, Parmesan 2006). Climate change will not only influence mean temperature, but also rainfall patterns, frequency of extreme events and other climate factors as well (IPCC 2014). The complexities of climatic change, and the mechanisms which drive it, make it difficult to predict outcomes for species with a degree of certainty. Hence,

methods which can measure the impact of multiple climate aspects on species and their distributions will be of great importance for determining how species will respond to a warming world.

Land-use change and deforestation are also human induced phenomena which threaten biological systems. Deforestation reduces habitat available for species, also frequently resulting in increased edge effects (Kapos 1989), levels of fragmentation (Laurance 2000), and loss of habitat or population connectivity (Dixo *et al*. 2009). Vegetation plays a key role in structuring patterns of thermal regimes across a landscape (Ashcroft 2006, Deo *et al*. 2009, Shoo *et al*. 2011, Storlie *et al*. 2013, Chapter 2); dense vegetation acts as an insulator, protecting species at ground level from exposure to temperature extremes (Scheffers *et al*. 2013a). At a local scale then, the effects of deforestation on thermal regimes will likely mirror those predicted by anthropogenic climate change. The frequency and intensity of weather events flow on to affect species survival rates (Welbergen *et al*. 2008) and seasonal distribution patterns (Reside *et al*. 2010), to name only a few potential impacts. Furthermore, deforestation releases carbon to the atmosphere and reduces the amount of vegetation available to absorb and store further carbon. In this way too, deforestation may act to further exacerbate the impacts of increased atmospheric carbon dioxide concentrations on climate and weather patterns. As human populations grow, demand for resources and land for agriculture will too, indicating that deforestation will continue to threaten biodiversity into the foreseeable future.

These processes and others threaten biodiversity and the ecological systems upon which we rely. Climate change is particularly insidious, since its effects will be felt across conservation boundaries, and can only be mitigated with united international effort. Considering the scale of threat, and pace at which new threats emerge; techniques to rapidly and accurately assess species response to threat are critical.

## Metrics and Frameworks for Assessing Vulnerability

The IUCN (2012) uses a number of metrics to define species' vulnerability; distributional area, habitat fragmentation, connectivity, or population abundance may be used. Singular, temporally discrete estimates of these metrics are rarely enough information to appropriately classify a species risk of extinction. Risk to threat is best

classified by combining a number of these metrics together across time and therefore obtaining a dynamic estimate of change across time. Threatening processes for biodiversity are unlikely to abate in the near term, and their potential to affect species must be documented. Therefore techniques which can be projected into the future to estimate one or more of these key population metrics will be essential for appropriately classifying species response to threat.

Traditional vulnerability assessments for species are often described as being a function of 'sensitivity' (the intrinsic properties of a species) and 'exposure' (the environmental conditions which a species experiences *in situ*) (Williams *et al.* 2008). Spatial environmental layers are frequently utilised as sources of exposure data for these vulnerability analyses (Storlie *et al*. 2014, Chapter 3). Unfortunately, *a-priori* assessment of layer accuracy and inclusion of species behavioural information which dictates exposure to any particular environmental aspect are often lacking (Storlie *et al* 2013, Chapter 2, Storlie *et al*. 2014, Chapter 3). When exposure data in vulnerability assessments do not represent or are poorly correlated with species exposure, the accuracy of the assessment must be drawn into question. One possible consequence of such a data mismatch could be species which are predicted to thrive outside their empirically determined fundamental thermal niche.

## Technology, Data, and Scientific Enquiry

The foundations of scientific enquiry are built upon observational studies. However, in the 20$^{th}$ century the emphasis shifted from observation to experimentation and hypothesis testing. Now a new shift is underway, emphasising the role of observational studies in determining ecological processes from observed patterns (Sagarin and Pauchard 2010). Although not a panacea, observational studies have several distinct advantages over those which are based on manipulative experiments.

Sources of data (e.g. species occurrences) for observational studies do not need to be collected inside a rigorous experimental design framework. So called 'citizen science' derived datasets have been able to document ecological and environmental processes in ways which experimental science cannot (e.g. Sagarin and Micheli 2001, Royle and Nichols 2005). Further, advances in GIS and satellite imagery have provided numerous data for diverse ecological analyses across a range of temporal and spatial resolutions.

Moreover, technology has increased our capacity to store data, resulting in large freely available datasets of spatial environmental layers (e.g. WORLDCLIM – Hijmans *et al*. 2005, AWAP – Jones *et al*. 2009,) and species occurrences (e.g. GBIF – Flemons *et al*. 2007). The new paradigm of observational studies is largely driven by technology, specifically the relationship between technology and data.

The development of statistical methods which combine these data sources (occurrence records and spatial layers of the environment) – examining pattern to estimate ecological process – has grown as a direct result of technology's influence over the availability of data. Statistical algorithms to cope with large datasets typically exhibit fewer constraints than more elegant models, which require smaller datasets with greater structural requirements. Simpler models may offer advantages over more complex ones, including their low cost and rapid parameterisation (owing to already available data). Frequently, however, simple algorithms with large datasets outperform complex algorithms with smaller datasets (Domingos 2012). Herein I will explore the relative advantages of simple models with large non-specific datasets in comparison to more complex modelling approaches with smaller datasets. By doing so, I hope to improve inference of ecological process derived from observational techniques, and more clearly define the appropriate use of large environmental and species datasets in ecological analyses.

## Species Distribution Modelling and Predicting Abundance

Species distribution models comprise a number of techniques which aim to predict the geographic pattern of species occupancy and can be divided into two main groups: correlative and process-based (Kearney 2006). Correlative models (hereafter, cSDMs) employ algorithms to identify associations between mapped environmental data and known locations of species presence (or absence). This effectively produces an estimate of the realised niche, which is then mapped back onto environmental layers to produce an estimate of occupancy in geographic space (Kearney 2006, Elith & Leathwick 2009). Mechanistic or process-based models (hereafter, mSDMs) relate environmental data to physiological rates of species, solving complex mass/energy balance equations to determine geographic areas suitable for the species (Kearney & Porter 2009, Kearney *et al*. 2009). Estimates of physiological rates for mechanistic models are combined to yield an approximation of the fundamental niche, since they are created without

reference to the species true occurrences and interactions with other species – the realised niche (Kearney & Porter 2009).

Both families of SDM techniques are underpinned by ecological niche theory, and have individual strengths and weaknesses. Extrapolation into novel conditions tends to reduce the accuracy of cSDMs more than mSDMs (Kearney & Porter 2009). Further, correlative models may be fit using environmental data which have little or no mechanistic impact on the species (Austin 2002). Therefore, variables important for model fitting may not be responsible for structuring the species distribution, merely correlated with it. Process-based models allow for a stronger understanding of mechanism on distribution. In doing so, they extrapolate better into novel spatial and temporal conditions, and may be better suited for forming climate change predictions (Kearney & Porter 2009). However, process-based models are based on an understanding of species physiological rates, which can be difficult to obtain without extensive investment of time and resources.

Correlative SDMs take many forms ranging from simple 'climate envelope' approaches like BIOCLIM, to commonly utilised generalised linear and additive models (GLMs and GAMs), variable detection rate logistic regression models, presence-only entropy-based techniques (MaxEnt), or complex machine-learning algorithms like Boosted Regression Trees and Artificial Neural Networks (Elith & Leathwick 2009). Each specific technique has unique algorithms, input data, model output, and assumptions (Elith *et al*. 2006). Many of the assumptions implicit in these SDMs are shared between techniques (Wiens *et al* 2009). SDMs assume that the complete extent of habitat suitable for the species will be occupied, in other words, that the species distribution has reached equilibrium (Austin 2007). Further, spatial layers of the environment utilised are assumed to relate to the fitness of individuals within the population, if only indirectly (Austin 2002). Hence, selection of environmental layers is an important undertaking at the beginning of any modelling procedure. Also, environmental factors not included in the model are assumed to not affect individual fitness. This assumption implies that interspecific interactions have no effect on species distribution (Guisan *et al*. 2006). When projecting models into the future, a key assumption of these techniques is niche conservatism (Wiens & Graham 2005), that the relationship between species and environment will not change either across time or space (Wiens *et al*. 2009).

Stated simply, these models assume all individuals within a population are the same regardless of demographic class or phenotype, and that evolution does not occur within the population over time. Finally, cSDMs often assume that detection rate for species is invariable and equal to 1 (MacKenzie *et al*. 2002); in other words species detection is not affected by the environment and species which are present will be detected. Despite these many assumptions, cSDMs have a wide variety of applications for ecological analyses, and can often be trained with relatively small datasets.

Many varieties of cSDM can be trained using presence-only data, which is typically more readily available (and simpler to gather) than presence-absence datasets. Gridded data-sets of the environment are now also widely available; particularly layers describing climate averages and vegetation across a range of spatial scales, which are frequently utilised in cSDMs. The differences between variable detection and detection invariant cSDMs can be important when attempting to estimate abundance. Since abundance and detection are often correlated (but not always, some species are abundant and cryptic), these two metrics can be difficult to untangle (MacKenzie *et al*. 2002, Royle & Nichols 2003). Adding further complications are the frequent correlations between detection rates and the environment, sampling technique, temporal phase, and/or observer (Royle *et al*. 2005). Models which utilise repeated presence-absence surveys to disentangle the influence of the environment on detectability may therefore generate more robust estimates of abundance (MacKenzie *et al*. 2002). However, long-term presence-absence datasets covering appropriate environmental and temporal gradients are expensive and time-consuming to procure. When rapid assessment of species distribution or abundance is required often our only option will be to utilise a detection invariant modelling technique based on presence-only data, for instance, MaxEnt (e.g. VanDerWal *et al.* 2009c).

There exists a paradigm in modelling literature ‘garbage in, garbage out’, which effectively describes that any model (e.g. an SDM) is only as accurate as the data upon which it is based. Many sources of error in cSDMs have been explored (see Elith *et al*. 2006, VanDerWal *et al*. 2009b, Gogol-Prokurat 2011, and Anderson & Gonzalez 2011) however few have examined the effects of inaccurate spatial layers on model outcomes. This will be a key emphasis of the following thesis. Spatial climate estimates, which are utilised in both cSDMs and mSDMs, are often inaccurate at fine-grain sizes in areas

with high topographic complexity and multiple vegetation types (Shoo *et al*. 2011, Scherrer *et al*. 2011, Storlie *et al.* 2013*,* Chapter 2) Further, climate estimates are frequently biased, and can vary as a function of the environment. Accurate, non-biased spatial data are necessary because these are being used in conjunction with location data to classify the species niche (Austin 2007). Misclassification of the niche will result in inaccurate inference regarding both habitat utilisation and geographic range of the species (Storlie *et al*. 2013, Chapter 2, Storlie *et al*. 2014, Chapter 3). Statistical downscaling techniques to improve the accuracy and resolution of spatial climate and weather layers have the capacity to improve SDM model outcomes.

## The Role of Climate Downscaling: Reducing Bias and Understanding the Influence of Vegetation

Understanding how current best-practice weather and climate layers are developed is key if we are to understand how to improve them. Typically, average monthly climate conditions are aggregated at a number of sites across the extent which is under consideration. Then thin-plate splines estimate the climate across a grid based on latitude, longitude, and elevation (McMahon *et al*. 1995, Jones *et al* 2009, Xu & Hutchinson 2011). Such splines are accurate at coarse spatial resolution (5km or greater), but densely saturating the environment with weather stations does allow to increase resolution to 250m or finer (e.g. ANUCLIM). While these techniques often achieve accurate climate estimates at coarse resolution in topographically homogeneous environments, they can be inaccurate and spatially biased at fine resolution in topographically complex environments (Loarie *et al.* 2009).

By incorporating further empirical records and finely resolved environmental layers known to decouple local from regional climates with extant climate or weather layers in a statistical downscaling framework, it is possible to improve their accuracy and spatial resolution. Multiple factors decouple regional from local climates, these include but are not limited to: slope, aspect, vegetation cover, distance to coast, and distance to stream (Shoo *et al*. 2011, Dobrowski 2011). Vegetation in particular can have profound influence on the thermal regimes of daily maximum temperatures (Shoo et al. 2011, Storlie *et al*. 2013, Chapter 2). Downscaled environmental layers which are based upon knowledge of vegetation cover provide an opportunity to examine the influence of

deforestation on thermal regimes to which species are exposed, and further enhance vulnerability assessments, SDMs, and targeted restoration activities.

Statistical downscaling techniques may be as simple as linear regression models (Schoof & Pryor 2001), or utilise more complex methods such as Artificial Neural Networks or Boosted Regression Trees (e.g. Poteau *et al*. 2011). Local and regional climate systems are complex and driven by a variety of environmental and biological factors working individually and in concert (Dobrowski 2011). Given this complexity, straight forward statistical methods such as linear regression are less accurate when statistically downscaling regional to local climate (e.g. Schoof & Pryor 2001). Boosted Regression Trees (BRTs) demonstrate great promise for this particular task, as they combine classic Classification and Regression Trees with a gradient boosting algorithm. Within the BRT framework, a single Regression Tree is used to relate mapped variables to empirical temperature records (Elith *et al*. 2008). Residual variance will usually remain after fitting a single tree, and further trees are fit to these residual values in an iterative fashion until the remaining error reaches some predefined stopping point. Importantly, BRTs can model interactions between multiple variables, select only those which explain the most variance, and ignore those which don't improve the model (Elith *et al*. 2008). Gradient Boosting allows BRTs to fit complex, non-linear relationships between predictor and explanatory variables, thus improving model fit over other simpler techniques (Elith *et al*. 2008).

## Summary of Thesis Aims

Herein, I will demonstrate that statistical downscaling techniques can be used to improve the accuracy of, and reduce bias associated with, spatial climate and weather layers. I proceed to show that biased climate layers can lead to the development of inaccurate SERs and patterns of spatial distribution – visible only in detail, not at coarse resolution. Further statistical downscaling of weather layers by including key behavioural information only increases their biological relevance; this insures measures of sensitivity and exposure for vulnerability analyses are concordant with one another and ecological niche theory. Further, I examine the use of spatial climate layers and SDMs to predict carrying capacity of a diverse group of species. Lastly, I leverage information concerning the impact of vegetation on thermal regimes from the statistical downscaling procedure to demonstrate the impact of deforestation on climate space.

## Summary of Thesis Chapters

### *Chapter 2: Improved Spatial Estimates of Climate Predict Patchier Species Distributions*

In this chapter, I aim to improve spatial estimates of climate and weather using statistical downscaling techniques. Spatial climate estimates are a key resource for SDM, which are available at global extent with differing levels of spatial and temporal resolution. However, the spatial resolution at which these models are currently available is quite coarse. This is problematic for two reasons: the climate layers are too large to represent the operative scale of most species and they are limited in their ability to represent temperatures experienced by species *in situ*. I proceed to downscale current best practice weather layers at 5km resolution against empirical temperature records and 250m resolution layers of environmental factors known to decouple local and regional climates. Statistical criteria reveal improved accuracy of downscaled layers, which are then incorporated into cSDM models for seven species of skink. Finally, I compare SDMs based on downscaled and best-practice climate layers, and demonstrate that downscaled layers predict patchier distributions, of approximately equal distributional area, for all species modelled.

> **Publication: Storlie C.J.**, B.L. Phillips, J.J. VanDerWal, and S.E. Williams. 2013. Improved spatial estimates of climate predict patchier species distributions. *Diversity and Distributions*, **9**:1106-1113.

### *Chapter 3: Stepping Inside the Niche: Microclimate Data are Critical for Accurate Assessment of Species Vulnerability to Climate Change*

In this chapter, I build on the outputs of Chapter 2 to further improve weather layers via the addition of species behavioural information. The utility of spatial weather layers to act as exposure surrogates for vulnerability analyses is directly dependent upon their ability to re-create the environment which species experience *in situ*. Many species utilise a variety of microhabitats across seasons and days, and as such their exposure to the environment will change accordingly. Here, I incorporate behavioural information of seven species of Microhylid frog to recreate their thermal exposure to diurnal maximum temperatures. To do this, I downscale the improved weather layers from

Chapter 2 against empirically recorded temperatures of these species preferred diurnal shelter (underneath logs in rainforest) using a simple linear model. Then a biological criteria, in this case the species critical thermal maxima, is used to assess layer accuracy. I demonstrate that non-downscaled layers indicate that these species thrive outside of their fundamental thermal niche, an impossible situation in nature. The downscaled layers, however, give results concordant with the species fundamental niche.

> **Publication: Storlie, C.J.**, A. Merino-Viteri, B.L. Phillips, J.J. VanDerWal, J. Welbergen, and S.E. Williams. 2014. Stepping inside the niche: microclimate data are critical for accurate assessment of species' vulnerability to climate change. *Biology Letters,* **10**(9):20140576.

### *Chapter 4: Can Less Be More? Comparing Predictions of Species Abundance Using Presence-Only and Presence-Absence Datasets.*

Here, I continue to build upon the results of Chapter 2, and apply improved spatial climate layers in two distinct modelling frameworks to estimate site specific carrying capacity for a group of rainforest skinks. Population abundance (and its rate of change) has long been recognised as a correlate for extinction risk of species. However, models which hope to estimate abundance across space require extensive presence-absence datasets across geographic and temporal gradients. These datasets can be expensive and time-consuming to gather, and may not be available when rapid assessment of species vulnerability is needed. With this in mind, I compare the effectiveness of presence-absence and presence-only modelling techniques to predict empirically measured abundance of seven species of rainforest skink. I demonstrate that both techniques produce accurate and nearly equivalent estimates of population carrying capacity.

> **Publication: Storlie, C.J.,** J.J. VanDerWal, and B.L. Phillips**.** In Prep. Can Less Be More? Comparing Predictions of Species Abundance Using Presence-Only and Presence-Absence Datasets, *Ecological Modelling*

### *Chapter 5: The Legacy of Past Land Clearing on Climate Space: A Primer for Global Warming.*

Here I apply the knowledge gained concerning the impact of vegetation cover on local climate and weather (from Chapter 2) to demonstrate the ecological effects of historic land-clearing. Within the Australian Wet Tropics, approximately 20% of rainforest areas have been cleared since European settlement. Nearly 90% of the cleared forest is concentrated in two subregions, including one identified as a potential climate change refugia for the region's endemic fauna. By combining past estimates of rainforest cover with current maps of vegetation density, I create a simulation of vegetation density before European settlement (c. 1750) when large scale land clearing began. When substituted into the statistical downscale procedure from Chapter 2, I recreate 30 years of daily temperature maps for the region under this pre-clear vegetation scenario. These maps indicate that forest clearing has caused an appreciable amount of warming since 1750, a shift nearly on par with that of global warming in the last century. Moreover, thermal environments are less regular in shape and are now less connected as a result of deforestation. Targeted restoration of forest habitat has the capacity to improve thermal buffering in upland areas and connectivity within the region, but because of the scale of land-clearing, cannot completely ameliorate its impact on thermal regimes.

> **Publication: Storlie, C.J.,** J.J. VanDerWal, and L.P. Shoo. In Prep. The Legacy of Past Land Clearing on Climate Space: A Primer for Global Warming. *Conservation Letters.*

### *Chapter 6: General Discussion and Synthesis*

Lastly, I proceed to integrate the findings of the previous four chapters and discuss the costs and benefits of algorithm complexity in relation to SDM analyses. The role of spatial bias in data layers and how this flows on to alter model outcomes is considered. Further the necessity of correcting spatial bias is considered in light of the purpose of the modelling exercise at hand. I conclude that simple model algorithms, with unbiased spatial data, can provide powerful insights into species environmental relationships, spatial distributions, and patterns of abundance.

# Chapter 2: Improved Spatial Estimates of Climate Predict Patchier Species Distributions.

---

**Publication: Storlie C.J.**, B.L. Phillips, J.J. VanDerWal, and S.E. Williams. 2013. Improved spatial estimates of climate predict patchier species distributions. *Diversity and Distributions*, **9**:1106-1113.

### *Introduction*

How can we know where a species occurs? This question is central to ecological theory and its answer is also of great concern to conservation planners. Increasingly, when faced with uncertainty in species' distributions, we rely on species distribution models (SDMs hereafter) to generate maps describing these distributions. SDMs utilise spatial layers of environmental variables to generate species-environment relationships (SERs) and to predict occurrence in geographic space (Austin 2007). As such, the quality and resolution of these spatial environmental layers will affect the quality of the resultant prediction of species distribution, irrespective of the modelling procedure used (e.g. Kearney & Porter 2004; Elith & Graham 2009).

The accelerated pace of SDMs in addressing diverse problems in applied ecology, including predicting species' invasions (Steiner *et al.* 2008), conservation planning (Gogol-Prokurat 2011), and vulnerability assessments under climate change (Thomas *et al.* 2004) highlights the importance of accurately quantified environmental spatial data. Numerous studies have assessed how SDM outcomes vary according to algorithm employed (Elith *et al.* 2006), extent of study area (VanDerWal *et al*. 2009b), grain-size (Gogol-Prokurat 2011) and regularisation procedure (Anderson & Gonzalez 2011). Despite reviews that acknowledge the importance of accurate environmental layers to the modelling process (e.g. Araujo & Guisan *et al*. 2006, Wiens *et al.* 2009), few studies conduct *a priori* assessments of layer accuracy, or consider the cascading effects of inaccurately quantified environment on model outcomes.

The most commonly used spatial environmental layers in SDMs are undoubtedly those describing climate, but their accuracy, and the effects of systematic errors in these data are rarely considered. Weather data interpolated at broad-scale, or without reference to vegetation density, is frequently inaccurate in topographically complex areas with dense

vegetation (e.g. Scherrer *et al.* 2011). Further exacerbating this effect is the incomplete coverage of environmental gradients (such as vegetation) by weather recording stations. The implementation of these commonly employed and commonly inaccurate spatial layers, represents a major weakness in the SDM paradigm, and must be addressed.

Here I use dense sampling of temperature variables across a topographically and biologically complex landscape (the Australian Wet Tropics) to examine the accuracy and extent of systematic bias in existing best-practice climate layers. To do this, I use a Boosted Regression Tree (hereafter, BRT) statistical downscaling approach to transform existing coarse weather layers to finer-scale weather layers corrected for biotic and abiotic factors. Resultant fine-scale weather layers are aggregated to be comparable to current best-practice climate layers (e.g. ANUCLIM, McMahon *et al.* 1995). I compare the downscaled climate layers (hereafter, 'accuCLIM') to ANUCLIM to determine the spatial pattern of temperature bias and proceed to demonstrate that these two alternative characterisations of climate result in species distribution models that are broadly similar, but which differ strongly in their level of patchiness.

### *Methods*

This study was conducted in the Australian Wet Tropics (AWT): a region exhibiting strong environmental gradients of temperature, rainfall, and vegetation (Nix 1991, Turton *et al*. 1999). Mountainous sub-regions are responsible for these strong gradients, and have also been implicated as the cause of patterns of vegetation stability (VanDerWal *et al*. 2009a) and sub-regional endemism (Graham *et al*. 2006). Seven species of rainforest skink (*Lampropholis coggeri*, *L. robertsi*, *Carlia rubrigularis*, *Gnypetoscincus queenslandiae*, *Saproscincus basiliscus*, *S. czechurai*, and *S. tetradactyla*) were chosen for the SDM exercise. These species are almost entirely constrained to rainforest, yet have contrasting functional ecologies and habitat use profiles.

BRTs were used to relate empirically measured temperatures to spatial estimates of weather and topography. The empirical (dependent) dataset for the BRT downscale procedure comprises 32,239 measurements of daily maximum temperature ($T_{max}$) and daily minimum temperature ($T_{min}$), gathered by field dataloggers at 54 sites across the AWT during the period June 2004 to June 2009 (Figure 2.1). A set of ten topographic,

weather, and environmental spatial layers were assembled to act as the independent variables for the BRT modelling procedure. These include 5km resolution estimates of daily $T_{max}$ and daily $T_{min}$ from the Australian Water Availability Project (hereafter, AWAP), which are interpolated from weather station data using a three-dimensional topographic spline technique (Jones *et al.* 2009). The independent topographic variables are slope, aspect, elevation, latitude, distance to coast, and distance to stream. Other environmental layers in the independent dataset include Foliage Projected Cover (FPC) and insolation. Detailed descriptions of the datasets for the BRT models can be found in Appendix S1.

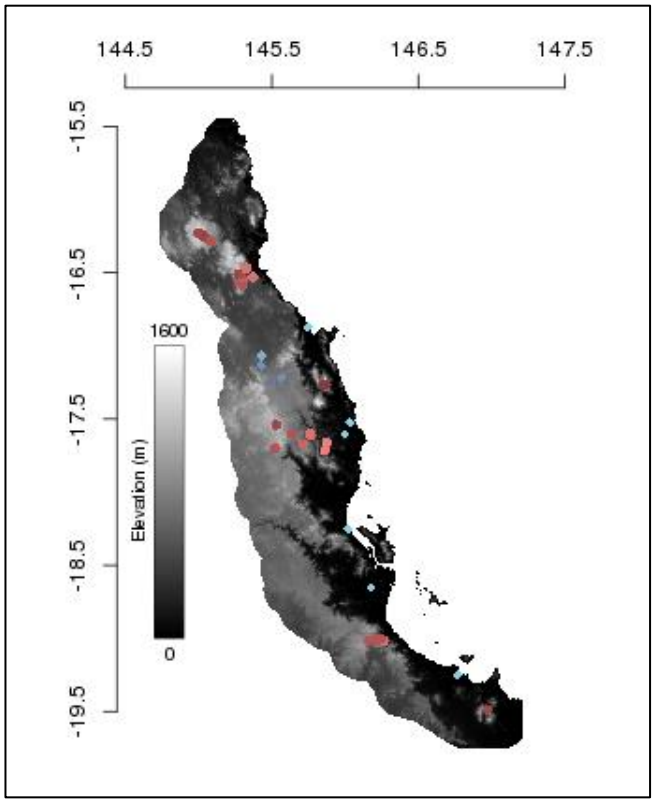



**Figure 2.1:** Map of the Australian Wet Tropics. Red points indicate climate monitoring locations established by the Centre for Tropical Biodiversity and Climate Change. Blue points indicate climate monitoring stations maintained by Australian Bureau of Meteorology. Shade of the coloured points corresponds to elevation of the monitoring sites; darker shades are located at high elevation.

Two separate BRT models (one for daily $T_{min}$ and one for daily $T_{max}$) were created using the above datasets. All model fitting was completed within the *R* Statistical Software package (*R* Development Core Team 2012) using a modified version of Ridgeway's (2010) *gbm* package found in Supplementary Appendix 3 of Elith *et al*. (2008). Root Mean Squared Error (RMSE) was the loss-function chosen to assess how well predictions fit the training dataset. Models were trained using the full dataset; however

tenfold cross-validation (Hastie *et al*. 2001) was used to avoid overfitting the model to the training data. Detailed methods for parameterising the BRT models can be found in the Appendix S2.

Once the optimal BRT models were fitted, they were projected into geographic space using Elith's *gbm.predict.grids* function (Elith *et al.* 2008) for every day between 1st January 1970 and 31st December 2005. I then used simple linear regression to test the accuracy and precision of the AWAP weather layers used for model fitting and their resultant downscaled BRT weather predictions. In both cases, daily weather predictions were regressed against empirically measured weather; accuracy was assessed by departure from expected regression coefficient values (slope of 1, y-intercept of 0) and precision was assessed using the adjusted $r^2$ measure of fit. Furthermore, a subset of 1,000 random BRT weather predictions from 13 sites between 1st January 1970 and 31st December 1979 were regressed against their respective empirical temperature measurements to assess the model's ability to predict into time periods not used for model fitting. The daily BRT predictions were then summarised to produce spatial layers congruent with those produced by ANUCLIM (McMahon *et al.* 1995). The end result of this process was four novel accuCLIM spatial layers at 250m resolution covering the entire extent of the study region which represent: Mean Annual Temperature (MAT), MAT Seasonality (MATS), Mean Maximum Temperature of the Warmest Month (MMTWM), and Mean Minimum Temperature of the Coldest Month (MMTCM).

Two separate SDMs were generated for the case study species using the Maximum Entropy method (Phillips *et al*. 2006). One model included the standard ANUCLIM climatic layers (MAT, MATS, MMTWM, MMTCM, Mean Annual Precipitation, Precipitation Seasonality, Precipitation of the Wettest Month, and Precipitation of the Driest Month) for the modelling regime established by Williams *et al*. (2010). The second model replaced ANUCLIM temperature layers with their accuCLIM equivalent. MaxEnt features were set to 'auto' and the number of model runs was capped at 500. Calculations of key distribution metrics were completed in *R* (*R* Development Core Team 2012) using the 'ClassStat' function within the package SDMTools (VanDerWal *et al*. 2012). Distribution metrics compared between models within species were: total distributional area, core distributional area (where a 'core' cell is one which has no

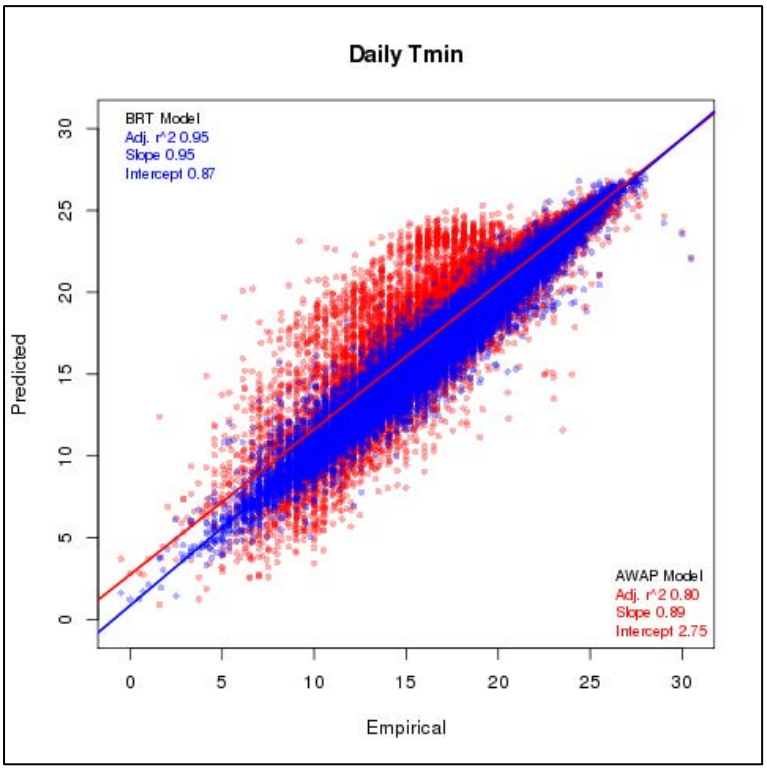
contact with a cell whose value falls below the selected MaxEnt threshold for occurrence), aggregation index, mean fractal dimension index, number of patches, and landscape shape index. For a more detailed description of occurrence records and MaxEnt model parameters see Williams *et al.* (2010).

### *Results*

The BRT procedure drastically improved prediction capability over a null model, decreasing the initial error function by an order of magnitude for both daily $T_{max}$ and daily $T_{min}$ models. For both BRT models of daily $T_{min}$ and daily $T_{max}$ three independent variables accounted almost 90% of all binary splits used to construct the decision tree prediction layers. For the daily $T_{max}$ BRT model these were coarse daily $T_{max}$ (51%), elevation (30%), and Foliage Projected Cover (6%); for the daily $T_{min}$ model they were coarse daily $T_{min}$ (81%), elevation (11%), and coarse daily $T_{max}$ (2%).

Linear regression indicates BRT estimates of empirically measured daily $T_{max}$ were far more precise and accurate (Figure 2.2 – top, slope = 0.95, y-intercept = 1.05, adj $r^2$ = 0.95) than those provided by AWAP weather estimates used for model fitting (Figure 2.2 – top, slope = 0.53, y-intercept = 14.46, adj $r^2$ = 0.57). Results of linear regression indicate that BRT also predicts empirically measured daily $T_{min}$ with greater accuracy and precision (Figure 2.2 – bottom, slope = 0.95, y-intercept = 0.87, adj $r^2$ = 0.95) than do AWAP estimates (Figure 2.2 – bottom, slope = 0.89, y-intercept =2.75, adj $r^2$ = 0.80), although in this case the difference is not as profound.

The results of regressing 1000 random site days of BRT predictions against their respective empirical measurements (which were not used in the training dataset) indicate that the accuracy and precision of the BRT weather layers are maintained when predicting into novel time periods (Daily $T_{max}$: slope = 0.97, y-intercept = 0.99, adj. $r^2$ = 0.97. Daily $T_{min}$: slope = 0.98, y-intercept = 0.3, adj. $r^2$ = 0.97). Importantly, the prediction bias of ANUCLIM climate layers shows clear spatial patterning across the study region, with systematic bias greatest in the coolest, high altitude parts of the study area where accuCLIM MMTWM often differed by 5 – 7 degrees and occasionally by as much as 9 degrees (Figure 2.3).



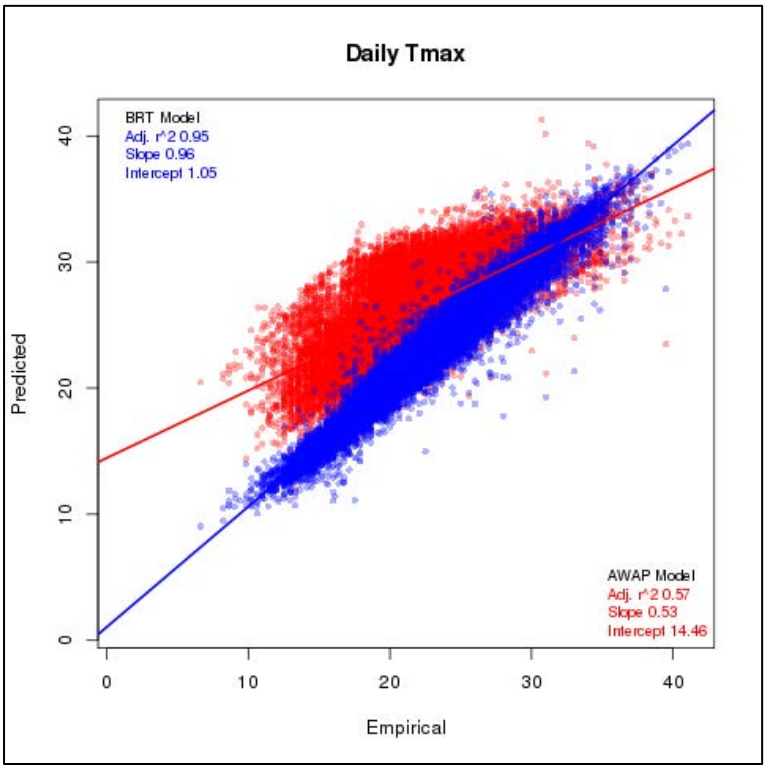




**Figure 2.2:** Scatterplots of empirically measured temperatures against predicted temperature. Blue points represent daily BRT predictions (this study) and red points represent daily broad-scale weather interpolations (from the Australian Water Availability Project). Bottom panel shows results for $T_{max}$, top panel shows results of $T_{min}$.

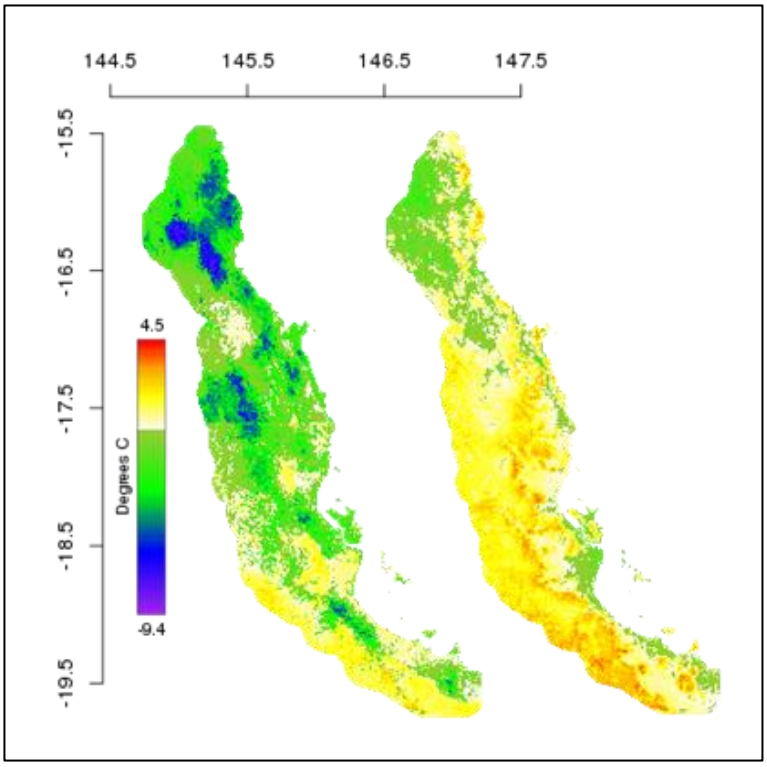



**Figure 2.3:** Left – Difference between accuCLIM and ANUCLIM Mean Maximum Temperature of the Warmest Month. Right – Difference between accuCLIM and ANUCLIM Mean Minimum Temperature of the Coldest Month. Dark blue and purple colours indicate areas where ANUCLIM is predicting temperatures much higher than expected.

Interestingly, when these different spatial climate layers were used in species distribution models, there was rarely much difference in total distributional area (Figure 2.4F). Importantly, the contribution of rainfall variables to the models (results not presented), varied little (<5%) between models within species, indicating that different model outcomes are associated with the accuracy of temperature data. MaxEnt model AUC was also universally high (greater than 0.9, regardless of species or the set of environmental layers utilised), indicating good model fit in all cases. Furthermore, model AUC showed very little discrepancy between models within species, never varying by more than .002 (results not shown). Metrics which describe fragmentation of the distribution (e.g. aggregation index – Figure 2.4A, and number of patches – Figure 2.4D) indicate a greater number of patches which are, on average, smaller when climate is characterised by the accuCLIM layers. Accordingly, the 'core' of these species distributions is predicted to be smaller under the more precise climate layers I derive here (Figure 2.4E). Landscape shape and mean fractal dimension indices also indicate the landscape as a whole and individual patches are more 'regularly' shaped (have fewer edges) when modelled with ANUCLIM (Figures 2.4B & 2.4C).

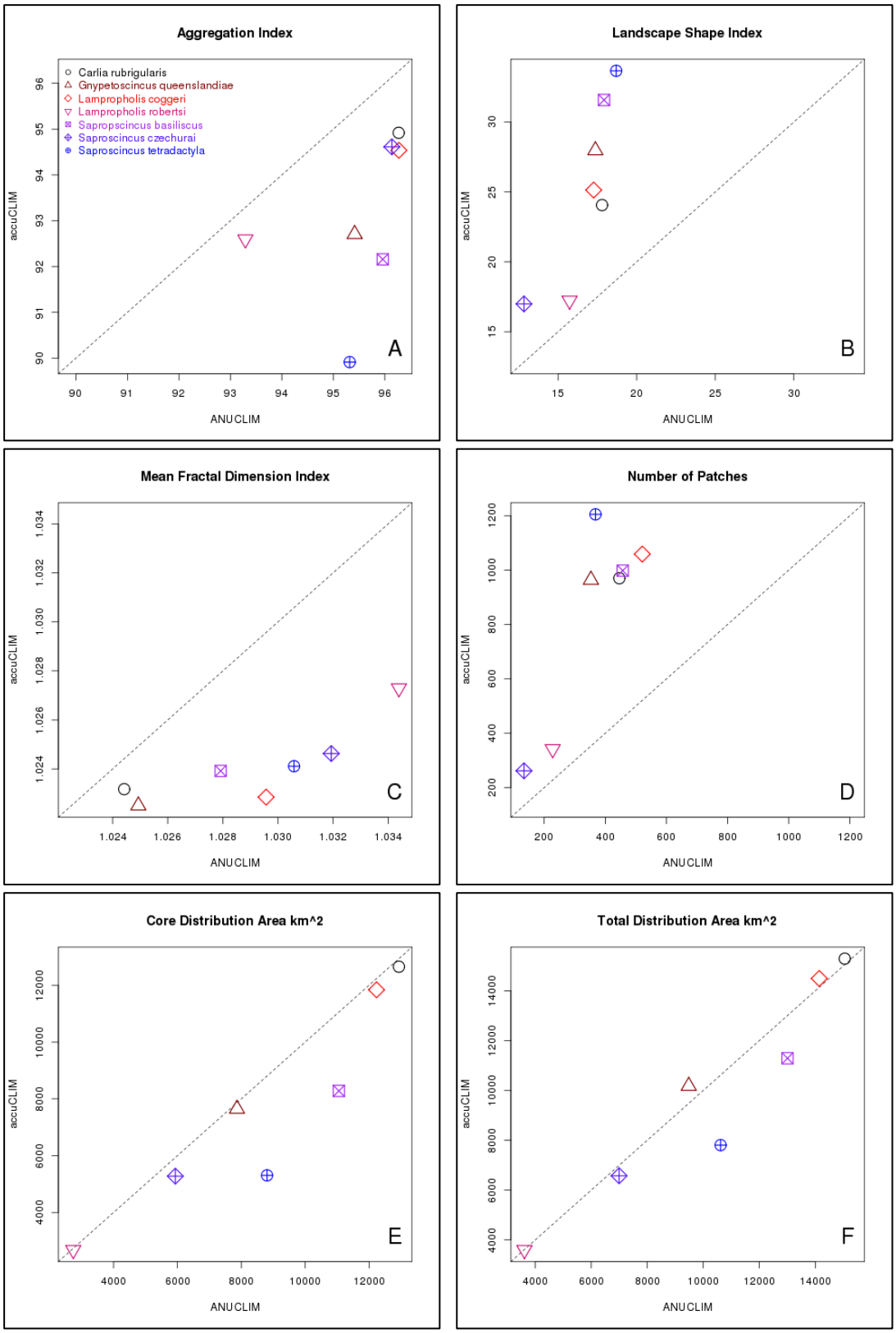



**Figure 2.4 A-F:** Comparison of distribution metrics from ANUCLIM SDM (x-axis) and accuCLIM SDM (y-axis) for 7 species. Top left to bottom right: (A) Aggregation Index, (B) Landscape Shape Index, (C) Mean Fractal Dimension Index, (D) Number of Patches, (E) Total Core Distribution Area, and (F) Total Distribution Area.

### *Discussion*

The BRT statistical downscaling procedure produced estimates of weather that were consistently more accurate than current best-practice weather layers (e.g. AWAP). This result was not unexpected considering the increased spatial resolution provided by statistical downscaling. More surprising, however, is the extent of the differences between climate layers: for MMTWM ANUCLIM layers differed by more than seven degrees from the accuCLIM layers. Importantly, the bias inherent in the ANUCLIM layers was distributed systematically in space (Figure 2.3), suggesting that the interpolation process behind the ANUCLIM data misses critical information in these places. Likely, this information dearth is caused by the incomplete placement of weather recording stations along the elevational and vegetation gradients of the region, coupled with the fact that spatial interpolation operates completely naïve to the buffering effect of vegetation on daily $T_{max}$; an effect shown to be particularly important in upland regions (Shoo *et al*. 2011). Thus, it seems that broad-scale climate data will likely show similar large inaccuracies in other high altitude, topographically complex regions around the world. My analysis shows that these inaccuracies can be corrected if placement of weather recording stations is made with explicit reference to environmental gradients, and if the interpolation procedure allows non-linear relationships and uses biotic as well as abiotic covariates. The necessity of a downscaling procedure to increase accuracy of weather maps, however, may not be warranted in areas with low topographic relief, and with adequate coverage of environmental gradients.

Given that correlative species distribution models simply build species environment relationships with the input data (naïve to ecology and physiology of the organism), I may not have expected SDMs built on biased ANUCLIM data to predict species distributions particularly different from those it would predict using the accuCLIM layers. One could expect that the species environment relationship imputed by the model to be inaccurate when the input data are inaccurate, but this might not lead to different predictions of species distribution given that the model is then projected back onto those inaccurate layers. Thus, I was not surprised when the area predicted for each of the seven modelled species was not particularly different between ANUCLIM and accuCLIM projections (Figure 2.4F). However, the alternative classification of climate

yielded very different estimates of core distribution area and distribution fragmentation (Figure 2.4E, 2.4A & 2.4D). The predicted distributions of the seven modelled species were consistently more convoluted (more edges) when modelled with the more accurate accuCLIM layers developed here (Figure 2.4 A-D). Why then would fragmentation between models be so different, while total area remains unaffected? The most likely explanation is that 'edge' cells from the ANUCLIM model were given artificially inflated Environmental Suitability values by the MaxEnt model because they did not account for the complex and non-linear interactions between topography, the environment, and local weather at these locations. The BRT downscale procedure however, accounted for these interactions and provided far more accurate temperature estimates in these locations. In this way, many edge cells predicted by ANUCLIM to be sub-optimal, but still suitable for a species, may have been predicted unsuitable by the accuCLIM model. Concentrations of these cells in habitat corridors (for instance connecting two mountains within a single range) would lead to greatly increased number of patches and fewer core areas, while having relatively little effect on total predicted distribution and population connectivity.

The effect of downscaled climate layers on fragmentation appears to be exacerbated in species that are elevationally restricted or have small distributional areas (i.e. *Lampropholis robertsi*). This increased patchiness in the predicted distributions is important, because more fragmented habitat typically make populations less resilient, as populations in small patches have increased vulnerability to local extinction (MacArthur and Wilson 1967, Laurance *et al.* 2002). Fragmented populations suffer from increased edge effects which may alter ecosystem composition, with concomitant and often unpredictable effects on population dynamics of individual species (Laurance *et al.* 2002). Furthermore, species existing in fragmented populations may find it more difficult to shift their range in response to climate change (Barton *et al.* 2012). Thus, in this case, inaccurate climate layers suggest a much more robust population than that predicted by the more accurate layers I developed. The extent to which I am overestimating the connectivity of populations generally remains to be seen, but these results sound a cautionary note that many species distributions may be more fragmented than we currently appreciate.

Past studies have examined effects of downscaled climate on SDM outcomes, however, few of these have compared across models with equivalent grain size and extent, as I have done here. Trivedi *et al*. (2008) and Randin *et al*. (2009) both compared continental scale distribution of vascular plants to local scale distribution via downscaling climate grids; interestingly these papers came to contrasting conclusions. Randin *et al*. (2009) found that increasing spatial resolution of climate layers led to increased persistence under future climate change scenarios, further they found a strong relationship between inaccurate prediction in continental scale cells and local topographic variation within those cells. Trivedi *et al*. (2008) reach the opposite conclusion, continental scale future climate change predictions predict fewer extinctions than those modelled at local scale. The cause of discrepancies arising between predicted distributions as a result of climate downscaling cannot, however, be properly assessed unless spatial extent and grain size are equivalent. My study holds all model factors (spatial extent, grain size) constant, allowing us to impute that differences between predictions were attributable to inaccuracies present in the spatial description of the environment.

The inclusion of topographic data in SDMs has been demonstrated to alter model outcomes under climate change scenarios (Luoto & Heikennen 2008). This is a conceptually similar approach to my technique; Luoto & Heikennen (2008) have effectively added another facet to the n-dimensional hypervolume used to represent the species fundamental niche and hence determine SERs for the model. My technique refines the n-dimensional hypervolume by increasing the accuracy of the environmental layers used to calculate it. As a result, SERs are altered, and model outcomes change. Both techniques effectively provide a more realistic model of species distribution in space, but ours allows altered model outcomes to be attributed directly to the inaccurate quantification of the spatial environment.

I have developed a highly accurate snapshot of daily $T_{max}$ and $T_{min}$ across the region between 1970 and 2009 by improving upon already existing spatial weather predictions (e.g. AWAP). Downscaling weather, as opposed to climate, holds multiple distinct advantages. In particular, we now live in a warming world where the occurrence of extreme temperature events is predicted to increase more rapidly than average temperature (Meehl & Tebaldi 2004). The ability of climate layers (e.g. ANUCLIM) to

represent these events is low, as they intentionally smooth over them to represent average conditions. The downscaling procedure I implement here allows the data to be aggregated in whatever fashion may be necessary; as a climate average or as a probability distribution, describing the frequency of extreme temperatures. Moreover, the fine spatial resolution afforded by downscaling allows climate and weather description across multiple scales from landscape scale buffering properties to the thermal exposure of a small population or individual, the latter being a critical step in process based models. It is clear, therefore, that increasing the accuracy and precision of spatial weather layers represents a substantial improvement for ecological analyses across spatial and temporal scales.

# Chapter 3: Stepping Inside the Niche: Microclimate Data are Critical for Accurate Assessment of Species' Vulnerability to Climate Change

---

**Publication: Storlie, C.J.**, A. Merino-Viteri, B.L. Phillips, J.J. VanDerWal, J. Welbergen, and S.E. Williams. 2014. Stepping inside the niche: microclimate data are critical for accurate assessment of species' vulnerability to climate change. *Biology Letters,* **10**(9):20140576.

### *Introduction*

Global climate change represents a threat to biodiversity across multiple biomes and organisational scales (Parmesan & Yohe 2003). In the face of this threat, robust estimation of species' vulnerability to climate change is necessary (Williams *et al*. 2008). Species' vulnerability can be seen as a function of the environmental regimes which a species experiences *in situ* (its exposure) and its physiological and adaptive responses to this environment (its sensitivity) (Williams *et al*. 2008). The use of inaccurate measures of exposure or sensitivity for creating vulnerability analyses can potentially lead to false inference and wasted conservation resources. Hence, we need to examine closely the analytical procedures used to derive estimates of exposure and sensitivity for species.

Coarse scale weather layers will likely make a poor surrogate for the microclimate experienced by an organism. Coarse scale weather layers are typically created from empirical point climate data (daily temperature maxima or minima, $T_{max}$ or $T_{min}$), which are then splined through un-sampled geographic space according to elevation, latitude, and longitude (e.g. Jones *et al*. 2009). Splining approaches do not take into account factors known to decouple local and regional thermal regimes and as a result are often inaccurate at the microclimate scale (Shoo *et al*. 2011, Scherrer *et al*. 2011, Storlie *et al*. 2013, Chapter 2). As well as this, coarse-scale weather takes no account of the species' behaviour: nocturnal and diurnal species, for example, may have very different exposures.

We can move from coarse weather layers to fine scale layers, and ultimately species' exposure, using either a correlative (statistical [Storlie *et al*. 2013, Chapter 2]) or a mechanistic (mass/energy balance [Kearney & Porter 2009]) downscaling approach. Mechanistic methods utilise complex energy balance equations which incorporate spatially mapped variables such as surface albedo, relative humidity, incoming solar radiation and wind speed (as well as conductivity and emission constants) to generate estimates of microclimate (Kearney & Porter 2009). While very flexible, particularly for predicting microclimates into the future, mechanistic approaches require a large number of parameters to be estimated. Alternatively a correlative approach may lend itself well to producing spatial estimates of microclimate under current conditions (Storlie *et al*. 2013, Chapter 2). Such an approach does not explicitly incorporate mechanism, but draws statistical associations between empirical microclimate observations and spatial layers of environmental factors (Storlie *et al*. 2013, Chapter 2). Unlike the mechanistic approach, the correlative approach does not require a large number of parameters to be specified *a priori*, and it makes use of large datasets on microclimate that already exist (Storlie *et al*. 2013, Chapter 2).

Microclimate estimates will need to meet three criteria to be relevant to the species in question and therefore suitable for constructing vulnerability analyses. First, factors that decouple regional and local microclimates will need to be explicitly incorporated. Second, microclimate estimates need to be temporally resolved enough to quantify the effect of short-term weather events on the focal species (Welbergen *et al*. 2008, Reside *et al*. 2010). Finally, microclimate estimates will need to be consistent with what we know of the current distribution and physiological constraints of the species in question. Meeting these criteria should produce estimates of microclimate which are biologically sensible and useful for vulnerability assessment.

Here, I assess the correlative approach to downscaling weather data against these three criteria. I compare three datasets that represent iteratively more complex statistical downscaling. The first dataset (AWAP) is of coarsely resolved (5km resolution) data on daily $T_{max}$ and $T_{min}$ (Jones *et al*. 2009). The second dataset (DS1) is a previously-described statistical downscale (to 250m resolution) of the first dataset that incorporates factors that decouple regional and local climates (Storlie *et al*. 2013, Chapter 2). I compare these two datasets against the known critical thermal maxima ($CT_{max}$, a hard

physiological limit on individual survival (Lutterschmidt & Hutchison 1997) of seven species of endemic frog (Family *Microhylidae*) in the Australian Wet Tropics (AWT) and find them wanting. Both indicate thermal regimes that routinely exceed the focal species' $CT_{max}$, yielding an impossible situation in which focal species thrive outside their fundamental thermal niche. I proceed to further statistically downscale DS1 using an independent dataset of microhabitat temperature records from the focal species' preferred diurnal shelter (underneath fallen logs in rainforest). The resultant dataset (DS2) accurately predicts observed temperatures and, importantly, projects temperature regimes compatible with the thermal niche of all the focal species. Indeed, the downscaled weather data place the focal species firmly within their fundamental thermal niche, suggesting that these results are now at a sufficient spatiotemporal scale as to be biologically valid and useful for constructing vulnerability analyses.

### *Methods*

I used the AWT as a case study region to demonstrate the problems inherent with utilising broad-scale weather layers to construct microclimates. To this end, I utilise a biological criterion to assess the accuracy of microclimate layers: their ability to accurately portray a species fundamental thermal niche. For the latter I ask a simple question: do thermal regimes predicted by weather layers at known species' occurrence points conform to known $CT_{max}$ of the species?

Comparisons are drawn between three representations of thermal regimes with increasing spatial resolution and complexity of statistical downscaling. The AWAP layers represent the daily open-air $T_{max}$ at ~1.5m above the ground at 5km daily resolution (Jones *et al*. 2009). The DS1 weather layers representing daily $T_{max}$ and $T_{min}$ are downscaled to 250m resolution from the AWAP layers, according to key environmental factors (Storlie *et al*. 2013, Chapter 2). Lastly, I use the DS1 weather layers (Storlie *et al*. 2013, Chapter 2) in conjunction with a paired empirical dataset of air and microhabitat (underneath fallen logs) temperature records. A linear model is employed to generate a relationship between daily air temperatures and microclimate $T_{max}$. Spatially mapped air temperature values from DS1 are then substituted into the model to form a region-wide spatial prediction of the daily thermal regime underneath a fallen rainforest log (DS2 layers). For further details of the linear downscaling model see Appendix S3.

$CT_{max}$ was determined for seven species of rainforest-restricted, terrestrially developing Microhylid frog: *Cophixalus aenigma, C. bombiens, C. exiguus, C. hosmeri, C. infacetus, C. mcdonaldi,* and *C. neglectus*. The focal species are known to be nocturnally active and shelter underneath fallen logs during daylight hours (Williams *et al.* 2010, C. Storlie pers. obs.). Adult male frogs were identified by call and tested within 24 hours using a dynamic methodology (Cowles & Bogert 1944) to determine $CT_{max}$. For further details on the determination of the focal species $CT_{max}$ see Appendix S4.

Location data for species occurrences were derived from Williams *et al.* (2010) and supplemented with recent records vetted for positional and observer accuracy. These occurrences were then overlaid on the weather layers in a GIS environment using the SDMTools package 1.13 (VanDerWal *et al* 2012) in the *R* Statistical Software Package v. 2.15.1 (*R* Core Team 2012) to extract absolute $T_{max}$ for all occurrence points of focal species. Thirty-eight years of absolute $T_{max}$ at known occurrence points were then standardised against each species' thermal limits by subtracting individual species' $CT_{max}$ from all records. I used kernel density estimation to represent the distribution of temperatures for all focal species. Density distributions were calculated across the entire range of predicted standardised temperatures and the total probability density falling above each species' $CT_{max}$ was calculated (i.e. the density above zero after standardisation). These values represent the proportion of time (given the data) that $CT_{max}$ is exceeded across the set of species occurrence locations for each species and are reported in Table 3.1.

### *Results*

In producing DS2, the linear model of microclimate $T_{max}$ fits the empirical temperature data well, giving us confidence in its ability to estimate microclimate $T_{max}$ from air temperature (Figure 3.1).

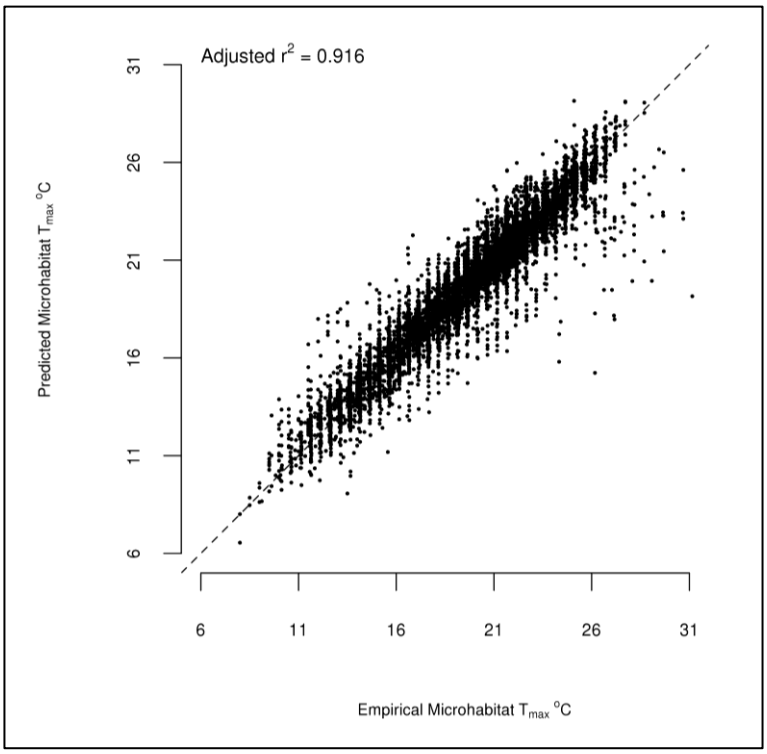
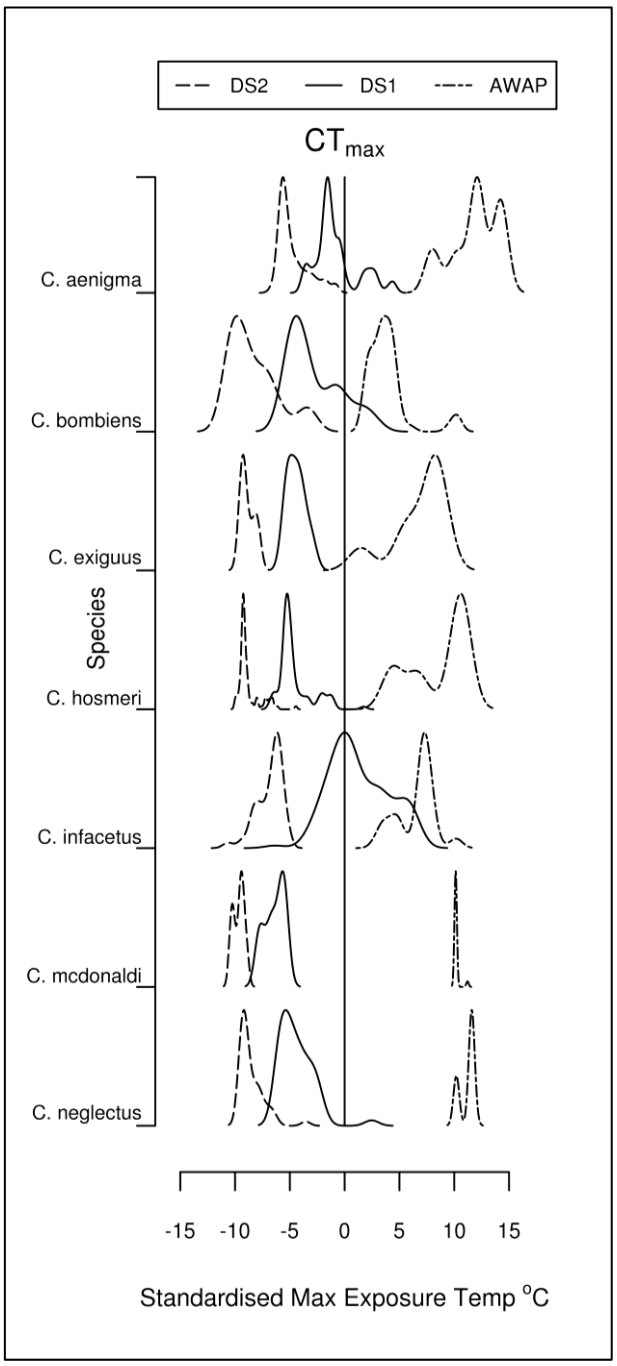



**Figure 3.1**: Relationship between empirical microclimate (underneath a fallen log) $T_{max}$ and predicted microclimate $T_{max}$ from the linear microclimate model. Black line represents a 1:1 relationship.

For all focal species, the thermal regimes predicted by AWAP and DS2 are non-overlapping; indicating substantial differences in temperature between datasets for all species (Figure 3.2). Importantly, the DS2 spatial layers rarely produce temperature estimates that exceed focal species' $CT_{max}$ (Table 3.1, Figure 3.2). This contrasts sharply with AWAP layers for which the majority of all occurrences exist in locations which exceed the focal species' $CT_{max}$ (Table 3.1, Figure 3.2). The DS1 layers produce estimates of thermal regimes that exceed species $CT_{max}$ for four of the seven target species (Table 3.1, Figure 3.2).

**Figure 3.2:** Thermal regimes predicted by three sets of weather layers. Temperature (x-axis) is standardised against individual species $CT_{max}$ (zero is $CT_{max}$ for each species). The y-axis shows the probability density of temperature given each dataset, scaled against the maximum density for each species set.




Summing kernel densities above individual species' $CT_{max}$ indicates an extremely low probability that any DS2 thermal regimes exceed the species known thermal limits ($p < 0.0002$ in all cases: Table 3.1); while the same procedure using the AWAP data shows an extremely high probability that conditions at known occurrence sites exceed the species fundamental thermal niche (Table 3.1).

**Table 3.1:** $CT_{max}$ values and standard deviations (in °C) for all focal species, the number of occurrence records, and the proportion of time that mean $CT_{max}$ is exceeded if I treat each of the datasets as truth.

| Species | $CT_{max}$ | SD of $CT_{max}$ | N ($CT_{max}$) | N (Occurrence Records) | AWAP Prop. Above $CT_{max}$ | DS1 Prop. Above $CT_{max}$ | DS2 Prop. Above $CT_{max}$ |
|---|---|---|---|---|---|---|---|
| *C. aenigma* | 28.09 | 2.31 | 3 | 148 | 0.9595 | .2316 | 0.0002 |
| *C. bombiens* | 32.36 | 3.16 | 10 | 58 | 0.9990 | .1692 | <0.0001 |
| *C. exiguus* | 35.93 | 1.00 | 6 | 8 | 0.9874 | <.0001 | <0.0001 |
| *C. hosmeri* | 31.73 | 0.41 | 4 | 86 | 0.9990 | .0111 | <0.0001 |
| *C. infacetus* | 35.08 | 0.71 | 5 | 121 | 0.9990 | .6355 | <0.0001 |
| *C. mcdonaldi* | 32.77 | 0.65 | 8 | 22 | 0.9999 | <.0001 | <0.0001 |
| *C. neglectus* | 30.70 | 2.56 | 13 | 45 | 0.9985 | .0213 | <0.0001 |

## *Discussion*

Currently available weather layers are powerful resources for ecological applications, yet in the context of vulnerability analyses they require adjustment to reflect microclimates species experience *in situ.* The addition of key environmental information to the AWAP layers via statistical downscaling allows for the DS1 layers to very accurately depict thermal regimes of open-air conditions (Storlie *et al*. 2013, Chapter 2). In some cases, this procedure alone is enough to produce estimates of thermal regimes that are within the species' fundamental thermal niche. Yet these layers are still not biologically sensible for a nocturnal species which shelters underneath logs during the day. The addition of the second, simpler, downscale procedure to the DS1 layers produces estimates of temperatures underneath logs during the day. The resultant dataset, DS2, produces a biologically meaningful depiction of thermal regimes that is consistent with the known thermal limits of all species.

I have presented a clear method for deriving microclimate layers which are accurate with respect to empirically measured temperatures and which predict thermal regimes within the known physiological limits of resident species. AWAP weather layers consistently predict thermal regimes at known occurrence points that exceed the focal

species' $CT_{max}$, effectively placing individuals outside of their fundamental niche. The microhabitat-specific DS2 weather layers generated here effectively never predict thermal regimes which exceed a focal species' $CT_{max}$ at a known occurrence point. This finding strongly supports the notion that increasing the biological relevance of weather layers – by including the important effects of behaviour (microhabitat use) and buffering at multiple scales – is a key aspect of robust vulnerability assessments.

As well as demonstrating the importance of downscaling coarsely resolved temperature data, this study also demonstrates the importance of using biological criteria (e.g. the physiological limits and distribution of species) to verify the resultant data. While both mechanistic and correlative approaches can be used to estimate microclimate (for a mechanistic approach, see the recent global dataset of Kearney *et al*. (2014), they both need to meet biological criteria to be convincing. In meeting biological criteria they can be more confidently used to estimate species' vulnerabilities.

Global climate change and other stressors will continue to threaten biodiversity into the foreseeable future, making the application of robust vulnerability analyses for species key to conservation outcomes. Estimates of species' exposure and sensitivity, which lie at the core of these analyses, must both then be accurate. Inaccuracies in these estimates may result in biologically nonsensical outcomes and call into question inference on exposure that flows from such data. Thus, we must generate accurate estimates of species' thermal regimes to be confident in the outcome of species' vulnerability assessments. These results show that correlative techniques with explicit consideration of key abiotic and biotic factors provide biologically meaningful estimates of thermal regimes.

# Chapter 4: Can Less Be More? Comparing Predictions of Species Abundance Using Presence-Only and Presence-Absence SDM Techniques.

### *Introduction*

Robust estimates of species' occurrence and abundance, and how these patterns vary through space, lie at the core of many ecological analyses. For this reason, ecologists have long been concerned with modelling these properties of a population. A major source of structural error in abundance datasets is species detectability, and so elegant models which account for variable detectability have been developed and can be used to infer the environmental drivers of abundance (e.g. MacKenzie *et al.* 2002, MacKenzie and Kendall 2002, Royle & Nichols 2003, Royle *et al*. 2005). Unfortunately, these techniques are data hungry and require large standard-effort presence-absence datasets to separate the drivers of population detectability and abundance. Detection invariant methods, on the other hand, have much looser requirements (e.g. ad hoc presence-only records can be used) and so there is typically more data available for detection invariant methods than for variable detection methods. When the aim is to predict distribution or abundance patterns through space, it is critical to have datasets that sample a broad range of environments (Austin *et al.* 1994, Hernandez *et al.* 2006). Because of their larger size, presence-only datasets will typically achieve a much broader sample of environments than presence-absence datasets. Yet it remains to be seen for spatial prediction if more data in a more flexible model is more powerful than fewer data run through a more elegant variable detection model.

Detection rates for all species are less than one (Kery & Schmidt 2008) therefore any modelling technique which does not account for detection is almost certain to underestimate abundance (McKenzie *et al.* 2002). Detection rates are influenced by a variety of factors including: habitat type, species ecology, observer effects, weather and more. To account for varying detection, occupancy models (which account for dynamic species detection rates as a function of environmental covariates) allow simultaneous estimation of species detection rates and abundance patterns (e.g. Royle & Nichols

2003). These methods (hereafter referred to as 'occupancy models') rely on repeated surveys of locations across time and produce measures of abundance, corrected for the effect of imperfect detection (MacKenzie *et al*. 2002). Occupancy models are a powerful technique for examining abundance patterns at sites with repeated survey effort, but logistical constraints mean that the number of sites surveyed in this manner will always be limited, so building robust associations between abundance and environmental covariates using an occupancy framework can be challenging.

Detection-invariant models, on the other hand, can still be employed to estimate species' distribution patterns without accounting for detectability (Austin 2002, Elith *et al*. 2006). Ignoring detectability means that data requirements are much looser – a much broader range of localities can be represented – so building associations between abundance and environment is more readily achieved. This flexibility means that detection-invariant models have become the tool of choice for modelling species' distributions, and a broad array of numerical machinery is now available for fitting these models (Elith *et al*. 2006). While detection-invariant occupancy models do not produce direct measure of abundance, the outputs of these models have been demonstrated to correlate well with abundance (VanDerWal *et al*. 2009c, Torres *et al.* 2012, Van Couwenberghe *et al*. 2012). In particular, presence-only detection-invariant models (*e.g.* MAXENT, Phillips *et al*. 2006) have been shown to predict empirically measured carrying capacity for a wide range of taxa (VanDerWal *et al.* 2009c). Further, these models fit complex relationships between environment and occurence, and so are good at representing complex realities.

Here, I draw comparisons between detection invariant MaxEnt models and variable detection rate abundance models and their capacity to predict empirical abundance data. I demonstrate the ability of these models to predict carrying capacity via comparison to a large dataset of repeated count surveys for a group of 7 skinks (Family: *Scincidae*) in the rainforests of the Australian Wet Tropics (AWT). For the purposes of comparison, I gave each technique the same input data and linear covariate structure. MaxEnt models were trained using only location data included in the occupancy models and allowing only linear and quadratic relationships between environmental variables and presence records. I then removed these constraints on the MaxEnt model, giving it the full presence-only dataset, and allowing it to fit complex relationships. When MaxEnt was

constrained to the same data and structure as the occupancy model, the occupancy model outperformed MaxEnt at predicting empirical abundance. When the data and structure constraints were removed, however, MaxEnt had similar performance to the occupancy models. In this case, then, the looser data and structural constraints of the detection invariant model made up for the fact that it didn't explicitly account for detection.

### *Methodology*

I proceeded to assess the ability of two distinct modelling approaches (detection variant / invariant) to predict measured count data for a group of 7 species. To achieve this goal, I used quantile regression (95$^{th}$ percentile, see Cade *et al*. 1999) to relate the upper limit of empirically measured abundance to the output of each model. Because detection is imperfect and varies through time, multiple surveys will tend to exhibit negative departures from some maximum abundance (i.e. carrying capacity). Hence I focused on the prediction of the 95$^{th}$ percentile of observed abundance, as opposed to the mean observed abundance.

Two-hundred twenty one sites were surveyed in the WTWHA which covered key environmental gradients (latitude, elevation, rainfall, and temperature) for the presence and abundance of a group of rainforest skinks. The focal species of these surveys were (Family *Scincidae*)*: Carlia rubrigularis, Gnypetoscincus queenslandiae, Lampropholis coggeri, L. robertsi, Saproscincus basiliscus, S. czechurai,* and *S. tetradactyla*. One-person hour active surveys (see Williams *et al.* 2010) were repeated at all sites (minimum 3 repeats, maximum 14) over a 13 year period, ending with a collection of 1,215 surveys with a count of all 7 species for each. Count data was averaged by individual species at each site across surveys; then average counts were made proportional to the maximum average count across all sites.

Variable detection occupancy models and detection invariant, presence-only MaxEnt models were employed to relate model outputs to the empirically measured abundance data. First, presence-only MaxEnt models are employed (Phillips *et al*. 2006). Although they predict 'environmental suitability' of the landscape for a species, prior research (VanDerWal *et al.* 2009c, Torres *et al*. 2012) has shown this metric relates to carrying capacity. MaxEnt models were developed for each species using presence-only data

from Williams *et al.* (2010). Initial MaxEnt models were trained using only presence records from the sites where repeated surveys were conducted, and were limited to linear and quadratic model features (i.e. a linear regression structure referred to hereafter as 'Occupancy Equivalent MaxEnt models'). Next I fit unconstrained MaxEnt models which were trained utilising complex features and the complete complement of presence-only records from Williams *et al.* (2010) (referred to hereafter as 'Saturated MaxEnt Models'). Finally, I compared these outputs to those from occupancy models fitted to the same data: using the detection / non-detection data from the 1,215 surveys conducted (referred to hereafter as 'Occupancy Models'). In this case, abundance is modelled with the technique of Royle & Nichols (2003) implemented in *R* Statistical Software Package v. 2.15.1 (*R* Core Team 2012) with the 'unmarked' package (Fiske *et al.* 2012).

Selected environmental covariates for the models (Occupancy, Occupancy Equivalent MAXENT, and Saturated MAXENT) were the same in all instances. Bioclimatic layers of Mean Maximum Temperature of the Warmest Month, Mean Annual Temperature Coefficient of Variation, Precipitation of the Wettest Month, and Mean Annual Precipitation Coefficient of Variation were selected from a pool of eight commonly employed on the basis of minimising co-variation (see Williams *et al.* 2010, VanDerWal *et al.* 2009c). Temperature and rainfall layers were derived from Xu & Hutchinson (2011). For the occupancy models, three daily environmental covariates were selected to model detection as a function of the environment: daily maximum temperature, daily temperature range, and daily rainfall derived from Jones *et al.* (2009). As a supplement, I also undertook this entire process (all models: Occupancy, Occupancy Equivalent MaxEnt, and Saturated MaxEnt) using an alternative set of climate and weather layers which have been statistically downscaled to increase their accuracy when predicting empirical temperatures (Storlie *et al.* 2013, Chapter 2). By constructing two versions of each model, where the only difference was the accuracy of input weather data, I was able to characterise the effect of inaccurate spatial layers on model outcome for all three model types.

For the occupancy model, all possible unique combinations of abundance (up to 4 covariates) and detection covariates (up to 3 covariates) were generated for a total of 1954 model formulae per species. AIC was calculated for each model, and the model

with the highest AIC is then chosen for each species. MaxEnt models were based on all four climate layers and optimised to balance omission and commission rates: AUC was calculated for all MaxEnt models to assess their fit to the data.

Finally, model output at each sampling site was assessed against the scaled count data from each site for all species. Since counts were scaled as a proportion of carrying capacity, quantile regression (95$^{th}$ percentile) is employed to relate the two metrics. Likelihood ratio pseudo-$r^2$ (Magee 1990) was used to demonstrate improvement over a null model for both modelling techniques. This measure of goodness-of-fit can be used to demonstrate which model has more explanatory power, since the independent variables for both (scaled counts) were the same.

### *Results*

Occupancy models of species abundance patterns varied idiosyncratically in their predictive power (as assessed by Magee's Pseudo-$r^2$ measure of fit) between 0.3 and 0.7 (Table 4.1). Presence-only (MaxEnt) models of species environmental suitability attained universally high AUC scores (above 0.85, indicating a good fit for all models) independent of features utilised and number of sites included (Table 4.2).

Both occupancy and MaxEnt models demonstrated variable fit to empirical count data across species (Table 4.3). Several notable trends were observed concerning the fit of models to the abundance data; first occupancy models outperformed (have higher Pseudo $r^2$ values) Occupancy Equivalent MaxEnt models (Figure 4.1, Table 4.3). The addition of further presence-only records and complex model features to the Saturated MaxEnt model improved their fit to the data in all cases except one (Figure 4.2). However, complex model features and additional sites did not allow detection invariant models to fit the data better than occupancy models (Figure 4.3).

The majority of the relationships between model output and empirical measured carrying capacity were positive for both occupancy and MaxEnt models (Figures 4.4 – 4.6). Predictions of abundance, as formed by relating occupancy and MaxEnt model outputs to empirically measured carry capacity, were nearly equivalent between the two approaches (Figures 4.7). Estimates of abundance were overwhelmingly positive departures from measured abundance (Figures 4.8 and 4.9). Both modelling approaches

predicted variable levels of abundance at sites where measured abundance is near zero (Figures 4.8 and 4.9). Results of models based on the downscaled layers of Storlie *et al.* (2013, Chapter 2), and how they compared to the above models, give qualitatively similar results and can be found in the Appendix S5.

**Table 4.1:** Variable Detection Occupancy Model Summary. The model formulae are abbreviated, the initial letter 'd' describes the detection variables which are: d1 – Rainfall of Day Before Survey, d2 – Max Temperature of Day Before Survey, d3 – Temperature Range of Day Before Survey. The letter 's' describes the state (abundance) variables which are: s1 – Mean Annual Temperature Coefficient of Variation, s2 – Mean Maximum Temperature of the Warmest Month, s3 – Precipitation of the Wettest Month, s4 – Mean Annual Precipitation Coefficient of Variation. Model $r^2$ values are determined by the likelihood ratio test of Magee (1990).

| Species | Model Formulae | Model $r^2$ | AIC | N Sites Found | N Surveys Found |
|---|---|---|---|---|---|
| LAMROBE | ~ d2:d3 ~ s1 + s2 +s3 +s4 + s2:s3 + s1:s4 | 0.389 | 148 | 15 | 29 |
| LAMCOGG | ~ d2+ d2:d3 ~ s1 + s4 +s1:s4 | 0.635 | 1405 | 206 | 715 |
| GNYQUEE | ~ d2 + d3 + d2:d3 ~ s1 +s2 +s3 +s4 | 0.704 | 1245 | 151 | 512 |
| SAPCZEC | ~ d2 + d2:d3 ~ s1 + s2 + s3 | 0.42 | 277 | 31 | 50 |
| SAPTETR | ~ d2 ~ s1 + s2 + s4 + s2:s3 | 0.409 | 649 | 71 | 118 |
| SAPBASI | ~ d1 + d2 + d3 + d2:d3<br>~ s1 + s2 + s4 +s2:s3 + s1:s4 | 0.711 | 924 | 114 | 270 |
| CARRUBR | ~ d2 + d2:d3 ~ s1 + s2 +s3 + s1:s4 | 0.559 | 1219 | 180 | 793 |

**Table 4.2:** Saturated MaxEnt Model Summary.

| Species | Threshold | Omission Rate | N Occurrences | AUC |
|---|---|---|---|---|
| CARRUBR | 0.099 | 0.009 | 470 | 0.861 |
| GNYQUEE | 0.078 | 0.003 | 324 | 0.919 |
| LAMCOGG | 0.081 | 0.009 | 331 | 0.883 |
| LAMROBE | 0.014 | <0.001 | 39 | 0.987 |
| SAPBASI | 0.061 | 0.021 | 291 | 0.889 |

| | | | | |
|---|---|---|---|---|
| SAPCZEC | 0.024 | 0.019 | 53 | 0.952 |
| SAPTETR | 0.056 | <0.001 | 108 | 0.926 |

**Table 4.3:** Quantile Regression Carrying Capacity Model Summary.

| Species | Occupancy Model Pseudo $r^2$ | Occupancy Equivalent MaxEnt Model Pseudo $r^2$ | Saturated MaxEnt Model Pseudo $r^2$ |
|---|---|---|---|
| LAMROBE | 0.998 | 0.472 | 0.852 |
| LAMCOGG | 0.561 | 0.067 | 0.343 |
| GNYQUEE | 0.711 | 0.668 | 0.464 |
| SAPCZEC | 0.836 | 0.531 | 0.656 |
| SAPTETR | 0.764 | 0.088 | 0.610 |
| SAPBASI | 0.775 | 0.253 | 0.395 |
| CARRUBR | 0.294 | 0.195 | 0.002 |

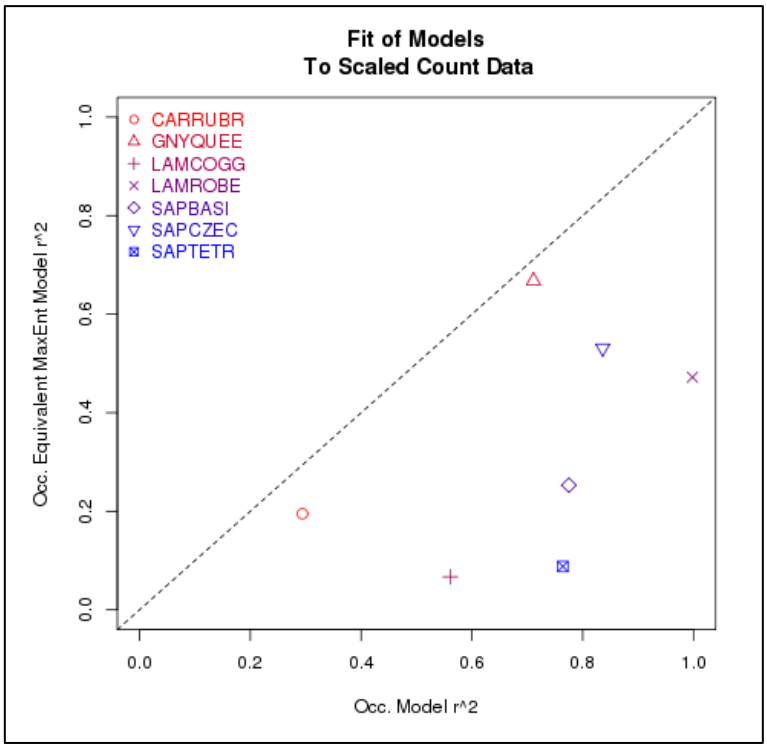



**Figure 4.1:** Comparing pseudo $r^2$ values for the fit of model outputs to the scaled count data, occupancy model values on the x-axis, occupancy equivalent MaxEnt model values on the y-axis.

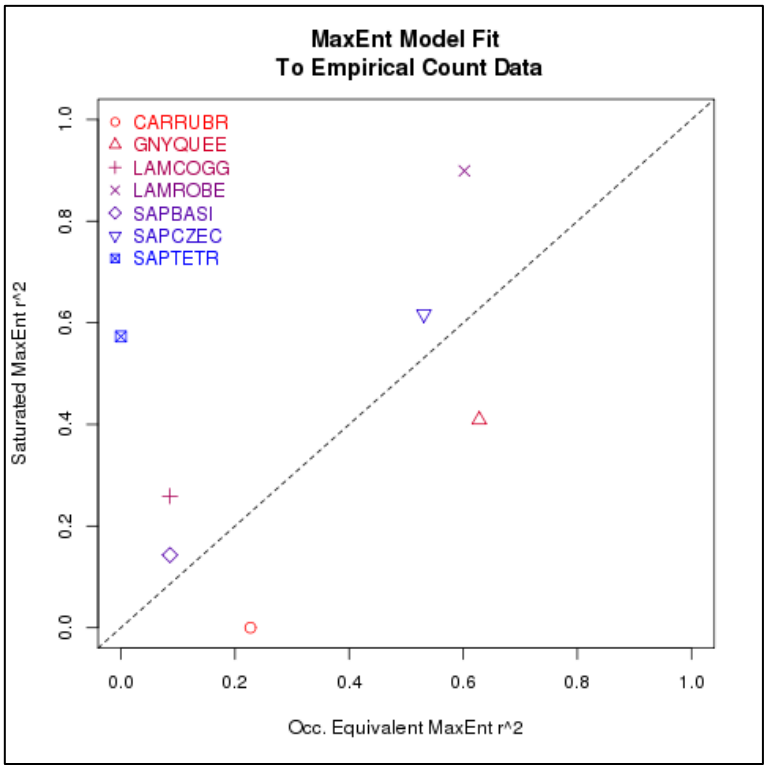



**Figure 4.2:** Comparing pseudo $r^2$ values for the fit of model outputs to the scaled count data, saturated MaxEnt model values on the x-axis, occupancy equivalent MaxEnt model values on the y-axis.

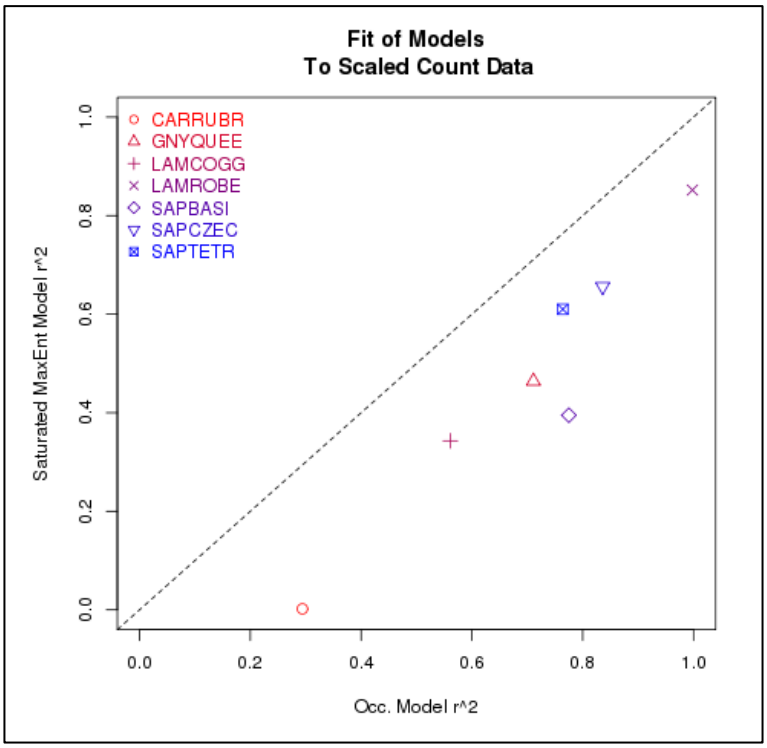



**Figure 4.3:** Comparing pseudo $r^2$ values for the fit of model outputs to the scaled count data, occupancy model values on the x-axis, saturated MaxEnt model values on the y-axis.

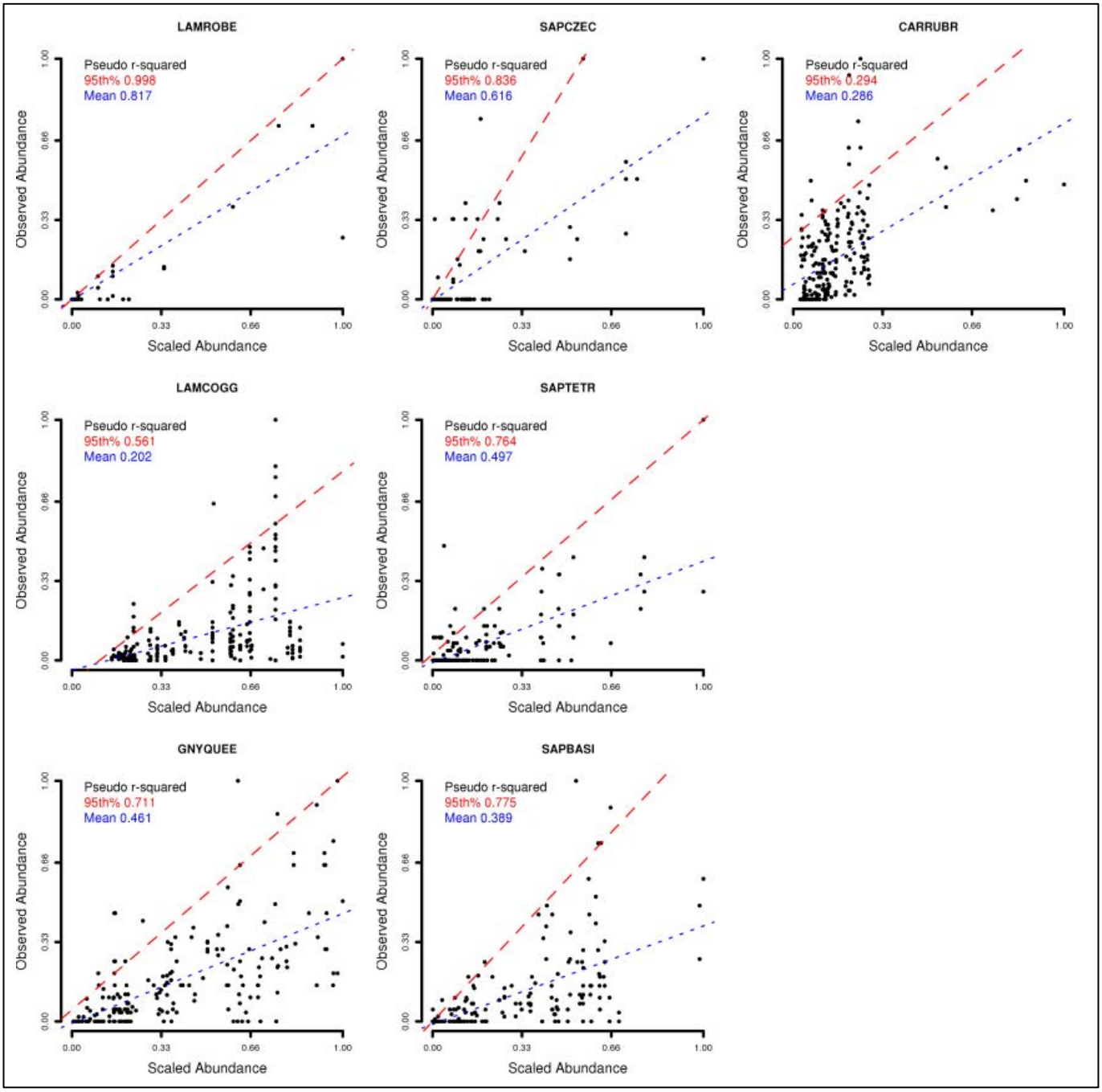



**Figure 4.4:** Fit of occupancy model output (x-axis) to scaled count data (y-axis). The dashed red line represents the fit of the model output to the 95$^{th}$ percentile of count data. The dashed blue lines represent the fit of the model output to the mean of the count data.

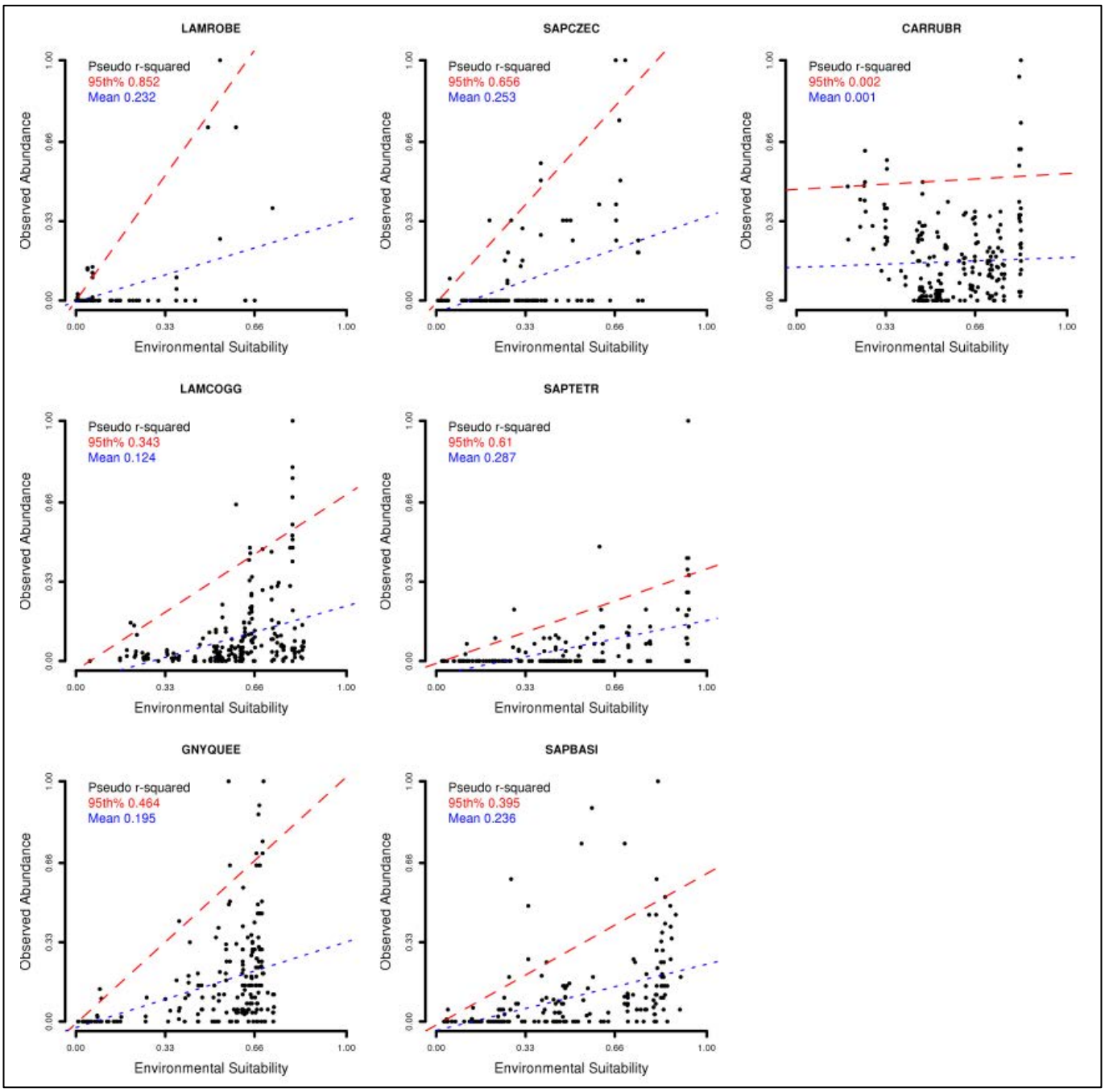



**Figure 4.5:** Fit of saturated MaxEnt model output (x-axis) to scaled count data (y-axis). The dashed red line represents the fit of the model output to the 95$^{th}$ percentile of count data. The dashed blue lines represent the fit of the model output to the mean of the count data.

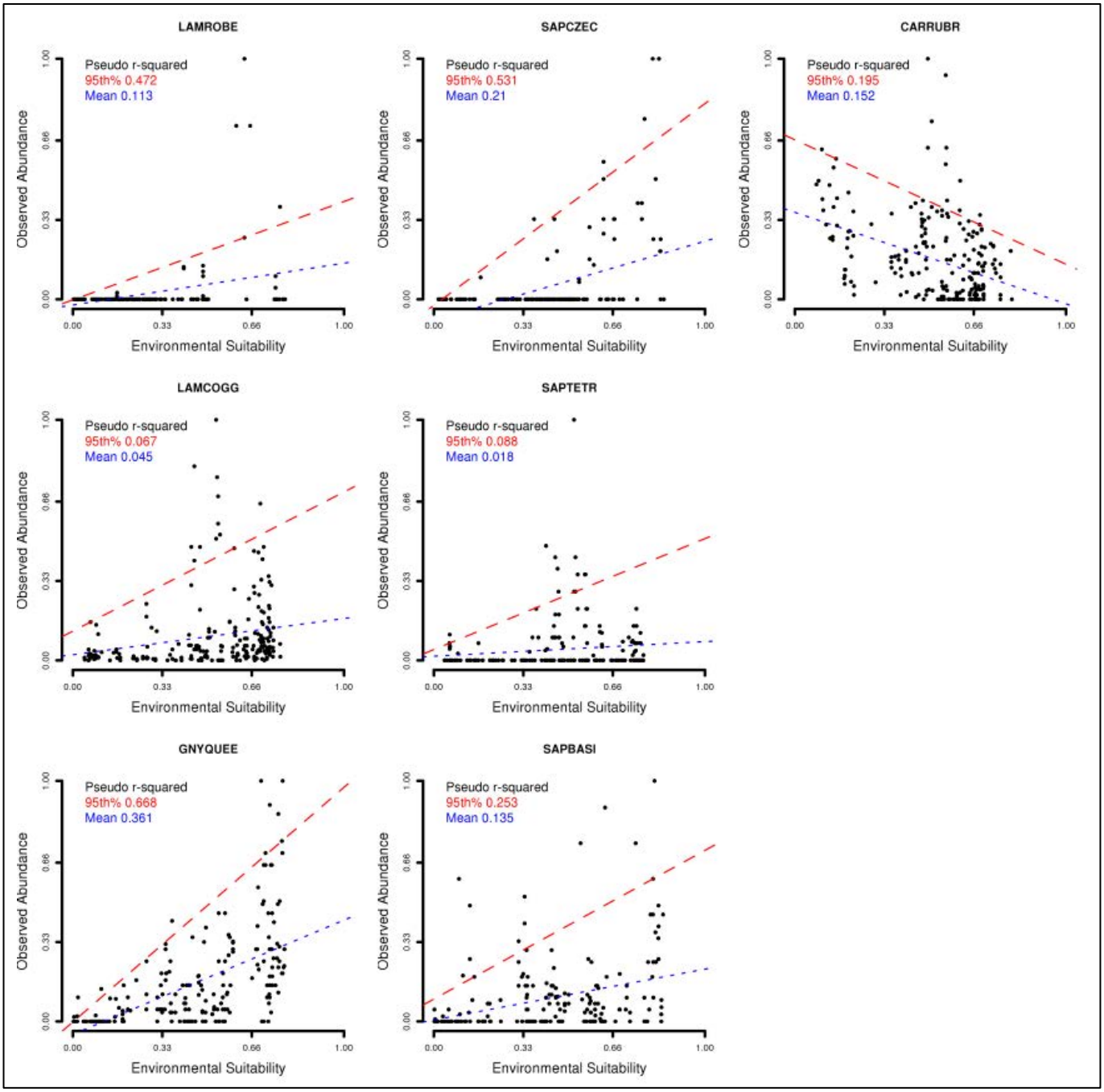



**Figure 4.6:** Fit of occupancy equivalent MaxEnt model output (x-axis) to scaled count data (y-axis). The dashed red line represents the fit of the model output to the 95$^{th}$ percentile of count data. The dashed blue lines represent the fit of the model output to the mean of the count data.

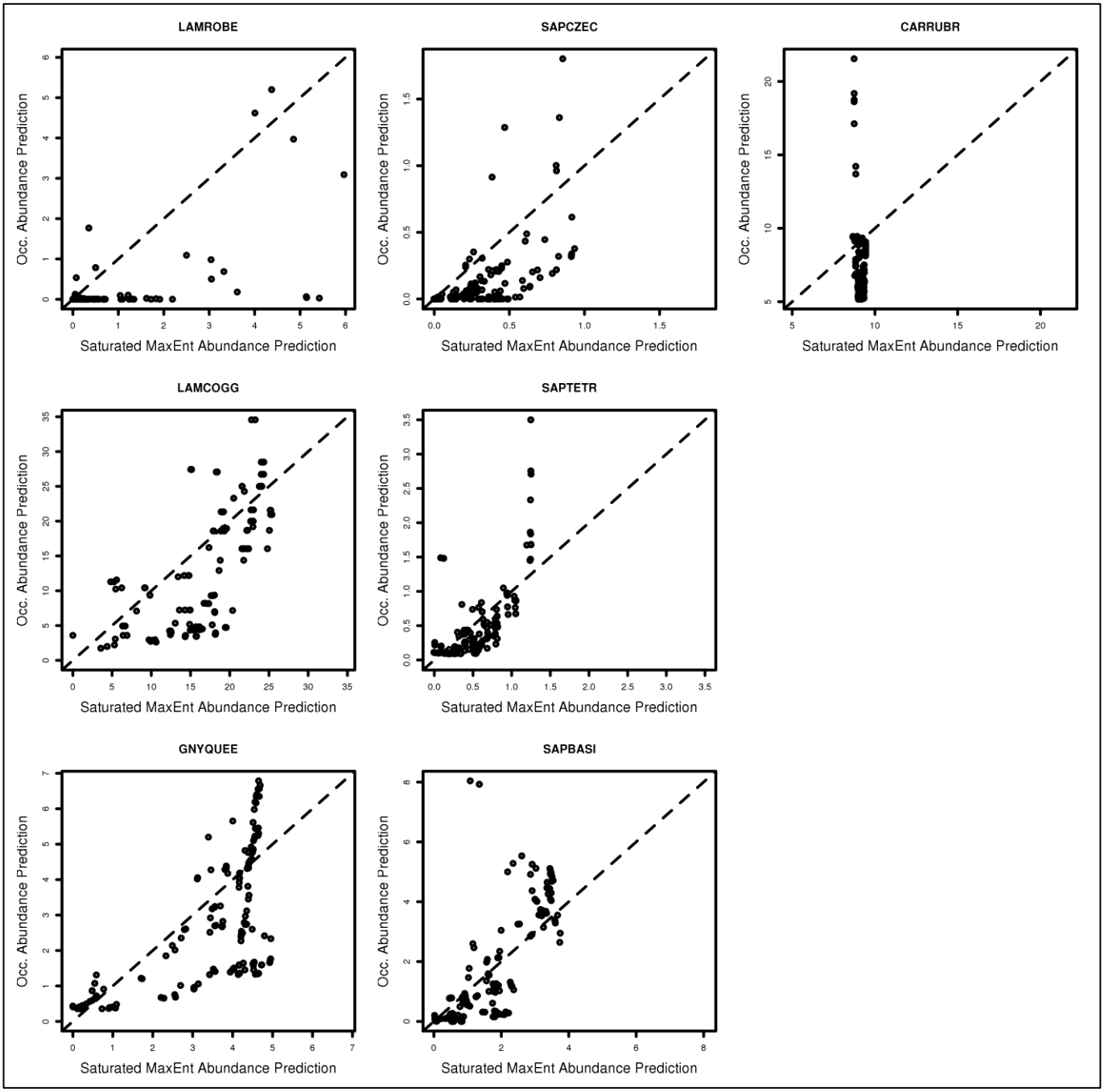



**Figure 4.7:** Comparison of abundance as predicted by the saturated MaxEnt model (x-axis) and abundance as predicted by the occupancy model (y-axis).

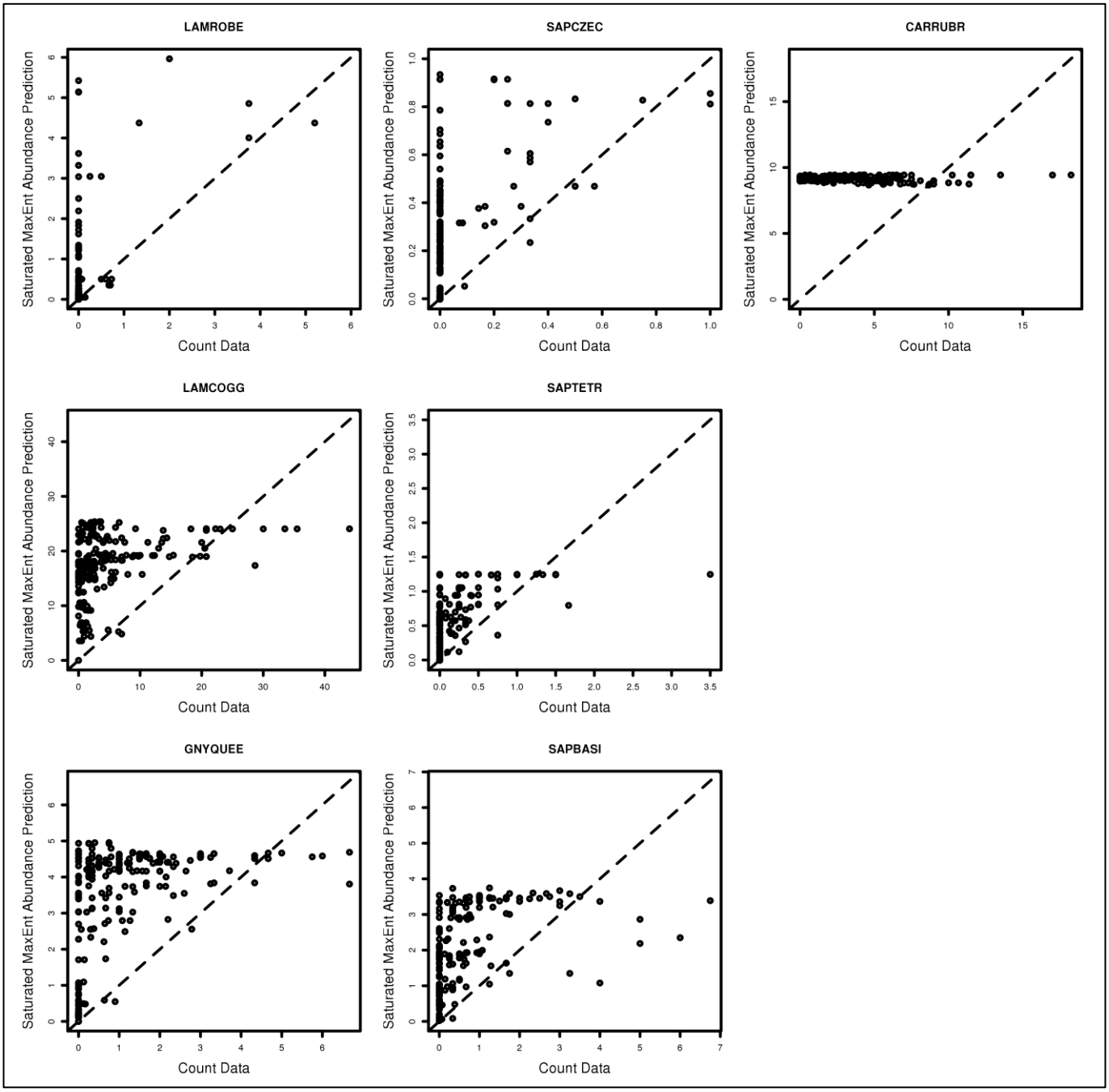



**Figure 4.8:** Comparison of measured carrying capacity (x-axis) to predicted abundance from the Saturated MaxEnt model (y-axis).

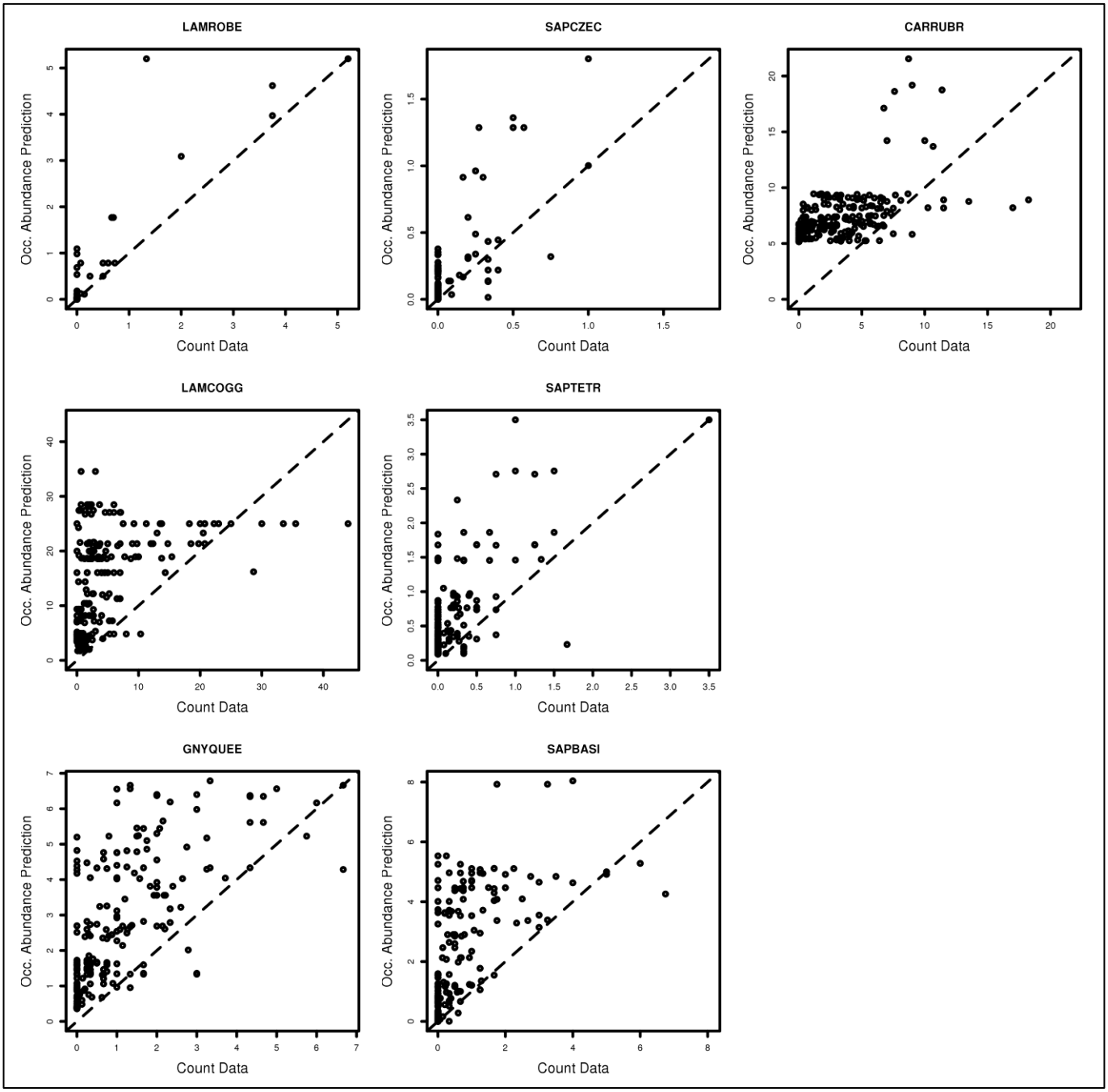



**Figure 4.9:** Comparison of measured carrying capacity (x-axis) to predicted abundance from the occupancy model (y-axis).

### *Discussion*

Increasingly, circumstances conspire to necessitate the rapid assessment of species abundance patterns across a number of biomes. Complicating this issue is the fact that many species which are believed vulnerable are little known and frequently located in areas with limited access to conservation resources. This 'perfect storm' of conditions should however not be viewed as roadblock to species' assessment. Correlative species distribution models (SDMs) are capable of extracting immense amounts of biologically relevant information from species about which otherwise little is known (Hernandez *et al.* 2006). These approaches work to estimate the relationships species share with their physical environment and the interplay between environment, individual fitness, and population vulnerability (Austin 2002, Austin 2007). Although limited in their ability to forecast into novel spatial and temporal environments, this and prior studies (VanDerWal *et al*. 2009c) have demonstrated the power of detection invariant, presence-only SDMs to predict population level features of species (i.e. carrying capacity).

'Presence only' SDMs (such as MaxEnt) rely on relatively little information to develop their predictions. All that is needed are occurrence records, as opposed to repeated sampling to demonstrate 'absence'; and spatial layers which describe environmental factors that influence fitness (Phillips *et al*. 2006, Elith *et al*. 2006). Both of these sources of data are frequently readily available, even for rare species. Furthermore, advances in statistical downscaling allow globally available coarse resolution climate layers to be utilised at a regional or local scale (Storlie *et al.* 2013, Chapter 2). 'Presence-absence' models of species abundance, on the other hand, rely on repeated surveys across seasons, years, and key environmental gradients to be appropriately parameterised.

Modelling variable detection rates for species requires repeat surveys as well as the inclusion of extra covariates in models; which will not be universally appropriate for all species under consideration (Austin 2002, Storlie *et al*. 2014, Chapter 3). For example, heliothermic species may respond to availability of sunspots on the rainforest floor, while those dwelling under fallen logs will likely not. This means parameterising models for individual species, or at least functional groups. Further, daily estimates of these variables can be difficult to find or generate. Still, logistic-regression techniques

dependent upon large presence-absence datasets may be necessary when abundance and detection co-vary strongly (MacKenzie & Kendall 2002).

Both the presence-absence and presence-only modelling techniques yield near equivalent estimates of abundance, despite differences in model structure and data for model training. This would indicate that species detection rates are not responding strongly to the environmental covariates used to model detection. While attempting to make the two models as equivalent as possible (in terms of covariates utilised) they are fundamentally different in terms of algorithms employed and their treatment of detectability. The difference between occupancy equivalent MaxEnt and Saturated MaxEnt model fit to data show that complex features and an availability of presence-only records allow presence-only models to perform almost as well as presence-absence models, when detection isn't strongly affected by the covariates.

This study demonstrates that both modelling techniques produce accurate estimates of carrying capacity for a range of species with varying ecological characteristics. Although not effective for all species modelled (e.g. *C. rubrigularis* and *L. coggeri*), this result may be explained by the geographic extent modelled for these species as opposed to the modelling techniques themselves (VanDerWal *et al.* 2009b, also see Brotons *et al*. 2004). Importantly, presence-only models of species with restricted geographic ranges and few presence records still correlate strongly with population carrying capacity. I conclude that presence-only SDMs are capable of forming near-equivalent predictions to the more data hungry presence-absence techniques and therefore are a viable alternative to predicting abundance patterns of species for which repeat survey data are not available.

# Chapter 5: The Legacy of Past Land Clearing on Climate Space: A Primer for Global Warming.

### *Introduction*

The spatial distribution of a species is often dependent upon, or correlated with, the climate to which that species is exposed. In particular, thermal exposure plays a key role in regulating metabolism which influences demographic processes including survivorship and reproductive success, amongst others (Kearney & Porter 2009). Global climate change has altered climate means and the frequency of extreme events, as well as altering the spatial distribution of climates upon which species depend (Meehl & Tebaldi 2004, IPCC 2014). As climates move through geographic space as a result of warming, concomitant shifts in species ranges (through both altitude and latitude) have been observed (Parmesan 2006). Worryingly, climate change proceeds largely unabated and is causing the disappearance of some climates at global and local scales (Burrows *et al*. 2014). Moreover, disappearing climates in tropical and subtropical regions are associated primarily with cool, upland locations (Williams *et al.* 2007), which are often associated with high levels of diversity and species endemism (Graham *et al*. 2006, VanDerWal *et al*. 2009a, Laurance *et al*. 2011b). Species associated with these climate zones will be forced then to adapt to changing circumstances, or face drastic population losses and the possibility of extinction.

Vegetation plays a key role in moderating the temperatures which species experience in the environment (Kearney *et al.* 2009). Climate models employed to predict the spatial distribution of shifting and disappearing climates are often formed at large scales and without reference to this key buffering factor (Williams *et al.* 2007, but see McAlpine 2009). At local scales, vegetation cover ameliorates extreme temperatures near the ground (Scheffers *et al*. 2013a). Rainforest canopies in particular are highly effective at buffering the land surface from heating (Shoo *et al*. 2011, Storlie *et al*. 2013, Chapter 2), and stratify temperatures experienced by species as a function of height, providing a mosaic of thermal habitats over small spatial scales (Scheffers *et al*. 2013a). At larger scales, vegetation plays a pivotal role in regulating carbon and water cycles, which in

turn influence regional climates (Pielke *et al*. 2002, Cramer *et al*. 2001). Based on this knowledge, it is reasonable to assume that deforestation may operate synergistically with climate warming to increase temperatures at the ground level, and accelerate the disappearance of already threatened climate spaces and species.

The impact or legacy of deforestation on climate space, and dependent species, will be a function of the extent and spatial pattern of deforestation with respect to extant climate space. The impacts of deforestation extend beyond the simple removal of habitat or climate space; since the spatial configuration of deforestation will influence levels of regional connectivity, fragmentation, and increase 'edge' space of available climates. The extent of fragmentation will in turn dictate species' ability to track their preferred climate through geographic space (Travis 2003, Thomas *et al*. 2012). Furthermore, the local impact of deforestation on climate may depend upon topographic factors such as elevation, slope, and aspect. These factors have a strong influence on local climate and weather patterns through their relationship with insolation and prevailing winds (Shoo *et al*. 2011). Knowledge of how environmental factors modify the impact of deforestation on climate envelopes merits explicit consideration when planning land-use or targeted restoration of vegetation to mitigate and reconnect already altered habitats.

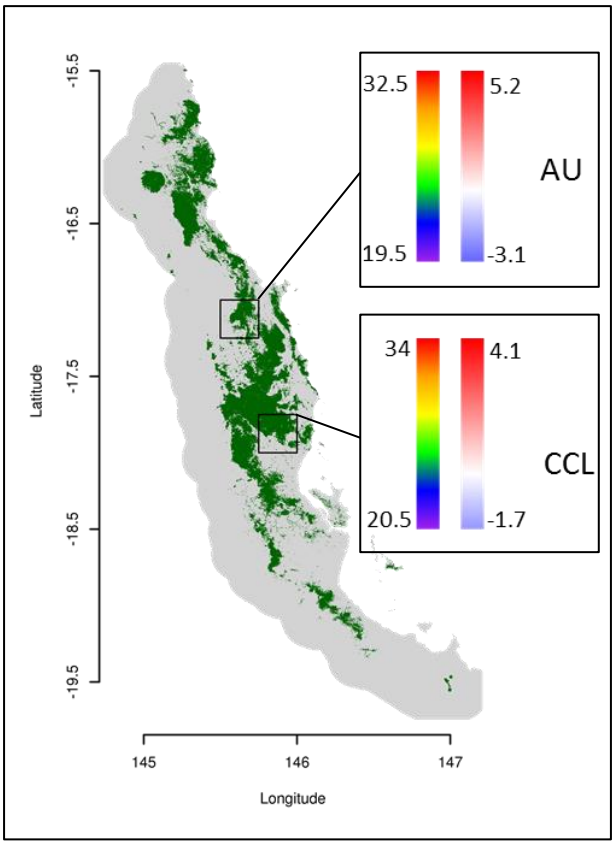
I proceed to quantify the legacy of deforestation on climate space using region wide spatial datasets for the Australian Wet Tropics – a highly diverse tropical forest system severely threatened by climate change (Williams *et al*. 2003). By combining past estimates of vegetation clearing for the region with novel statistically downscaled weather maps that are dependent upon vegetation density, I estimate the spatial pattern of daily maximum temperature ($T_{max}$) that would have existed had rainforests not been cleared from the region. I demonstrate the fragmentation and loss of connectivity between climate spaces as a result of forest clearing. Further, I describe differential effects of deforestation based on the location and extent of clearing. The alteration of climatic envelopes resulting from deforestation will exacerbate the effect of anthropogenic climate warming on species distributions leaving them further fragmented and disconnected. However, well planned targeted restoration of degraded habitats remains a viable option for restoring climate space and buying species time to adapt to changing climatic conditions.

### *Methods*

The Australian Wet Tropics (AWT) is located approximately between 15 and 20 degrees south latitude and 144.5 and 147.5 degrees east longitude (Figure 5.1). The vegetation of the AWT consists largely of a dry sclerophyll matrix which isolates mountainous rainforest patches from one another (Nix 1991). Montane rainforests are known regionally for high levels of biodiversity and endemism, associated with the cool, stable climate which dense canopy cover provides (Graham *et al*. 2006). The AWT experienced large scale land clearing (~22% of regional cover, DSITIA 2014) of rainforest areas between ca. 1750 and 1988, after which remaining rainforest areas were largely protected by declaring the area a World Heritage site. The majority (~87%) of this clearing is located in two spatially disjunct subregions, the Atherton Uplands (hereafter, AU) and the Cairns-Cardwell Lowlands (hereafter, CCL) (Figure 5.1).

My goal is to determine the legacy effect of land-clearing on available climate space for dependent species. To achieve this I combined interpolated models of weather which explicitly link vegetation cover to daily maximum temperature ($T_{max}$) with vegetation clearing scenarios. Two vegetation scenarios were employed to quantify the effect of vegetation on daily weather and available climate space; a 'preclear' scenario, which describes distribution of rainforest across the study region before European settlement (ca. 1750) and a 'remnant' scenario detailing the extant distribution of rainforests (DSITIA 2014).

Daily $T_{max}$ is estimated for both scenarios for the study region for the period 1976 to 2005 according to the method of Storlie *et al.* (2013, Chapter 2). Here, a Boosted Regression Tree statistical downscaling procedure linked 5km daily weather estimates to 250m resolution environmental factors and empirical point measurements of daily temperature (Storlie *et al*. 2013, Chapter 2). The influence of vegetation cover was modelled using a 30 year average of Foliage Projective Cover (FPC) product derived from satellite imagery for the period 1976 to 2005 (DNRW 2008). The preclear scenario required estimation of FPC for areas subject to deforestation and was calculated by randomly sampling from a frequency distribution of FPC in areas which are classified as rainforest in both the preclear and remnant scenarios.



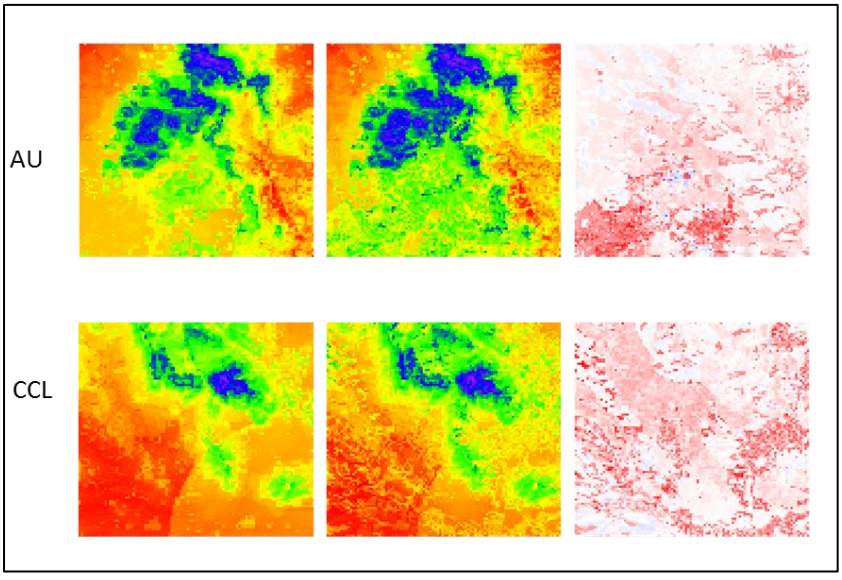




**Figure 5.1:** Left panel, green areas represent the current extent of rainforest in the AWT. The upper box within the map encompasses a portion of the AU subregion, while the lower box within the map encompasses a portion of the CCL subregion. AU subregion, left to right: Current Mean Maximum Temperature of the Warmest Month (MMTWM), Preclear MMTWM, Difference (Current – Preclear). CCL subregion, left to right: Current MMTWM, Pre-Clear MMTWM, Difference (Current – Preclear). Lefthand legends are MMTWM, righthand legends are MMTWM difference. All units are °C.

Daily weather layers for both scenarios were then aggregated to represent a 30 year average of Mean Maximum Temperature of the Warmest Month (MMTWM). I proceeded to derive estimates of percentage change in climate space (characterised by 0.5°C bands of MMTWM) using the preclear scenario as a baseline. This is done for the whole region, and separately for two heavily deforested subregions; AU and CCL. All mapped layers used in analyses are sourced or derived at a resolution of 0.0025 degrees (≈250m).

The cumulative area below each 0.5°C band of MMTWM was calculated in hectares for both scenarios of vegetation extent. Finally, metrics describing the fragmentation of thermal environments were calculated at the subregional level for both vegetation scenarios using the 'ClassStat' function in the SDMTools package (VanDerWal *et al.* 2012) in the statistical software *R* (*R* Core Team 2012).

### *Results*

Deforestation post ca. 1750 resulted in the loss of 22% of rainforest cover in the region. The losses of rainforest are not evenly distributed throughout the region, but are concentrated into two geographically distinct subregions. These were the AU and CCL subregions, which together experienced 87% of the collective deforestation within the region, corresponding to a loss of 29 and 42% of rainforested areas in these two subregions respectively (Table 5.1).

Deforestation has a discernible effect on the distribution of regional and subregional climates. At the regional level MMTWM increased by 0.67°C; greater shifts were apparent for localised areas subjected to more extensive deforestation (~0.86-0.90°C, Table 1, Figure 5.1). Deforestation results in small proportional losses of the coolest climate spaces in the region, and very large proportional gains in the hottest parts (Figure 5.1). This pattern is evident at both the regional and subregional level.

The extent to which deforestation manifested as a change in the available area below a particular maximum temperature depends on the critical thermal threshold but is non-trivial in some circumstances (Figure 5.2). For example, deforestation reduces the available area below the historical mean MMTWM for AU (i.e. 25.5°C ) and CCL (i.e. 29.5°C) by more than 50,000 ha (40%) and 60,000 ha (50%) respectively (Figure 5.3).

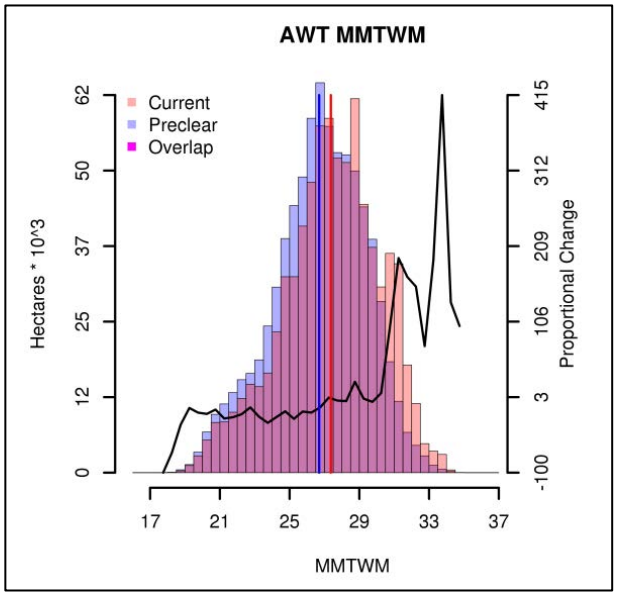
Deforestation renders thermal environments more fragmented than they were in the past (Figure 5.4). Smaller values of aggregation index and larger values of landscape division index for the remnant rainforest scenario indicate a landscape with fewer similar adjacencies of climate space. This indicates that remnant rainforest climate envelopes pose greater barriers to dispersal than those of the preclear scenario (but note the similar values of aggregation index for the CCL subregion, Figure 5.4).

**Table 5.1:** Summary of rainforest cover for preclear and remnant vegetation scenarios, separated by subregion.

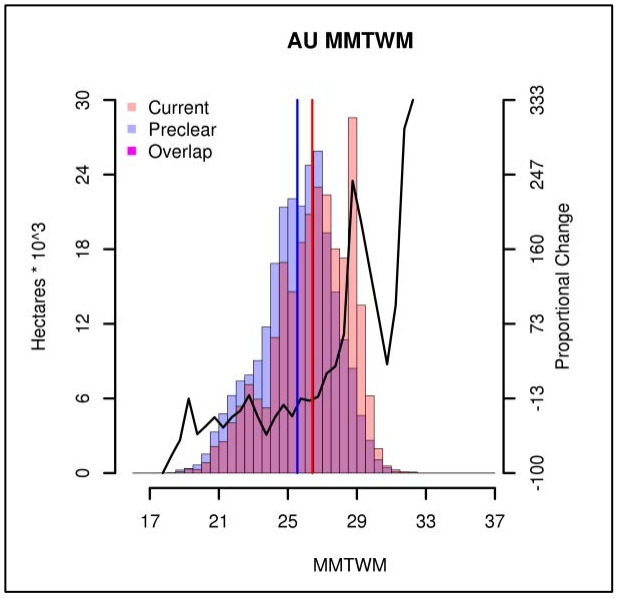
| | **Subregion** | | |
|---|---|---|---|
| | **AWT** | **CCL** | **AU** |
| Preclear Rainforest Extent (10^3 Hectares) | 770 | 192 | 247 |
| Current Rainforest Extent (10^3 Hectares) | 593 | 110 | 174 |
| Percent Deforested | 22 | 42 | 29 |
| Current MMTWM (°C) | 27.37 | 30.24 | 26.42 |
| SD of Current MMTWM (°C) | 2.81 | 1.41 | 2.19 |
| Preclear MMTWM (°C) | 26.70 | 29.33 | 25.55 |
| SD of Preclear MMTWM (°C) | 2.66 | 1.33 | 2.12 |
| Difference of MMTWM (Current – Preclear) | 0.67 | 0.91 | 0.86 |
| Difference of SD of MMTWM (Current – Preclear) | 0.15 | 0.08 | 0.08 |

**Figure 5.2:** Comparative distribution of Mean Maximum Temperature of the Warmest Month (MMTWM) for (A) the whole region, and (B-C) two subregions where deforestation has been most extensive under two alternative scenarios of vegetation extent: preclear (blue bars) and current (red bars). Solid vertical lines represent means of the respective distributions. The black line represents the proportional change in the extent of climatic envelopes characterised by 0.5°C bands, using the preclear extent as a baseline.

**A**




**B**




**C**

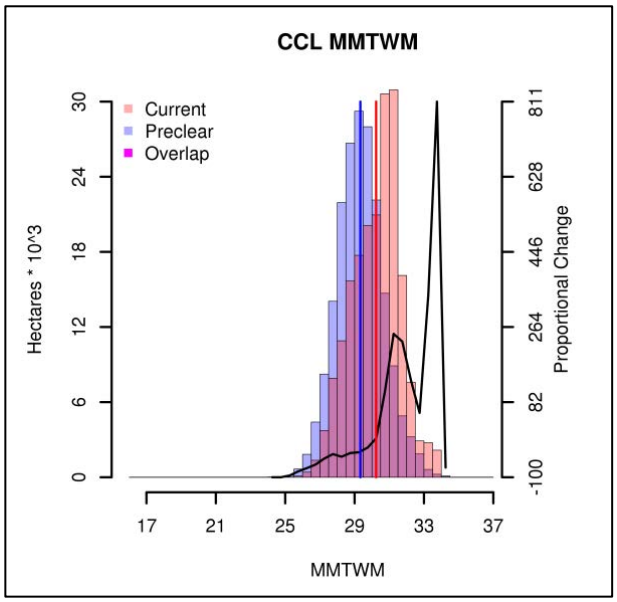
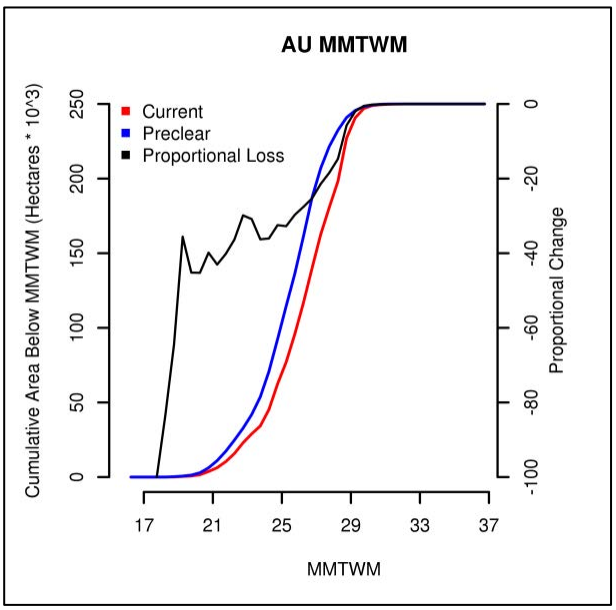
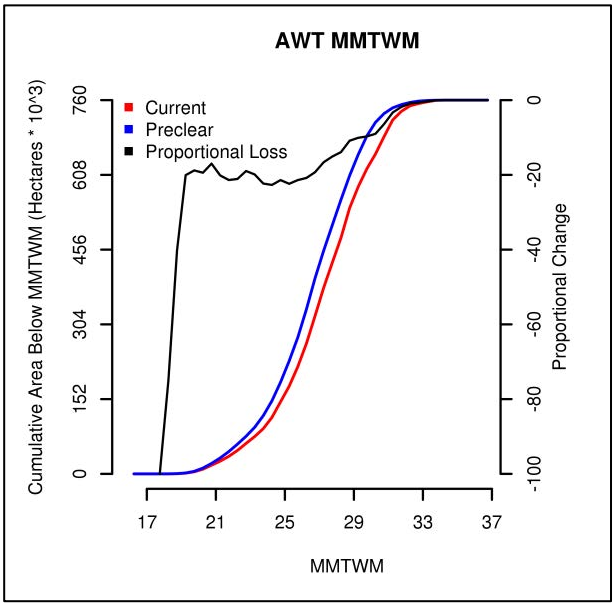


**Figure 5.3:** Impact of deforestation on available area of thermal habitat below different thresholds of Mean Maximum Temperature of the Warmest Month (MMTWM) under two alternative scenarios of vegetation extent: preclear (blue lines) and current (red lines). The black line represents the proportional change in the extent of cumulative area below a climatic threshold, using the preclear extent as a baseline.

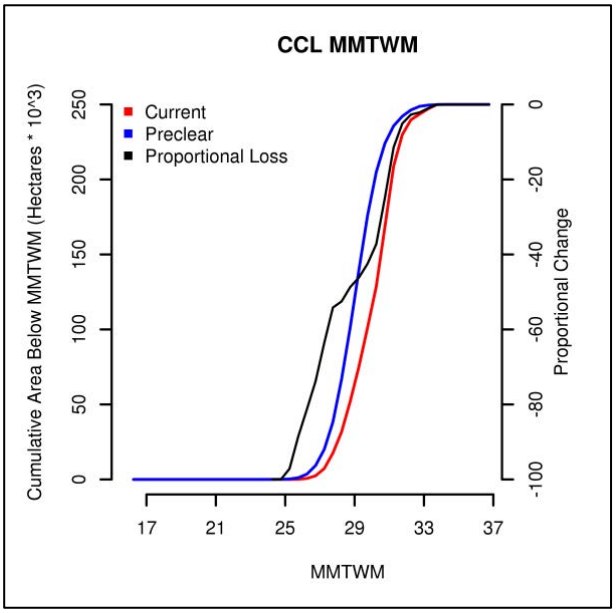
**A**




**B**




**C**



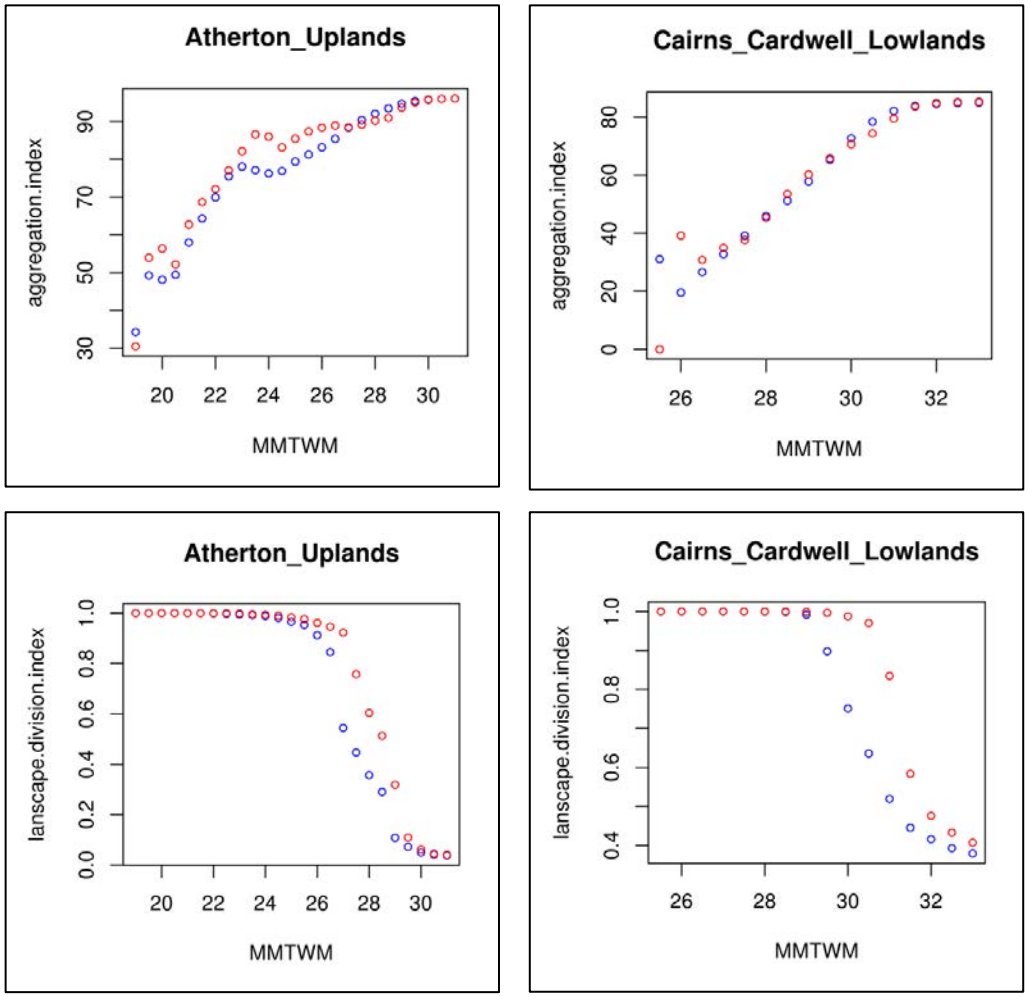



**Figure 5.4:** Impact of deforestation on the fragmentation of thermal environments as measured by Mean Maximum Temperature of the Warmest Month (MMTWM) under two alternative scenarios of vegetation extent: preclear (blue dots) and current (red dots).

### *Discussion*

By combining alternate vegetation scenarios and weather maps in a statistical downscaling framework, I demonstrate that deforestation has a strong effect on climate space in the study region. Moreover, I demonstrate that the change in climate space resulting from deforestation is one of overall warming and fragmentation; these changes act as a primer for global warming, further exacerbating its effects on species distribution especially when the region has already seen ~1°C of warming due to climate change (IPCC 2014). Although this study is limited to a small and well protected region, the AWT, it is likely the effects of deforestation on climate envelopes will scale with the proportion of the landscape which has been cleared in other less protected or more deforested areas (Bush 2002, Feddema *et al*. 2005, Bala *et al*. 2007). This indicates a clear role for forest protection in minimising potential climate change impacts on species.

The environmental conditions which species experience *in situ*, as well as their 'exposure' play a key role in determining their distributional patterns. Localised deforestation alters climatic exposure, increasing the $T_{max}$ that species experience. Thermal tolerance will then determine if a species continues to utilise the altered local climate, or in the case of thermally sensitive species, is forced to shift through geographic space to remain within its fundamental thermal niche. Changes to the exposure of a species $T_{max}$ regime will likely affect species even before they reach their critical thermal maxima. Specifically, exposure to daily weather patterns has been demonstrated to affect species distributions across seasons and years (Reside *et al*. 2010). Moreover, differential sensitivity between demographic classes means individuals within species may react differently to altered exposure (Welbergen *et al*. 2008, Sheffers *et al.* 2013b).

Deforestation has greatly reduced the climatic area available for species to utilise. The proportion by which a climatic envelope has been reduced is a function of the geographic location of clearing with respect to other environmental factors (e.g. elevation, slope, aspect). For thermally sensitive species, a reduction in the area of available climate will result in a concomitant decrease in distributional area and population size for the species. In the AWT, this is particularly concerning because the AWT is home to a variety of endemic species with narrow thermal tolerances and small

distributional areas. Prior studies suggest this is a result of the relative cool and stable climates these species have experienced over the past ~120,000 years of glacial expansion and contraction (Graham *et al*. 2010). How the loss of climatic envelopes will translate to species distributional areas is still somewhat unclear, since many species thermal tolerances are not understood (but see Krockenberger *et al*. 2012, Moritz *et al*. 2012, Storlie *et al*. 2014, Chapter 3).

The fragmentation of suitable habitat as a result of deforestation may have profound and differential impacts on species distributions. One key factor that will determine the impact of climate fragmentation on species is dispersal ability (e.g. Barton *et al*. 2012). Increased levels of fragmentation will make it more difficult for non-vagile species to track their thermal environments as conditions change (Travis 2003). Additionally, fragmentation often drives the rapid evolution of lowered dispersal, so, following fragmentation, connectivity continues to diminish over time (Murrell *et al*. 2002). Species which exhibit seasonal habitat preference may also experience climatic barriers to dispersal. Further, species which must adapt to changing climate will also have greater difficulty migrating altitudinally or latitudinally to do so (Atkins & Travis 2010). The loss of cumulative area below any particular climate threshold as a result of deforestation may also have profound impacts on species abundance patterns and their dynamics. Reduction in available climatic area will also likely result in reduced availability of resources. Decreases in population size as an effect of smaller areas and fewer resources will leave populations more vulnerable to loss of genetic diversity or extinction in smaller patches, potentially creating 'sink' populations. Moreover, an increasingly fragmented thermal landscape may result in a reduction of successful immigrants, destabilising metapopulation stability and hastening the process of extinction for populations remaining in small patches.

I show that the legacy of deforestation is one that is likely to exacerbate range losses and accelerate projected impacts of climate change. It is notable that shifts in temperature resulting from deforestation are on par with, but additional to, current levels of global warming in the region (IPCC 2014). Results may be conservative because our local climate models do not yet account for other factors such as moisture stress or climate-land surface feedbacks (McAlpine *et al.* 2009). Nonetheless, our results reaffirm the importance of retaining extant vegetation cover (e.g. Mackey *et al*.

2014), but also point to the importance of reinstating vegetation to reduce range losses and buy time for climate change adaptation. Clearly, there are major barriers to realising large scale restoration needed to effectively reverse the legacy of deforestation (e.g. resources and land-use). Even in this well protected system, the extent of deforestation (~177,000 hectares) greatly exceeds the scale of most current reforestation initiatives – which are typically a few hectares in extent (Catterall & Harrison 2006). However, a more tangible goal might be to increase the local extent of cool environments that are expected to have a strong bearing on the persistence of species within contracted ranges, or to revegetate corridors and thereby decrease fragmentation and reduce climatic barriers to dispersal (e.g. Shoo *et al*. 2011).

# Chapter 6: General Discussion & Synthesis

While there exists an immense variety of spatial data available for use in ecological analyses (in particular gridded climate and vegetation data) its use must be considered carefully before any analysis is undertaken. I have demonstrated that 'available' data and 'appropriate' data are not synonymous. For example, distributional areas predicted by SDMs using freely available spatial layers showed little difference when modelled with more accurate, non-spatially biased data (Chapter 2). However, the same freely available layers would indicate entire populations of species thriving outside their fundamental thermal niche (Chapter 3). In this instance, improving data before modelling had a small effect on model outcome, but a large effect on increasing concordance with ecological theory. This is not to say that currently existing resources such as AWAP (Jones *et al*. 2009) and ANUCLIM (Xu & Hutchinson 2011) layers are without use in deriving species-environment relationships. My work has shown that currently available resources such as these are vital stepping stones on the path to deriving accurate and biologically relevant data for spatially explicit ecological questions. Moreover, statistical downscaling techniques which improve the accuracy of these source data reveal important associations between environmental factors, such as vegetation cover and local climatic regimes.

## Summary of Major Findings

### *Statistical Downscaling to Increase Accuracy and Reduce Bias in Climate Layers*

I clearly demonstrate that spatial weather estimates created at a fine resolution (250m or finer), are unlikely to be accurate in topographically complex systems unless they take into account key factors which decouple local and regional processes (Figure 2.2). Moreover, the bias displayed by current best practice climate layers is not distributed equally in geographic space (Chapter 2, Figure 2.3). While local and regional climatic patterns are tightly correlated, vegetation is an important factor that buffers localities from high temperatures (Chapter 2, Results, Shoo *et al*. 2011). Elevation is also identified as an important factor that decouples broad and fine scale climate (Chapter 2,

Results, Shoo *et al*. 2011). However, this bias is likely associated with the placement of weather recording stations and a local association between elevation and dense rainforest vegetation.

In a statistical downscaling framework, spatial variation in climate is no longer determined solely by Euclidean distance (as is the case for the thin-plate spline technique of ANUCLIM). Instead, spatial correlations of climate are mediated by a number of environmental factors, resulting in climate models with greater levels of turnover over smaller distances than a spline technique. Stated simply, statistically downscaled weather and climate layers reveal the environment to be more heterogeneous than previously estimated. The environmental factors which structure thermal regimes at a local scale will likely depend on the peculiarities of the landscape, and as such, statistical downscaling of climate models in other regions may rely upon environmental factors that differ from those selected here. The necessity of downscaling climate depends upon the scale of the analysis in question, for example, the costs of improving continental scale assessments of biodiversity patterns by downscaling climate to 250m resolution may exceed the benefits of improvement in predictive strength at a finer scale.

The spatial bias of climate layers identified herein will not be confined to this study region only, but to other systems which have similar environmental characteristics or sparsely placed climate observatories (e.g. Chen *et al.* 1999, Scherrer *et al.* 2011). Statistical downscaling represents an excellent platform to reduce spatial bias, but the capacity for downscaling to improve weather layers is still heavily dependent upon saturating key environmental gradients with climate recording stations. When identified spatial bias of climate layers is minimal, statistical downscaling may be unnecessary, even when accuracy of climate layers to represent empirically measured temperatures is low. When inaccurate, but non-biased climate layers are used in SDM the species-environmental responses (SERs) identified will be incorrect; but they should not affect the spatial distribution predicted when mapped back onto the climate layers used to construct them.

### *Impacts of Spatially Biased Climate Layers on SDM Outcomes*

Spatially biased climate layers cause two primary problems when implemented in SDM analyses: Firstly, they mischaracterise the SERs if the species in question is distributed across the gradient of bias. Secondly, and as a result of mischaracterising SERs, these SDMs are likely to incorrectly estimate the species distribution in geographic space. My research reveals that biased spatial climate layers result in species distributions which are of approximately similar distributional area, but which differ markedly in their shapes, level of fragmentation, and connectedness when compared with distributions predicted by non-biased spatial layers (Figure 2.4).

When the goal of an SDM exercise is to gain further understanding of how species' utilise environmental space, the use of spatially biased climate layers to infer SERs is not advisable. Biased SERs skew our understanding of a species niche, potentially leading to misinterpretation of environmental factors which limit abundance and distribution of species. Correlative SDMs may be used to gain an appreciation of environmental factors which limit species fitness, distribution, or abundance before undertaking more complex mechanistic models. In these cases, spatially biased climate layers may misidentify the magnitude of a variables importance, resulting in parameterising physiological models that may have limited influence over individual fitness and hence species distributional limits. Hence, biased SERs and altered distributional patterns may flow on to affect other follow-on analyses.

Species vulnerability is strongly associated with the spatial structure of their distribution. When other factors like distributional area and population abundance are equal, a more fragmented, irregular, or less connected distribution indicates a species or population at a higher level of threat (Laurance *et al.* 2002, Laurance *et al.* 2011a). Fragmented or irregular distributions will possess more 'edges', which frequently display reduced environmental suitability for species when compared with locations within the core of their range. Further, less connected populations may suffer from decreased gene flow (Travis 2003, Dixo *et al*. 2009), potentially leading to more rapid shifts in population genetic structure. A loss of connectivity will also suppress dispersal, making immigrants and emigrants less common (Murrell *et al*. 2002). This may result in localised extinctions, particularly when isolated populations have a growth rate less than 1 (sink populations). Disconnected landscapes will also make it more

difficult for species to track their thermal environmental niche across altitudes and latitudes as climate warming becomes more prevalent (Travis 2003). Much of the land within the study region I have chosen is already protected with a network of national parks, and even though species distributions are predicted to be more fragmented when based on downscaled climate layers, further threats from deforestation and land-clearing are minimal. However, where habitat fragmentation and deforestation exist at higher levels or continue unabated, this newer more patchy understanding of species distributions could further exacerbate already existing effects of habitat clearing and warming climate on species distributions.

### *Spatially Biased Layers as a Source of Exposure Data*

In Chapter 2, I demonstrate that complex statistical models (BRTs) can relate broad scale weather layers to empirically measured temperatures via fine grained environmental layers, producing accurate estimates of daily thermal regimes. In Chapter 3, I demonstrate that simpler statistical models (OLS regression) can be used to relate ambient air temperature to thermal regimes of nearby microhabitats (Figure 3.1). This allows for the production of spatially explicit maps of microhabitat thermal regimes, which unlike their non-downscaled counterparts, place species firmly within their range of thermal tolerance (Figure 3.2). Downscaling from broad scale weather layers directly to microhabitats will not only involve a greater number of more complex models, but may also be ineffective at producing spatially explicit estimates. This may arise when local factors responsible for decoupling microhabitat temperatures from broad scale ambient weather patterns are unavailable in a spatially explicit form.

Spatial layers of climate and weather have applications beyond SDMs that may also be affected by the application of biased spatial layers. When constructing vulnerability analyses for species, many conventional approaches call for estimates of both species environmental exposure and their sensitivity or capacity for adaptation in the face of change. Thermal niche breadth of a species may be misrepresented by inaccurate exposure data, resulting in species which seem to be living closer to their upper thermal tolerance than they truly are. In this instance, biased exposure data combined with accurate sensitivity data will likely result in false inference of species vulnerability. This is problematic, since conventional vulnerability frameworks are frequently used to assess species threat to a number of processes and assign conservation priorities. In this

way, spatially biased weather layers as sources of exposure data may contribute to wasted conservation prioritisation efforts and resources.

Downscaling weather, as opposed to climate layers, is a key component of generating exposure data which are biologically relevant. Utilising a biological criterion to test the accuracy of spatial layers as exposure data would have not been possible utilising temporally aggregated climate data. For example, an area with an average maximum temperature of 30°C will not likely be suitable for an animal with a $CT_{max}$ of 30°C. Climate layers smooth over weather anomalies (hot and cold days), reducing measured environmental variance, and falsely indicating suitable habitat for species. Weather layers at a daily time scale however, compress environmental variance far less, and are a better indicator of species potentially suitable habitat as a function of their instantaneous interactions with weather conditions.

### *Predicting Patterns of Abundance with Presence Only Species Distribution Models and Downscaled Climate Layers*

Presence-only based SDM techniques can be used to predict population level metrics of species (carrying capacity), though they do not fit the empirically measured carrying capacity data as well as presence-absence techniques (Figures 4.2 and 4.3). Presence-only and presence-absence techniques both display overwhelmingly positive relationships with carrying capacity. In this instance, it seems downscaled climate layers did not improve the predictive power of the models with respect to carrying capacity, or improve their fit to the occupancy data used to construct the model (Appendix S5, Figures S5.1-S5.5).

Good estimates of abundance were attained for species with otherwise disparate ecological roles, including those that are geographically rare or hard to detect. Skinks used as study species in Chapters 2 and 4 have varying microhabitat preferences, ranging from sunspot utilisation, to shady leaf litter dwellers, and those dwelling in cool, sheltered microhabitats. Yet, without the use of species-specific environmental layers, quantile regression models of abundance (based on both presence-only and presence-absence model output) achieve a good fit to the data. These techniques, if further refined, may represent an alternative approach to parameterising species-specific demographic models, which rely on estimating fecundity and survivorship for multiple

demographic classes within a species (Akcakaya 2000). Results may improve further by downscaling air temperatures to the microhabitat of preference for species, or perhaps developing spatial layers that describe other finely resolved environmental variables, such as sunspot availability on the forest floor.

The lack of improvement in model fit when utilising downscaled weather layers may indicate that species responses to climate and weather regimes are mediated via another mechanism, such as the availability of food resources or activities of competitors, which may themselves be responding in a non-linear fashion with thermal regimes (Austin 2007). We may expect thermal regimes to have a strong effect on abundance when species are near the limit of their thermal tolerance, but this effect should be reduced when a species environment is near their thermal optima, or tolerant of a wide range of thermal conditions.

Although presence-only SDM outputs correlate less strongly with empirical abundance data than outputs from the presence-absence occupancy models, they may still be a powerful tool for predicting carrying capacity when detection based datasets are not available. The potential to predict abundance with readily available presence-only datasets saves valuable conservation resources and time. Further, the carrying capacities of geographically rare species (e.g. *S. czechurai a*nd *L. robertsi*) are predicted with high accuracy, indicating a potential to more easily assess the distribution of rare or cryptic species.

### *The Role of Vegetation in Buffering Species from a Warming World*

Removal of rainforest vegetation has increased temperatures which species are exposed to in certain localities (Figure 5.1). At a subregional level, the impact of deforestation is nearly as severe as climate warming during the last century (Figure 5.2, IPCC 2014). Furthermore, deforestation has fragmented climate envelopes available for species to utilise, leaving the thermal landscape less connected that it was before land-clearing began (Figure 5.3). This indicates that climate warming and deforestation are operating synergistically to alter thermal environments, decreasing available habitat and time for adaptation for thermally sensitive species. The placement of vegetation clearing within the landscape is of particular importance. Clearing of forest in the upland regions of the study area has contributed to the loss of climate space which has been identified as a

potential for refugia for species under global warming scenarios (Shoo *et al*. 2011). While clearing in lowland areas of the study region may not immediately affect thermally tolerant lowland species, it may eventually accelerate the process of lowland biotic attrition as climate warming continues.

All species possess a minimum distributional area required to maintain a population growth rate greater than 1 (Schaffer 1981), and therefore to avoid sliding into extinction. The cumulative area approach utilised herein gives a strong indication of the level of thermal adaptation a species must undergo as a result of deforestation, based upon its thermal tolerance and minimum distributional area. In this way, we can measure the rate at which species must adapt to changing climate, and determine what (if any) measures are necessary to help a species adapt *in situ*. Altered thermal regimes will reinforce barriers to dispersal put in place by habitat removal, potentially resulting in the formation of small isolated populations that ultimately become sinks (Murrell *et al*. 2002).

The time frame for adaptation to altered thermal regimes resulting from deforestation and global warming via evolution is likely too small for many vertebrate species (Quintero & Wiens 2013). This indicates that the restoration of forest landscapes may become a necessary tool to buy time for species to adapt to a warming world. The scale of land-clearing, however, is orders of magnitude larger than most targeted restoration activities (Catterall & Harrison 2006). The use of statistically downscaled weather layers which specifically link thermal regimes to vegetation cover allows us to identify the areas which are most severely affected. We may then proceed to cross-reference these with the areas of greatest ecological significance, and target restoration activities there to minimise fragmentation and maximise connectivity.

## Future Directions

This research highlights methods to improve spatial climate and weather layers for ecological studies and also reveals avenues for further exploration. In particular, spatial models of weather need to relate thermal regimes to position within the habitat (edge vs. core) to properly describe edge effects on climate. This sort of technique will provide even more detailed information of land surface / climate feedbacks and in turn a more accurate picture of how vegetation regulates climate locally and regionally.

The demonstration of SDMs capability to predict carrying capacity for multiple species provides potential for rapidly assessing abundance of threatened species. However, the veracity of these models could not be ascertained without a large datasets of species counts covering geographic, environmental, and temporal gradients. The capacity of SDMs to predict abundance may be improved by demonstrating concordance of modelled relationships to abundance within taxa or functional groups. This may help in forming predictions of abundance for species about which otherwise little is known, while still maintaining a measure of confidence in the predictions.

## Concluding Remarks

The costs and benefits of selecting algorithms and input data for observational modelling studies depend largely on the question being asked. Complex models with rigorous statistical assumptions are not a requisite when attempting to derive spatial properties of species distributions or their associations with the environment. Simple models can extract meaningful information concerning species distributions and abundances from the patterns found in presence-only datasets, without restrictive assumptions on the data. The accuracy of the data used as inputs in models must be examined carefully, however, as inaccurate spatial data can lead to false inference of species-environmental responses and mischaracterisation of a species fundamental niche.

Threats to species and ecosystems persist and multiply, as do methods to assess and mitigate these threats. Spatial climate and weather data represent a powerful resource for determining how species interact with their environment, a key aspect of vulnerability assessment and SDMs. Statistical downscaling methods provide new insights into the links between the physical environment and the thermal regimes to which species are exposed. Downscaling also reduces spatial bias in weather and climate layers, making them more accurate at fine resolution, and creating biologically relevant sources of exposure data.

# References


Akçakaya, H.R. 2000. Population viability analyses with demographically and spatially structured models. *Ecological Bulletins*, **48**:23-38.

Anderson R.P. and I. Gonzalez Jr. 2011. Species-specific tuning increases robustness to sampling bias in models of species distributions: An implementation with Maxent. *Ecological Modelling*, **222**:2796-2811.

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

Ashcroft, M.B. 2006. A method for improving landscape scale temperature predictions and the implications for vegetation modelling. *Ecological Modelling*, **197**:394–404.

Ashcroft, M.B., M. Cavanagh, M.D.B. Eldridge, and J.R. Gollan. 2014. Testing the ability of topoclimatic grids of extreme temperatures to explain the distribution of the endangered brush-tailed rock-wallaby (*Petrogale penicillata*). *Journal of Biogeography*, **41**(7):1402-1413.

Atkins, K.E., and J.M.J. Travis. 2010. Local adaptation and the evolution of species' ranges under climate change. *Journal of Theoretical Biology*, **266**:449-457.

Austin, M. 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological Modelling*, **157**(2-3):101–118.

Austin, M. 2007. Species distribution models and ecological theory: A critical assessment and some possible new approaches. *Ecological Modelling*, **200**:1-19.

Austin, M.P., A.O. Nicholls, M.D. Doherty, and J.A. Meyers. 1994. Determining species response functions to an environmental gradient by means of a beta function. *Journal of Vegetation Science*, **5**:215–228.

Bala, G., K. Caldeira, M. Wickett, T.J. Phillips, D.B. Lobell, C. Delire, and A. Mirin. 2007. Combined climate and carbon-cycle effects of large-scale deforestation.

*Proceedings of the National Academy of Sciences of the United States of America*, **104**(16):6550-6555.

Bartoń, K.A., T. Hovestadt, B.L. Phillips, and J.M.J. Travis. 2012. Risky movement increases the rate of range expansion. *Proceedings of the Royal Society B: Biological Sciences,* **279**:1194-1202.

Brotons, L., W. Thuiller, M.B. Araújo, and A.H. Hirzel. 2004. Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. *Ecography*, **27**(4):437-448.

Burrows, M.T., D.S. Schoeman, A.J. Richardson, J G. Molinos, A. Hoffmann, L.B. Buckley, P.J. Moore, C.J. Brown, J.F. Bruno, C.M. Duarte, B.S. Halpern, O. Hoegh-Guldberg, C.V. Kappel, W. Kiessling, M.I. O'Connor, J.M. Pandolfi, C. Parmesan, W.J. Sydeman, S. Ferrier, K.J. Williams, and E.S. Poloczanska. 2014. Geographical limits to species-range shifts are suggested by climate velocity. *Nature*, **507**:492-495.

Bush, M.B. 2002. Distributional change and conservation on the Andean flank: a palaeoecological perspective. *Global Ecology and Biogeography*, **11**:463–473.

Cade, B.S., J.W. Terrell, and R.L. Schroeder. 1999. Estimating effects of limiting factors with regression quantiles. *Ecology*, **80**:311–323.

Catterall, C.P. and D.A. Harrison. 2006. Rainforest Restoration Activities in Australia's Tropics and Subtropics. Cooperative Research Centre for Tropical Rainforest Ecology and Management, Rainforest CRC, Cairns, Australia.

Chen, J., S.C. Saunders, T.R. Crow, R.J. Naiman, K.D. Brosofske, G.D. Mroz, B.L. Brookshire, and J.F. Franklin. 1999. Microclimate in forest ecosystem and landscape ecology. *Bioscience*, **49**:288–297.

Cowles R.B. and Bogert, C.M. 1944 A preliminary study of the thermal requirements of desert reptiles. *Bulletin of the American Museum of Natural History*, **83**:261-296.

Cramer, W., A. Bondeau, F.I. Woodward, I.C. Prentice, R.A. Betts, V. Brovkin, P.M. Cox, V. Fisher, J.A. Foley, A.D. Friend, C. Kucharik, M.R. Lomas, N. Ramankutty, S. Sitch, B. Smith, A. White, and C. Young-Molling. 2001. Global response of terrestrial

ecosystem structure and function to CO2 and climate change: results from six dynamic global vegetation models. *Global Change Biology*, **7**:357–373.

Deo, R.C., J.I. Syktus, C.A. McAlpine, P.J. Lawrence, H.A. McGowan, and S.R. Phinn. 2009. Impact of historical land cover change on daily indices of climate extremes including droughts in eastern Australia. *Geophysical Research Letters*, **36**(8):L08705.

Dixo, M., J.P. Metzger, J.S. Morgante, and K.R. Zamudio. 2009. Habitat fragmentation reduces genetic diversity and connectivity among toad populations in the Brazilian Atlantic Coastal Forest. *Biological Conservation*, **142**(8):1560-1569.

DNRW. 2008. Landcover Change in Queensland 2006–07: A Statewide Landcover and Trees Study (SLATS) Report, December 2008. Department of Natural Resources and Water, Brisbane.

DSITIA. 2014. Pre-clearing & Remnant Regional Ecosystem Map, or Broad Vegetation Group Map, Version 8.0. Online RE Map, The Department of Science, Information Technology, Innovation and the Arts, Brisbane. URL - http://www.ehp.qld.gov.au/ecosystems/biodiversity/re-broad-veg-group-request.php. Accessed on 09/06/2014.

Dobrowski, S.Z. 2011. A climatic basis for microrefugia: the influence of terrain on climate. *Global Change Biology*, **17**:1022-1035.

Domingos, P. 2012. A few useful things to know about machine learning. *Communications of the Association for Computer Machinery*, **55**(10):78-87.

Elith, J. and C.H. Graham. 2009. Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. *Ecography*, **32**:66-77.

Elith, J., C.H. Graham, R.P. Anderson, M. Dudík, S. Ferrier, A. Guisan, R.J. Hijmans, F. Huettmann, J.R. Leathwick, A. Lehmann, J. Li, L.G. Lohmann, B.A. Loiselle, G. Manion, C. Moritz, M. Nakamura, Y. Nakazawa, J.M. Overton, A.T. Peterson, S.J. Phillips, K. Richardson, R. Scachetti-Pereira, R.E. Schapire, J. Soberón, S.E. Williams, M.S. Wisz, and N.E. Zimmermann. 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**:129-151.

Elith, J. and J.R. Leathwick. 2009. Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annual Review of Ecology, Evolution, and Systematics*, **40**(1):677–697.

Elith, J., J.R. Leathwick, and T. Hastie. 2008. A working guide to boosted regression trees. *Journal of Animal Ecology*, **77**:802-13.

Feddema, J.J., K.W. Oleson, G.B. Bonan, L.O. Mearns, L.E. Buja, G.A. Meehl, and W.M. Washington. 2005. The importance of land-cover change in simulating future climates. *Science*, **310**(5754):1674-1678.

Fiske I, R. Chandler, D. Miller, A. Royle, and M. Kery. 2012. unmarked: Models for Data from Unmarked Animals, R-Package version 0.9-9. URL - http://CRAN.R-project.org/package=unmarked

Flemons, P., R. Guralnick, J. Krieger, A. Ranipeta, and D. Neufeld. 2007. A web-based GIS tool for exploring the world's biodiversity: The Global Biodiversity Information Facility Mapping and Analysis Portal Application (GBIF-MAPA). *Ecological Informatics*, **2**(1):49-60.

Furrer, R., D. Nichka, and S. Sain. 2012. Fields: Tools for Spatial Data, R-Package version 6.7. URL - http://CRAN.R-project.org/package=fields

Gogol-Prokurat, M. 2011. Predicting habitat suitability for rare plants at local spatial scales using a species distribution model. *Ecological Applications*, **21**:33-47.

Graham, C.H., C. Moritz, and S.E. Williams. 2006. Habitat history improves prediction of biodiversity in rainforest fauna. *Proceedings of the National Academy of Sciences of the United States of America*, **103**:632-6.

Graham, C.H., J. VanDerWal, S.J. Phillips, C. Moritz, and S.E. Williams. 2010. Dynamic refugia and species persistence: tracking spatial shifts in habitat through time. *Ecography*, **33**:1062-1069.

GRASS Development Team. 2012. Geographic Resources Analysis Support System (GRASS) Software, Version 6.4.1. Open Source Geospatial Foundation. URL - http://grass.osgeo.org

Guisan, A., A. Lehmann, S. Ferrier, M. Austin, J. Overton, R. Aspinall, and T. Hastie. 2006. Making better biogeographical predictions of species' distributions. *Journal of Applied Ecology*, **43**(3):386-392

Hastie, T., R. Tibshirani, and J. Friedman. 2001. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer-Verlag, New York.

Hernandez, P. A., C.H. Graham, L.L. Master, and D.L. Albert. 2006. The effect of sample size and species characteristics on performance of different species distribution modelling methods. *Ecography*, **29**(5): 773-785.

Hijmans, R.J., S. Cameron, J. Parra, P.G. Jones, and A. Jarvis. 2005. WorldClim: Very high resolution global terrestrial climate surfaces for monthly temperature and precipitation. *International Journal of Climatology*, **25**(15):1965-1978.

Illán, J. G., D. Gutiérrez, and R.J. Wilson. 2010. The contributions of topoclimate and land cover to species distributions and abundance: fine-resolution tests for a mountain butterfly fauna. *Global Ecology and Biogeography*, **19**:159-173.

IPCC. 2014. *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* [Field, C.B., V.R. Barros, D.J. Dokken, K.J. Mach, M.D. Mastrandrea, T.E. Bilir, M. Chatterjee, K.L. Ebi, Y.O. Estrada, R.C. Genova, B. Girma, E.S. Kissel, A.N. Levy, S. MacCracken, P.R. Mastrandrea, and L.L. White (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, New York, USA.

IUCN. 2012. IUCN Red List Categories and Criteria: Version 3.1. Second edition. IUCN, Gland, Switzerland and Cambridge, UK.

Jones, D.A., W. Wang, and R. Fawcett. 2009. High-quality spatial climate data-sets for Australia. *Australian Meteorological and Oceanographic Journal*, **58**:233-248.

Kapos, V. 1989. Effects of isolation on the water status of forest patches in the Brazilian Amazon. *Journal of Tropical Ecology,* **5**:173-185.

Kearney, M. 2006. Habitat, environment, and niche : what are we modelling? *Oikos,* **116**(1):186-191

Kearney, M. and W.P. Porter. 2004. Mapping the fundamental niche: physiology, climate, and the distribution of a nocturnal lizard. *Ecology*, **85**:3119–3131.

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

Kearney, M., R. Shine, and W. P. Porter. 2009. The potential for behavioral thermoregualtion to buffer "cold-blooded" animals against climate warming. *Proceedings of the National Academy of Sciences of the United States of America*, **106**(10):3835-3840.

Kearney, M., A.P. Isaac, and W.P. Porter. 2014. microclim: Global estimates of hourly microclimate based on long-term monthly climate averages. *Scientific Data*, **1**:14006.

Kery, M. and B.R. Schmidt. 2008. Imperfect detection and its consequences for monitoring for conservation. *Community Ecology*, **9**:207–216.

Krockenberger, A., W. Edwards, and J. Kanowski. 2012. The limit to the distribution of a rainforest marsupial folivore is consistent with the thermal intolerance hypothesis. *Oecologia*, **168**:889-899.

Laurance, W.F. 2000. Do edge effects occur over large spatial scales? *Trends in Ecology and Evolution*, **15**:134–135

Laurance, W.F., T.E. Lovejoy, H.L. Vasconcelos, E.M. Bruna, R.K. Didham, P.C. Stouffer, C. Gascon, R.O. Bierregaard, S.G. Laurance, and E. Sampaio. 2002. Ecosystem Decay of Amazonian Forest Fragments: a 22-Year Investigation. *Conservation Biology*, **16**:605-618.

Laurance, W.F., J.L.C. Camargo, R.C.C. Luizão, S.G. Laurance, S. L. Pimm, E.M. Bruna, P.C. Stouffer, G.B. Williamson, J. Benitez-Malvido, H.L. Vasconcelos, K.S. VanHoutan, C.E. Zartman, S.A. Boyle, R.K. Didham, A. Andrade, and T.E. Lovjoy. 2011a. The fate of Amazonian forest fragments: a 32-year investigation. *Biological Conservation*, **144**(1): 56-67.

Laurance, W.F., D. Carolina Useche, L.P. Shoo, S.K. Herzog, M. Kessler, F. Escobar, G. Brehm, J.C. Axmacher, I.C. Chen, L.A. Gámez, P. Hietz, K. Fiedler, T. Pyrcz, J. Wolf, C.L. Merkord, C. Cardelus, A.R. Marshall, C. Ah-Peng, G.H. Aplet, M. del Coro Arizmendi, W.J. Baker, J. Barone, C.A. Brühl, R.W. Bussmann, D. Cicuzza, G. Eilu, M.E. Favila, A. Hemp, C. Hemp, J. Homeier, J. Hurtado, J. Jankowski, G. Kattán, J. Kluge, T. Krömer, D.C. Lees, M. Lehnert, J.T. Longino, J. Lovett, P.H. Martin, B.D. Patterson, R.G. Pearson, K.S.H. Peh, B. Richardson, M. Richardson, M.J. Samways, F. Senbeta, T.B. Smith, T.M.A. Utteridge, J.E. Watkins, R. Wilson, S.E. Williams, and C.D. Thomas. 2011b. Global warming, elevational ranges and the vulnerability of tropical biota. *Biological Conservation*, **144**:548-557.

Loarie, S.R., P.B. Duffy, H. Hamilton, G.P. Asner, C.B. Field, and D.D. Ackerly. 2009. The velocity of climate change. *Nature*, **462**(7276):1052-1055.

Lookingbill, T. and D. Urban. 2003. Spatial estimation of air temperature differences for landscape-scale studies in montane environments. *Agricultural and Forest Meteorology*, **114**:141–151.

Luoto, M. & R.K. Heikkinen. 2008. Disregarding topographical heterogeneity biases species turnover assessments based on bioclimate models. *Global Change Biology*, **14**:483-494.

Lutterschmidt, W.I. and V.H. Hutchison. 1997. The critical thermal maximum: history and critique. *Canadian Journal of Zoology*, **75**:1561-1574.

MacArthur, R.O. and E.O. Wilson. 1967. *The theory of island biogeography*. Princeton University Press, Princeton, New Jersey.

MacKenzie, D.I., J.D. Nichols, G.B. Lachman, S. Droege, J.A. Royle, and C.A. Langtimm. 2002. Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, **83**(8):2248-2255.

MacKenzie, D.I. and W.L. Kendall. 2002. How should detection probability be incorporated into estimates of relative abundance? *Ecology,* **83**(9):2387-2393.

Mackey, B., D.A. DellaSala, C. Kormos, D. Lindenmayer, N. Kumpel, B. Zimmerman, S. Hugh, V. Young, S. Foley, K. Arsenis and J.E.M. Watson. 2014. Policy options for

the world's primary forests in multilateral environmental agreements. *Conservation Letters*, doi: 10.1111/conl.12120.

McAlpine, C.A., J. Syktus, J.G. Ryan, R.C. Deo, G.M. McKeon, H.A. McGowan, and S.R. Phinn. 2009. A continent under stress: interactions, feedbacks and risks associated with impact of modified land cover on Australia's climate. *Global Change Biology*, **15**:2206-2223.

McMahon, J.P., M.F. Hutchinson, H.A. Nix, and K.D. Ord. 1995. *ANUCLIM user's guide. Version 1*. Centre for Resource and Environmental Studies, Australian National University, Canberra.

Magee, L. 1990. $R^2$ measures based on Wald and likelihood ratio joint significance tests. *The American Statistician*, **44**(3):250-253.

Meehl, G. and C. Tebaldi. 2004. More intense, more frequent, and longer lasting heatwaves in the 21st century. *Science*, **305**:994-7.

Moritz, C., G. Langham, M. Kearney, A. Krockenberger, J. VanDerWal, and S. Williams. 2012. Integrating phylogeography and physiology reveals divergence of thermal traits between central and peripheral lineages of tropical rainforest lizards. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **367**:1680-1687.

Murrell, D.J., J.M.J. Travis, and C. Dytham. 2002. The evolution of dispersal distance in spatially-structured populations. *Oikos*, **97**:229-236.

Nix, H. 1991. Biogeography: pattern and process. In: *Rainforest Animals: Atlas of Vertebrates Endemic to Australia's Wet Tropics* [Nix H.A. and M.A. Switzer (eds)], Australian National Parks and Wildlife Service, Canberra, pp. 11–40.

Parmesan, C. 2006. Ecological and evolutionary responses to recent climate change. *Annual Review of Ecology & Systematics*, **37**:637-669.

Parmesan, C. and G. Yohe. 2003. A globally coherent fingerprint of climate change impacts across natural systems. *Nature*, **421**:37-42.

Phillips, S.J., R.P. Anderson, and R.E. Schapire. 2006. Maximum entropy modelling of species geographic distributions. *Ecological Modelling*, **190**(3):231-259.

Pielke, R.A., G. Marland, R.A. Betts, T.N. Chase, J.L. Eastman, J.O. Niles, and S.W. Running. 2002. The influence of land-use change and landscape dynamics on the climate system: relevance to climate-change policy beyond the radiative effect of greenhouse gases. *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, **360**(1797):1705-1719.

Pouteau, R., S. Rambal, J. Ratte, F. Gogé, R. Joffre, and T. Winkel. 2011. Downscaling MODIS-derived maps using GIS and boosted regression trees: The case of frost occurrence over the arid Andean highlands of Bolivia. *Remote Sensing of Environment*, **115**(1):117–129.

Quintero, I. and J.J. Wiens. 2013. Rates of projected climate change dramatically exceed past rates of climatic niche evolution among vertebrate species. *Ecology Letters,* **16**(8), 1095-1103.

*R* Development Core Team. 2012. *R*: A language and environment for statistical computing. *R* Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. URL - http://www.R-project.org

Randin, C., R. Engler, S. Normand, M. Zappa, N. Zimmerman, P. Pearman, P. Vittoz, W. Thuiller, and A. Guisan. 2009. Climate change and plant distribution: local models predict high-elevation persistence. *Global Change Biology*, **15**:1557-1569.

Reside, A.E., J.J. VanDerWal, A.S. Kutt, and G.C. Perkins. 2010. Weather, not climate, defines distributions of vagile bird species. *PLoS One*, **5**(10):e13569.

Ridgeway, G. 2010. Generalized Boosted Regression Models, *R* Package version 1.6-3.1. URL - http://CRAN.R-project.org/package=gbm

Royle, J.A. and J.D. Nichols. 2003. Estimating abundance from repeated presence-absence data or point counts. *Ecology*, **84**(3):777-790

Royle, J.A., J.D. Nichols, and M. Kery. 2005. Modelling occurrence and abundance of species when detection is imperfect. *Oikos*, **110**(2):353–359.

Sagarin, R. and A. Pauchard. 2010. Observational approaches in ecology open new ground in a changing world. *Frontiers in Ecology and the Environment*, **8**(7): 379-386.

Sagarin, R. and F. Micheli. 2001. Climate change in non-traditional data sets. *Science*, **294**:811.

Shaffer, M.L. 1981. Minimum population sizes for species conservation. *BioScience*, **31**(2):131-134.

Scheffers, B. R., R.M. Brunner, S.D. Ramirez, L.P. Shoo, A. Diesmos, A., and S.E. Williams. 2013a. Thermal buffering of microhabitats is a critical factor mediating warming vulnerability of frogs in the Philippine biodiversity hotspot. *Biotropica*, **45**(5):628-635.

Scheffers, B.R., B.L. Phillips, W.F. Laurance, N.S. Sodhi, A. Diesmos, and S.E. Williams. 2013b. Increased arboreality of frogs with increasing altitude suggests a novel biogeographical dimension. *Proceedings of the Royal Society B-Biological Sciences*, **280**:20131581.

Scherrer, D., S. Schmid, S., and C. Korner. 2011. Elevational species shifts in a warmer climate are overestimated when based on weather station data. *International Journal of Biometeorology*, **55**:645-654.

Schoof, J.T. and S.C. Pryor. 2001. Downscaling temperature and precipitation: a comparison of regression-based methods and artificial neural networks. *International Journal of Climatology*, **21**(7):773-790.

Shoo, L.P., C.J. Storlie, J.J. VanDerWal, J. Little, and S.E. Williams. 2011. Targeted protection and restoration to conserve tropical biodiversity in a warming world. *Global Change Biology*, **17**:186-193.

Steiner, F.M., B.C. Schlick-Steiner, J.J. VanDerWal, K.D. Reuther, E. Christian, C. Stauffer, A.V. Suarez, S.E. Williams, and R.H. Crozier, R.H. 2008. Combined modelling of distribution and niche in invasion biology: a case study of two invasive Tetramorium ant species. *Diversity & Distributions*, **14**:538–545.

Storlie, C.J., A. Merino-Viteri, B.L. Phillips, J.J. VanDerWal, J. Welbergen, and S.E. Williams. 2014. Stepping inside the niche: microclimate data are critical for accurate assessment of species' vulnerability to climate change. *Biology Letters*, **10**(9):20140576.

Storlie C.J., B.L. Phillips, J.J. VanDerWal , and S.E. Williams 2013. Improved spatial estimates of climate predict patchier species distributions. *Diversity and Distributions*, **9**:1106-1113.

Thomas, C.D., A. Cameron, R.E. Green, M. Bakkenes, L.J. Beaumont, Y.C. Collingham, B.F.N. Erasmus, M.F. De Siqueira, A. Grainger, L. Hannah, L. Hughes, B. Huntley, A.S. Van Jaarsveld, G.F. Midgley, L. Miles, M.A. Ortega-Huerta, A.T. Peterson, O.L. Phillips, and S.E. Williams. 2004. Extinction risk from climate change. *Nature*, **427**:145-8.

Thomas, C.D., P.K. Gillingham, R.B. Bradbury, D.B. Roy, B.J. Anderson, J.M. Baxter, N.A.D. Bourn, H.Q.P. Crick, R.A. Findon, R. Fox, J.A. Hodgson, A.R. Holt, M.D. Morecroft, N.J. O'Hanlon, T.H. Oliver, J.W. Pearce-Higgins, D.A. Procter, J.A. Thomas, K.J. Walker, C.A. Walmsley, R.J. Wilson, and J.K. Hill. 2012. Protected areas facilitate species' range expansions. *Proceedings of the National Academy of Sciences of the United States of America*, **109**(35):14063-14068.

Torres, M.T., P.D.M. Junior, T. Santos, L. Silveira, A.T.A. Jacomo, and J.A.F. Diniz-Filho. 2012. Can species distribution modelling provide estimates of population densities? A case study with jaguars in the Neotropics. *Diversity and Distributions*, **18**:615-627.

Travis, J.M.J. 2003. Climate change and habitat destruction: a deadly anthropogenic cocktail. *Proceedings Of The Royal Society Of London Series B-Biological Sciences*, **270**:467-473.

Trivedi, M., P. Berry, M. Morecroft, and T. Dawson. 2008. Spatial scale affects bioclimate model projections of climate change impacts on mountain plants. *Global Change Biology*, **15**:1089-1103.

Turton, S., M. Hutchinson, A. Accad, P. Hancock, P, and T. Webb. 1999. Producing fine-scale rainfall climatology surfaces for Queensland's wet tropics region. In: *Geodiversity: Readings in Australian Geography at the Close of the 20th Century, Special Publication Series No. 6*. [Kesby, J.A., J.M. Stanley, R.F. McLean, and L.J. Olive (eds)] School of Geography and Oceanography, University College, ADFA, Canberra, pp. 415–428.

Van Couwenberghe, R., C. Collet, J-C. Pierrat, K. Verheyen, and J-C. Gegout. 2012. Can species distribution models be used to describe plant abundance patterns? *Ecography*, **36**(6):665-674.

VanDerWal, J.J., L.P. Shoo, and S.E. Williams. 2009a. New approaches to understand late Quaternary climate fluctuations and refugial dynamics in Australian wet tropical rain forests. *Journal of Biogeography*, **36**:291-301.

VanDerWal, J.J., L.P. Shoo, C. Graham, and S.E. Williams. 2009b. Selecting pseudo-absence data for presence-only distribution modeling: How far should you stray from what you know? *Ecological Modelling*, **220**:589-594.

VanDerWal, J.J., L.P. Shoo, C.N. Johnson, and S.E. Williams. 2009c. Abundance and the environmental niche: environmental suitability estimated from niche models predicts the upper limit of local abundance. *The American Naturalist*, **174**(2):282–91.

VanDerWal, J.J., L.P. Shoo, S. Januchowski, L. Falconi, and C.J. Storlie. 2012. SDMTools: Species Distribution Modelling Tools: Tools for processing data associated with species distribution modelling exercises, R-Package version 1.1.3. URL - http://CRAN.R-project.org/package=SDMTools

Welbergen, J.A., S.M. Klose, N. Markus, and P. Eby. 2008. Climate change and the effects of temperature extremes on Australian flying-foxes. *Proceedings of the Royal Society: Series B, Biological Sciences*, **275**:419-25.

Wiens, J. and C.H. Graham. 2005. Niche conservatism: Integrating evolution, ecology, and conservation biology. *Annual Review of Ecology, Evolution, and Systematics*, **36**:519–539.

Wiens, J., D. Stralberg, D. Jongsomjit, C.A. Howell, and M.A. Snyder. 2009. Niches, models, and climate change: assessing the assumptions and uncertainties. *Proceedings of the National Academy of Sciences of the United States of America*, **106**(Supplement 2): 19729-19736.

Williams, J.W., S.T. Jackson, and J.E. Kutzbacht. 2007. Projected distributions of novel and disappearing climates by 2100 AD. *Proceedings of the National Academy of Sciences of the United States of America*, **104**:5738–5742.

Williams, S.E., E.E. Bolitho and S. Fox. 2003. Climate change in Australian tropical rainforests: an impending environmental catastrophe. *Proceedings of the Royal Society of London: Series B, Biological Sciences*, **270**:1887-1892.

Williams S.E., L.P. Shoo, J. Isaac, A.A. Hoffmann, and G. Langham. 2008. Toward an integrated framework for assessing the vulnerability of species to climate change. *PLoS Biology*, **6**(12):2621-2626.

Williams, S.E., J.J. VanDerWal, J. Isaac, L.P. Shoo, C.J. Storlie, S. Fox, E.E. Bolitho, C. Moritz, C. Hoskin, C., and Y.M. Williams. 2010. Distributions, life-history specialization, and phylogeny of the rain forest vertebrates in the Australian Wet Tropics. *Ecology*, **91**:2493.

Xu, T., and M.F. Hutchinson. 2011. *ANUCLIM Version 6.1.* Fenner School of Environment and Society, Australian National University, Canberra.

# Appendix S1: Spatial Surfaces for the Boosted Regression Tree Downscale Procedure

Temperature data were collected from a network of weather stations (n=27) distributed throughout the study region across a latitudinal and altitudinal gradient. Weather stations were positioned underneath dense rainforest canopy at a height of approximately 1.3m above the ground. Measurements of temperature were recorded every 30 minutes using a HOBO 8-Bit Temperature Sensor[1] or an iButton thermochron[2] from November 2006 to June 2009. Additional rainforest sites (n=14) were also monitored intermittently over the period June 2004 to June 2009 using thermochron temperature loggers sampling at the same height and time interval. Loggers were protected from exposure to rainfall and direct sunlight by a housing which covered them from above, but allowed air to move freely around all sides. Twice hourly data were summarised to the two climate variables of interest, daily $T_{max}$ and daily $T_{min}$ for all sites. This empirical dataset was supplemented with climate data provided by the Australian Bureau of Meteorology (BoM)[3]. Daily $T_{max}$ and $T_{min}$ were obtained from all BoM weather stations in the study region (n=13) encompassing the study period from November 2006 to June 2009. The final empirical data set comprises daily measurements of $T_{max}$ and $T_{min}$ from 54 sites, representing 32,239 site-days.

A suite of 10 environmental surfaces were selected as independent variables for the BRT downscaling, based on their ability to mediate the broad-scale/fine-scale climate relationship. The broad-scale surfaces selected for the BRT downscaling procedure were daily $T_{max}$ and daily $T_{min}$ from the Australian Water Availability Project (AWAP, Jones *et al.* 2009) at a 5km resolution. These AWAP surfaces of $T_{max}$ and $T_{min}$ and were created from observations of weather by BoM stations across Australia; spatially interpolated by latitude, longitude, and elevation. AWAP surfaces were interpolated to 250m resolution before model-fitting with the 'interp.surface' function from the *R* package 'fields' (Furrer *et al*. 2012). A surface of insolation (hours of sun exposure)

[1] http://www.microdaq.com/occ/hws/micro_station.php

[2] http://www.maxim-ic.com/products/ibutton/

[3] http://www.bom.gov.au

was also calculated at 250m resolution for each day of the year using the GRASS r.sun command (GRASS Development Team 2012). A surface of elevation was obtained from Geoscience Australia at 250m resolution over the entire extent of the study region[4]. From this surface I calculated spatial surfaces of slope and aspect, using the r.slope command in GRASS. Surfaces of distance to coast and distance to stream were created using Spatial Analyst in ESRI ARCGIS and maps from Geoscience Australia's Global Map Australia 1M 2001 product[5]. Distance to stream was converted to *ln* (distance +1) to emphasise the role that very small distance from stream has in influencing microclimate patterns (Ashcroft 2006, Lookingbill & Urban 2003). A spatial surface of foliage projected cover (FPC) was obtained from the Queensland Department of Natural Resources and Water (DNRW 2008). This surface details a 30 year average of FPC on a scale of 0 (no vegetation cover) to 100 (complete vegetation cover) for the entire study region at a spatial resolution of 250m.

[4] Resampled from GEODATA 9S DEM Version 2; Geoscience Australia, http://www.ga.gov.au/

[5] https://data.gov.au/dataset/global-map-global-map-elevation-of-australia-1-million-scale-2001

# Appendix S2: Parameterising the Boosted Regression Tree Dowsnscaling Models

Two separate BRT models (one for daily $T_{min}$ and one for daily $T_{max}$) were created using the above datasets. The two BRT models use the full set of predictor variables to develop a statistical association between empirically observed $T_{min}$ and $T_{max}$ respectively. All model fitting was completed using Elith's modified *gbm.step* function (Elith *et al.* 2008, Supplementary Appendix 3) *in* the R-Statistical Software package (*R* Development Core Team 2009). Errors of the dependent data approximate a normal distribution, so a Gaussian error distribution was used in both models. The loss function chosen to evaluate model fit was Root Mean Squared Error (RMSE). Tuneable parameters in the model included bag fraction (the proportion of training data to be randomly sampled for model fitting at each iteration), interaction depth (the number of potential interactions between predictors) and learning rate (a scaling factor used to weight the contribution of each tree as it is added to the model). The bag fraction was set at a constant value of 0.5, in accordance with the recommendations from Elith *et al*. (2008). In order to select the tuneable parameters giving the best predictive power, 16 models were created for each dependent variable ($T_{max}$ or $T_{min}$), each of these models possessing a unique combination of learning rate (0.05, 0.01, 0.025 or 0.001) and tree complexity (4, 8, 12 or 16). The 'optimal model' selected was the model which had the smallest RMSE value as reported by the tenfold cross-validation procedure (Hastie *et al.* 2001). The optimal combination of learning rate and tree complexity for the BRT $T_{max}$ model were .05 and 8 respectively, yielding a model which minimised the RMSE at 1,500 trees. The optimal combination of learning rate and tree complexity for the BRT $T_{min}$ model were .05 and 16 respectively, yielding a model which minimised the RMSE at 700 trees.

# Appendix S3: Downscaling Weather Layers to Exposure Surrogates

In the AWT landscape, iButton thermochrons[6] were placed in a paired design underneath the rainforest canopy at 21 sites which largely cover key environmental gradients of latitude, elevation, and moisture regimes. Two loggers were placed in each site, one underneath a minimally disturbed *in situ* fallen log (the primary diurnal retreat of the focal species). A second logger was suspended above the forest floor, sheltered according to the method of Storlie *et al.* (2013, Chapter 2) and located within 20m of the under-log thermochrons. Loggers recorded temperature data over three and half years from December 2006 to February 2010 and the resultant dataset was aggregated to daily maximum ($T_{max}$) and minimum ($T_{min}$) temperature for each microclimate (under-log and open-air) representing 14,834 unique combinations of site and date. Simple linear regression was used to relate under-log temperature maxima to open-air $T_{max}$ and $T_{min}$ with the following equation: $T_{MAX\ UL} = 0.88*T_{MAX\ AIR} - 0.36*T_{RANGE\ AIR} - 0.003*T_{MAX\ AIR}*T_{RANGE\ AIR} + 2.632$. By substituting mapped air temperature values (Storlie *et al*. 2013, Chapter 2) into the linear model for under-log temperature, I generated maps estimating 38 years of daily under-log temperature to estimate the thermal exposure of species using this microhabitat. This form of model was chosen for two reasons; first, selected variables needed to be spatially explicit to form a prediction of exposure at sites of known species occurrence which lack paired empirical dataloggers, hence the use of daily air temperatures estimated by Storlie *et al.* (2013, Chapter 2). Second, the independent variables were selected based on the principles of process-based microclimate modelling (Kearney & Porter 2009). Atmospheric forcing (air temperature) is known to have a strong effect on microclimate temperatures (Kearney & Porter 2009). Other variables known to effect microclimate temperatures (e.g. surface albedo, radiation exposure (Kearney & Porter 2009) are regionally homogeneous and were therefore not included in the model. Ultimately, the linear

[6] http://www.maxim-ic.com/products/ibutton

microclimate model formed accurate predictions of the empirical microclimate (under-log) $T_{max}$ achieving an adjusted-$r^2$ of 0.916.

# Appendix S4: Determination of Microhylid Frog Critical Thermal Maxima

Measurement of the species' $CT_{max}$ were achieved using a dynamic methodology (Cowles & Bogert 1944) using the loss of righting response as the end point. Calling males were collected from the wild, in order to ensure accurate identification of specimens for determining physiological tolerances. Individuals were held for a short period, less than 24 hours, and the experiment was performed *in situ* to avoid complications of individuals acclimating to laboratory temperatures (Lutterschmidt & Hutchison 1997). It should be noted that this $CT_{max}$ metric may not be applicable to other age/stage classes within the focal species (e.g. eggs and juveniles). The principles of biophysical modelling (Kearney & Porter 2009) indicate that adult individuals should be less thermally sensitive than their younger counterparts. Therefore, this metric is both accurate for adult males and could be considered an underestimate of the sensitivity of the entire population.

Individuals were placed in a small cylindrical chamber, with a Sable Systems TC-2000 Thermocouple meter measuring chamber temperature continuously. The chamber was sealed and placed in a water bath which was then heated at a rate of 1°C $min^{-1}$ using a heating plate, in accordance with the recommended heating rate of Lutterschmidt and Hutchison (1997). The chamber was rotated slowly and continuously, forcing the animal to make postural corrections for the duration of the experiment. When an individual failed to initiate a righting response on two consecutive occasions the experiment was completed, and the animal was deemed to have reached its $CT_{max}$. Upon experiment completion animals were placed in a cool water bath to restore a functional operative body temperature and then released at their point of capture.

# Appendix S5: Modelling Abundance with Downscaling Climate and Weather Layers

The use of the alternate, downscaled climate layers in the modelling exercise does not improve the fit of occupancy models to the empirical occupancy data used to train them (Figure S5.1). There is some indication that the simpler, occupancy equivalent MaxEnt models have improved AUC scores when trained with downscaled climate layers, however, this is not the case with the Saturated MaxEnt model (Figure S5.2A and S5.2B). Similarly, inclusion of downscaled climate layers does not markedly improve either of the three models ability to fit the count data (Figures S5.3 to S5.5).

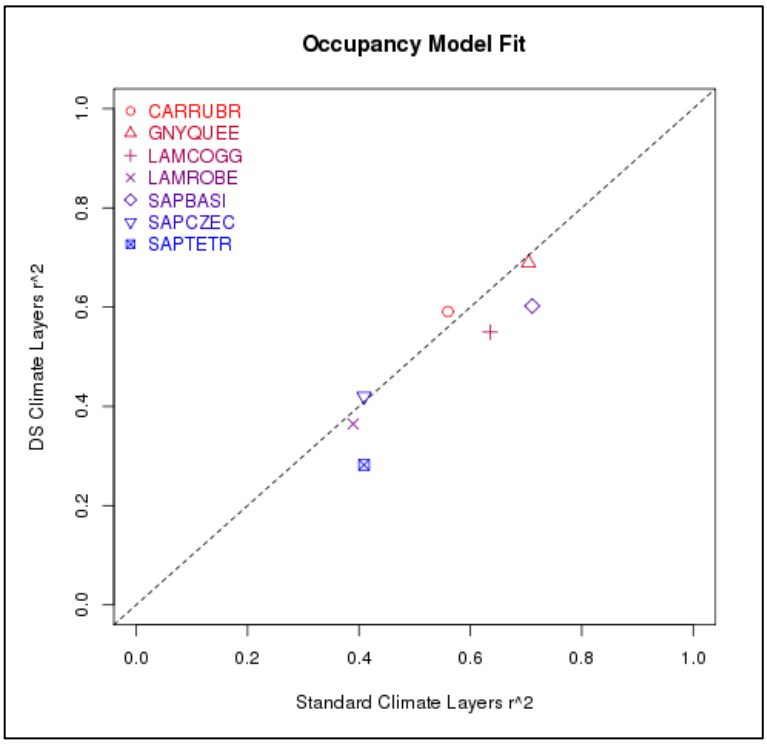
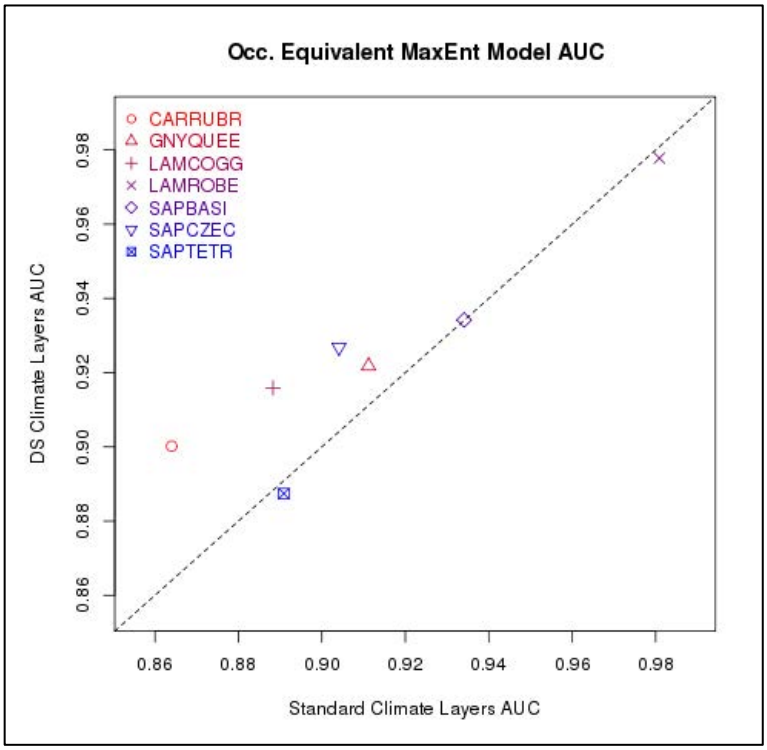



**Figure S5.1:** Comparison of pseudo $r^2$ values for the fit of occupancy modesl to the training data (species presences and absences), occupancy model $r^2$ values on the x-axis were trained on non-downscaled climate and weather layers (Jones *et al*. 2009, Xu & Hutchinson 2011), occupancy model $r^2$ values on the x-axis were trained on non-downscaled climate and weather layers (Storlie *et al*. 2013, Chapter 2).

A




B

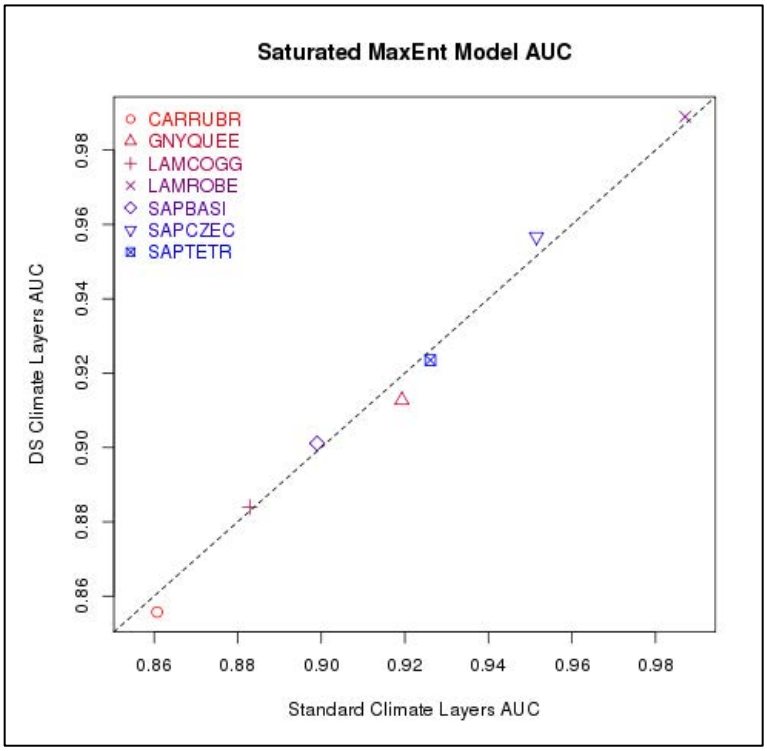



**Figure S5.2**: Fit of presence-only MaxEnt model Environmental Suitability to the training data (species occurrences); assessed using AUC scores. X-axis represents AUC scores of layers utilising non-downscaled spatial layers, y-axis represents AUC scores of models trained using the downscaled layers of Storlie *et al.* (2013, Chapter 2). (A) Occupancy equivalent MaxEnt model. (B) Saturated MaxEnt model.

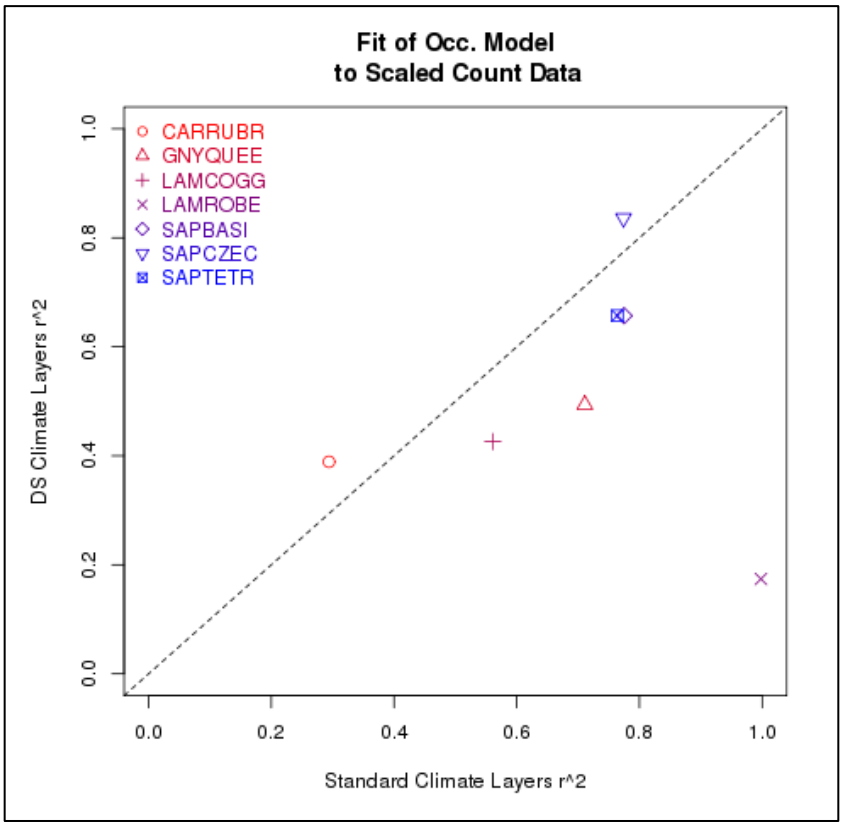



**Figure S5.3:** Comparison of pseudo $r^2$ values for the fit of occupancy model outputs to the scaled count data, occupancy model $r^2$ values on the x-axis were trained on non-downscaled climate and weather layers (Jones *et al*. 2009, Xu & Hutchinson 2011), occupancy model $r^2$ values on the x-axis were trained on non-downscaled climate and weather layers (Storlie *et al*. 2013, Chapter 2).

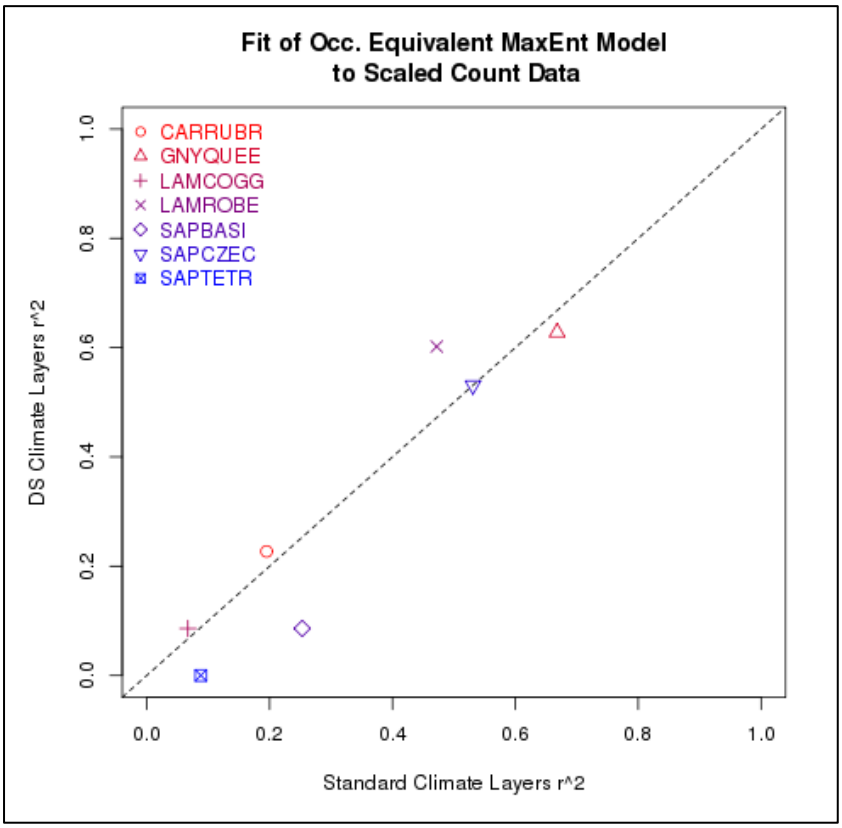



**Figure S5.4:** Comparison of pseudo $r^2$ values for the fit of occupancy equivalent MaxEnt model outputs to the scaled count data, model $r^2$ values on the x-axis were trained on non-downscaled climate surfaces (Xu & Hutchinson 2011), model $r^2$ values on the x-axis were trained on non-downscaled climate surfaces (Storlie *et al*. 2013, Chapter 2).

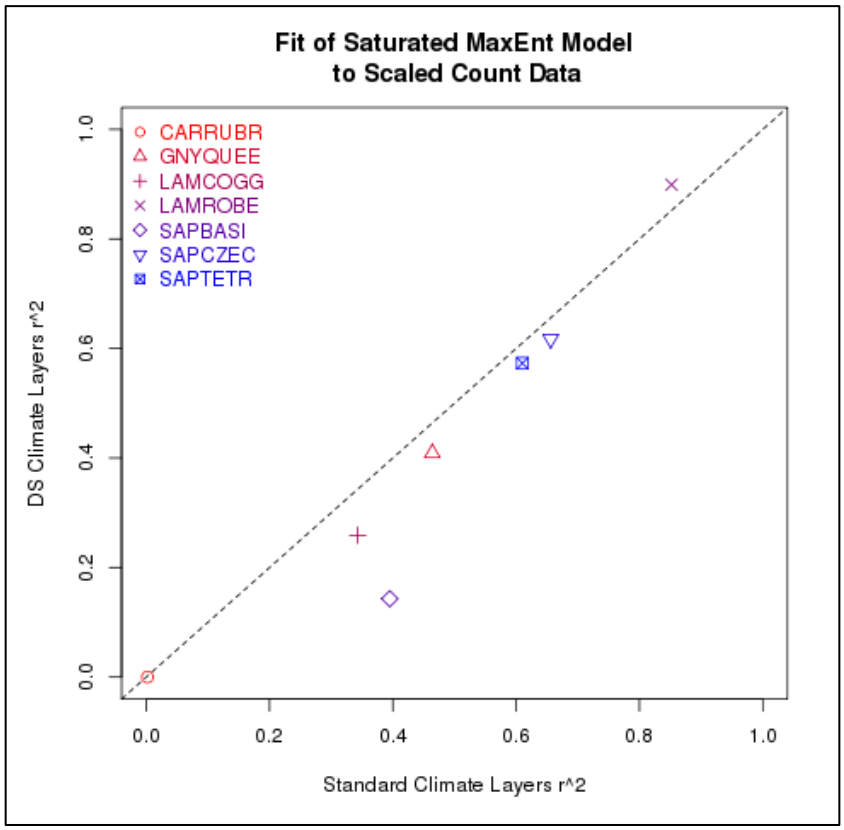



**Figure S5.5:** Comparison of pseudo $r^2$ values for the fit of Saturated MaxEnt model outputs to the scaled count data, model $r^2$ values on the x-axis were trained on non-downscaled climate surfaces (Xu & Hutchinson 2011), model $r^2$ values on the x-axis were trained on non-downscaled climate surfaces (Storlie *et al*. 2013, Chapter 2).

The relationships between model output and scaled count data are still overwhelmingly positive in nature, although the widespread species *C. rubrigularis* has negative or no relationship between both MaxEnt outputs and the count data (Figures S5.6 to S5.8). There are no clear trends between the abundance predictions of the MaxEnt models and the occupancy model, although the MaxEnt models tend to predict larger population sizes at sites than do the occupancy models (Figures S5.9). The majority of sites are predicted to experience positive departures from the estimated count data (Figures S5.10 and S5.11).

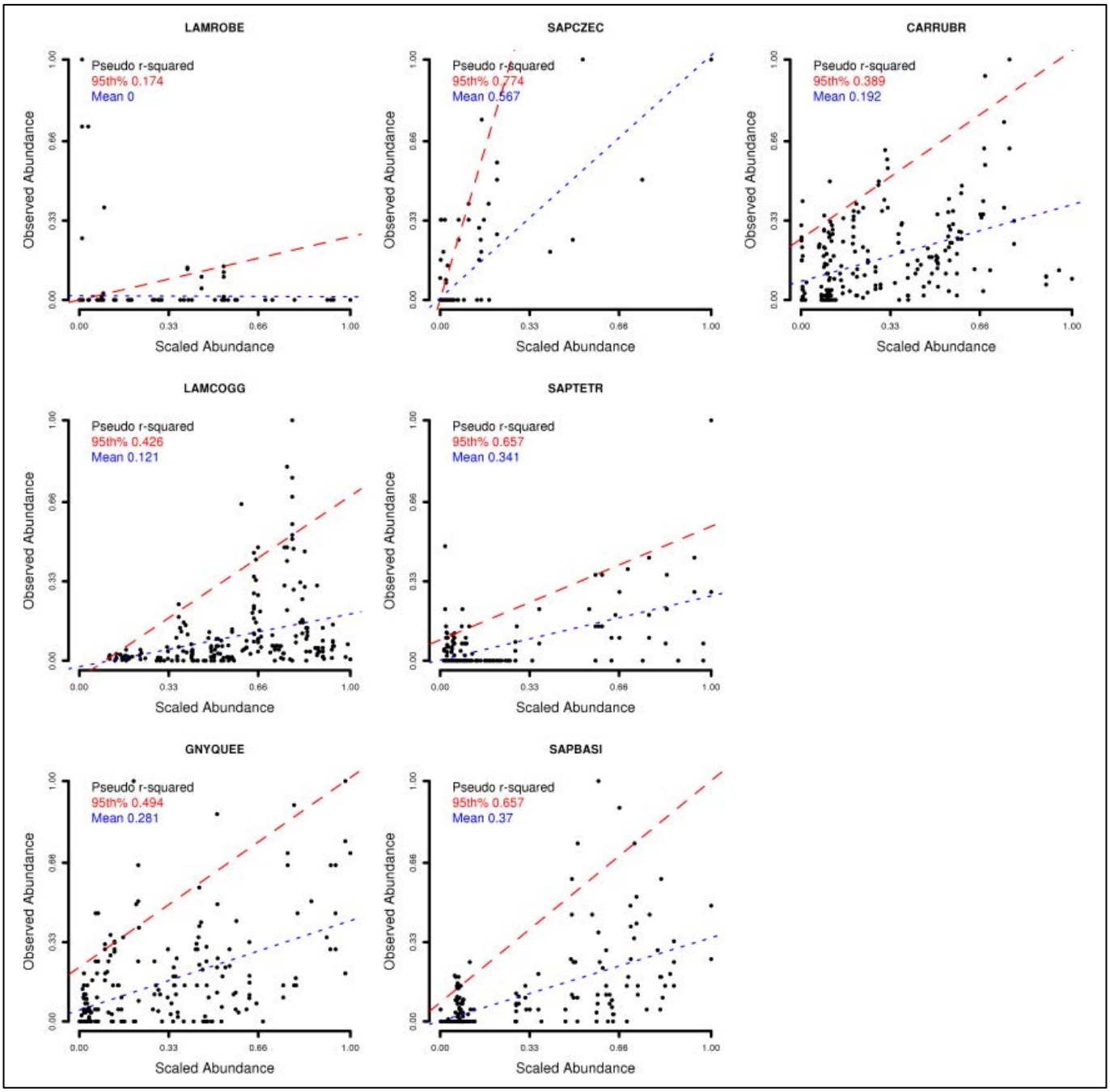



**Figure S5.6:** Fit of occupancy model output (x-axis) to scaled count data (y-axis). The dashed red line represents the fit of the model output to the 95$^{th}$ percentile of count data. The dashed blue lines represent the fit of the model output to the mean of the count data.

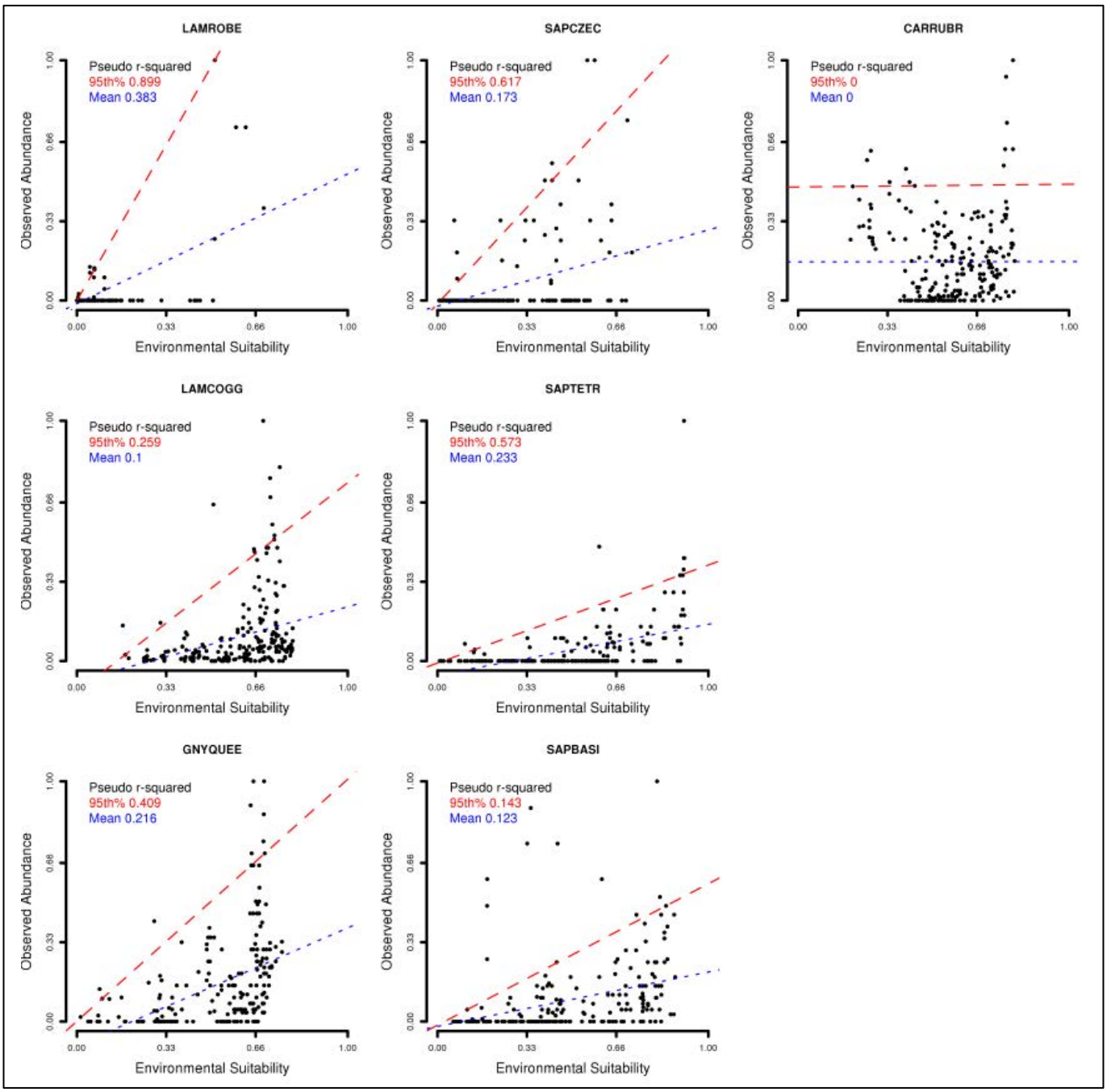



**Figure S5.7:** Fit of saturated MaxEnt model output (x-axis) to scaled count data (y-axis). The dashed red line represents the fit of the model output to the 95$^{th}$ percentile of count data. The dashed blue lines represent the fit of the model output to the mean of the count data.

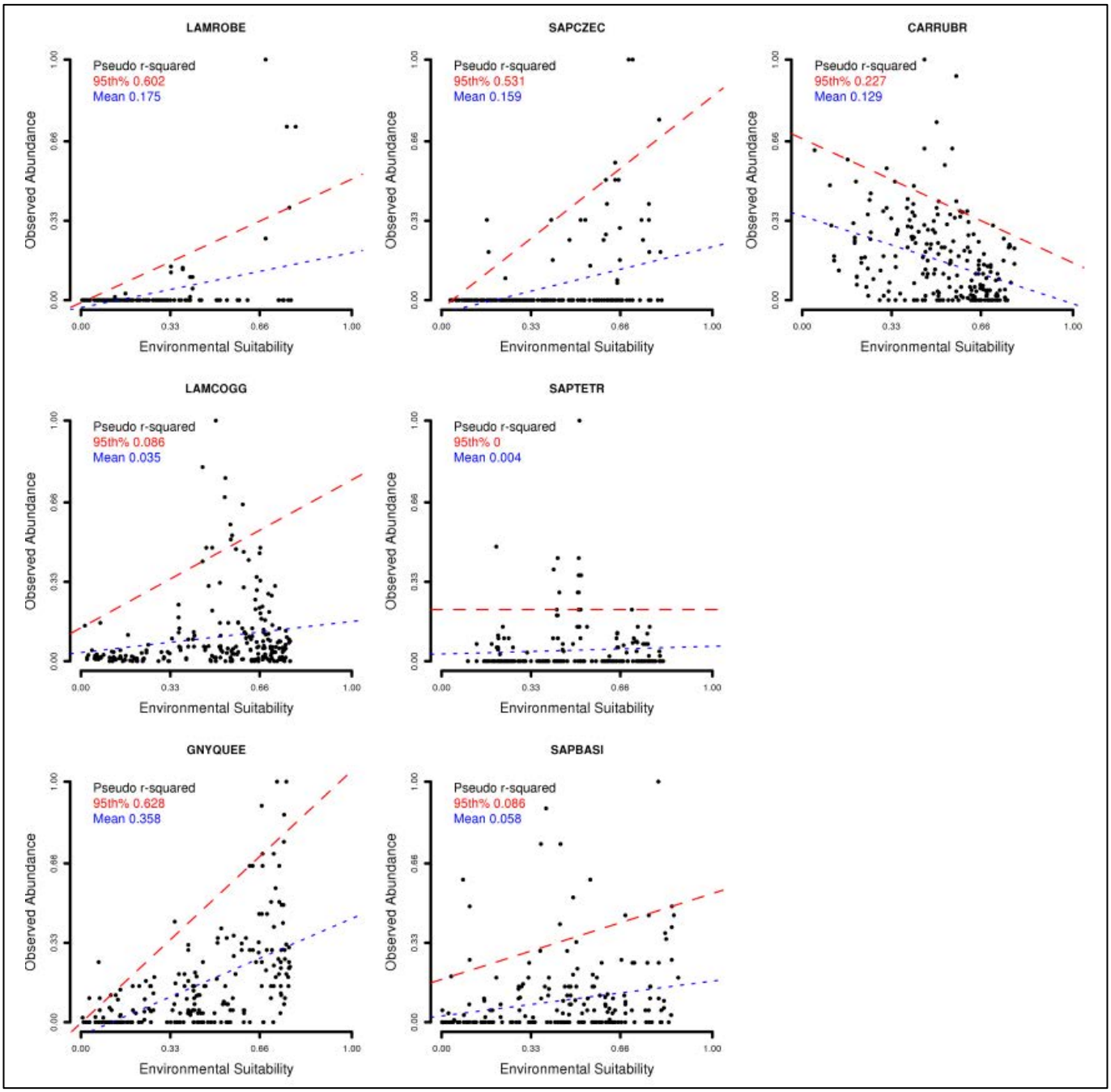



**Figure S5.8:** Fit of occupancy equivalent MaxEnt model output (x-axis) to scaled count data (y-axis). The dashed red line represents the fit of the model output to the 95$^{th}$ percentile of count data. The dashed blue lines represent the fit of the model output to the mean of the count data.

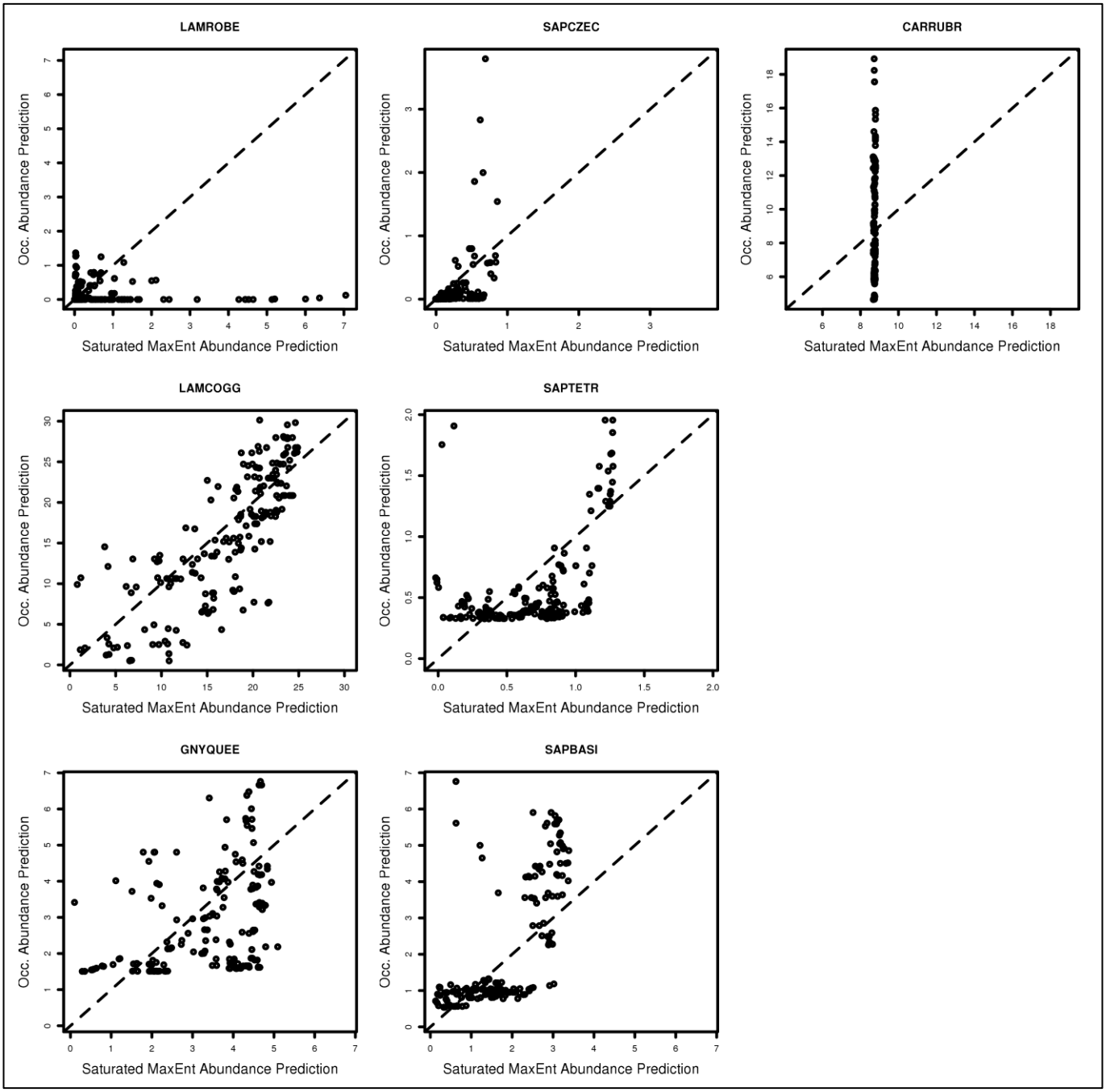



**Figure S5.9:** Comparison of abundance as predicted by Saturated MaxEnt model (x-axis) and occupancy model (y-axis).

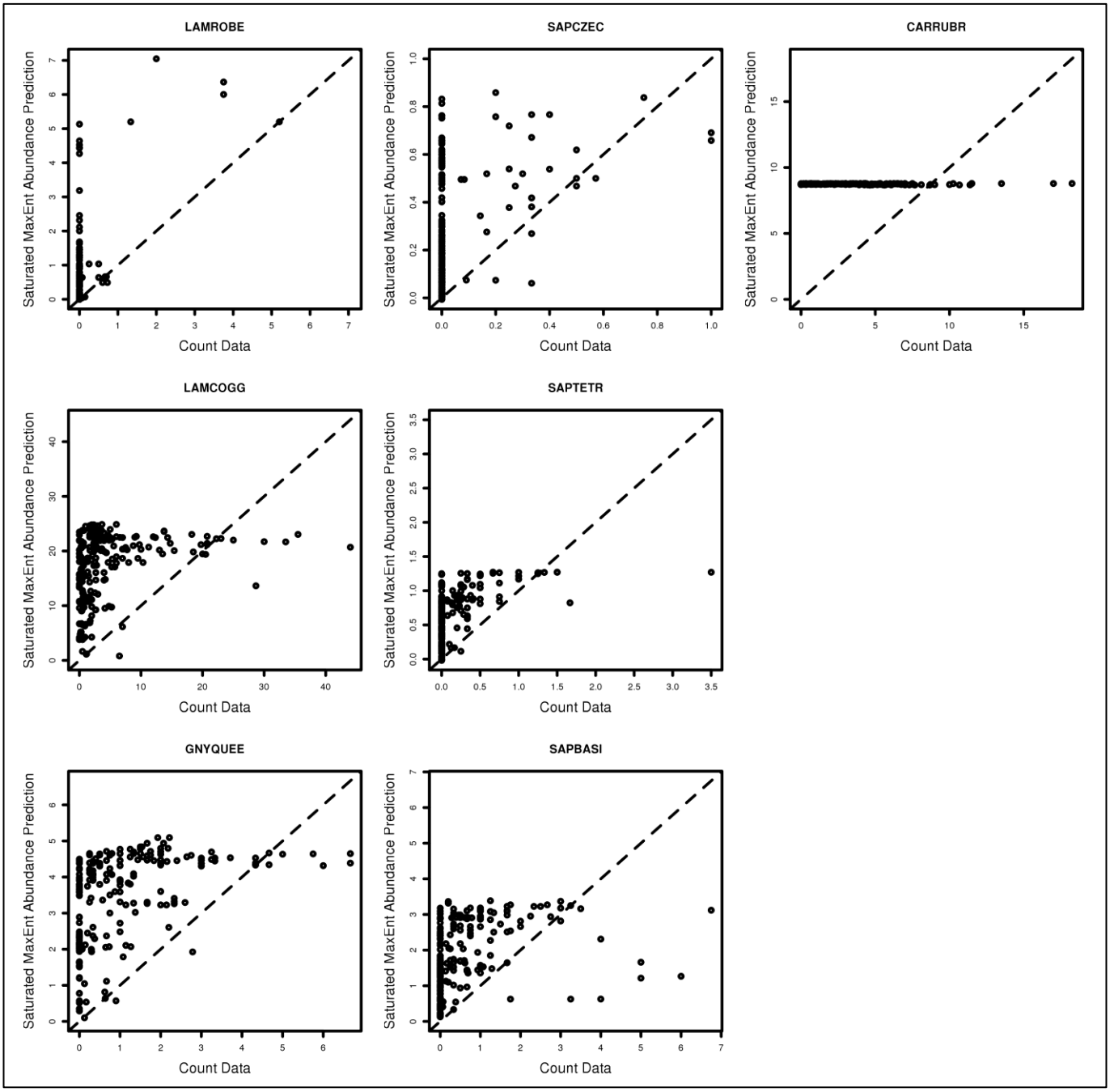



**Figure S5.10:** Comparison of measured carrying capacity (x-axis) to predicted abundance from the Saturated MaxEnt model (y-axis).

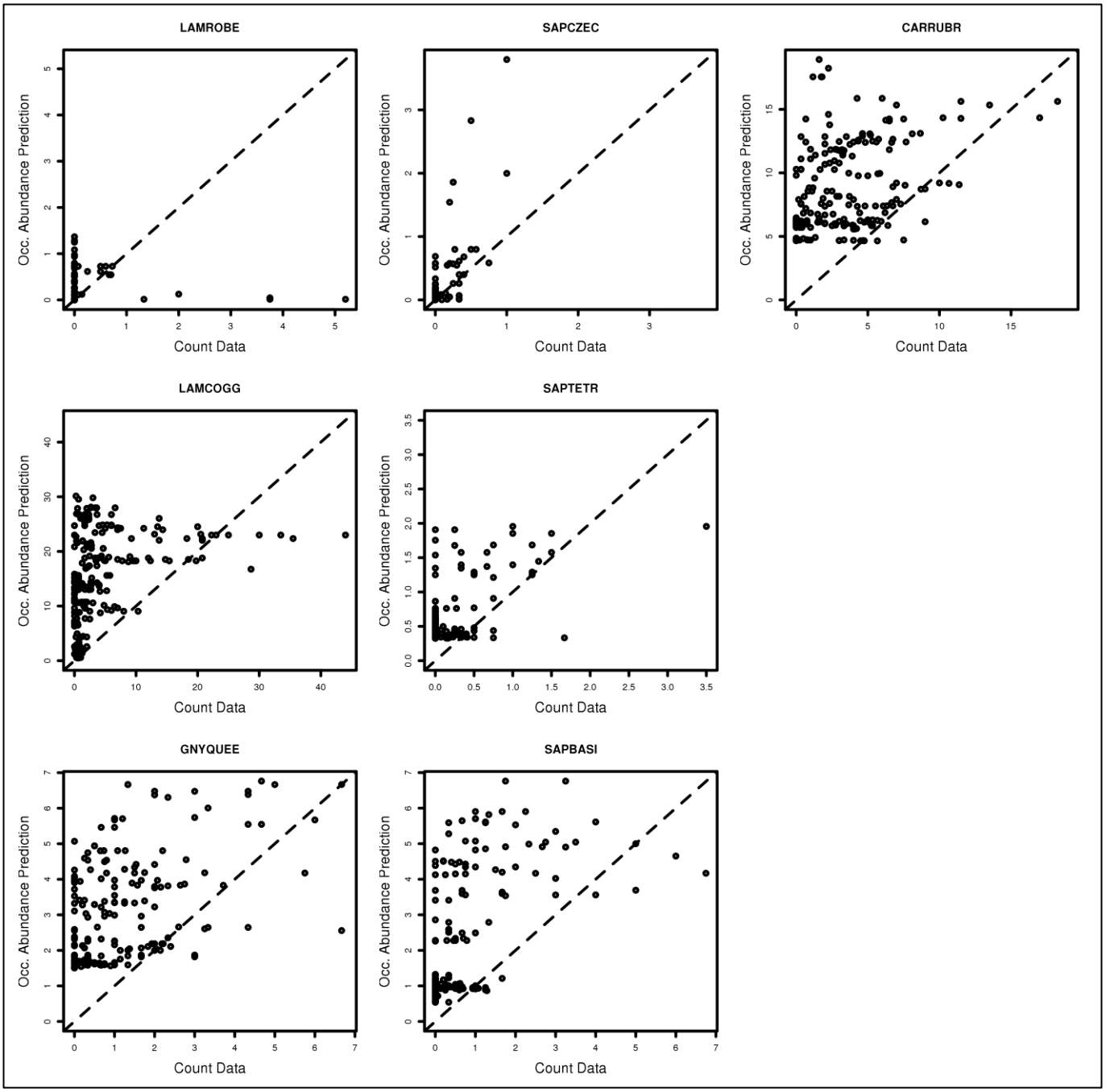



**Figure S5.11:** Comparison of measured carrying capacity (x-axis) to predicted abundance from the occupancy model (y-axis).

Models based on downscaled and current-best practice climate layers are both producing outputs with adequate fit to the data. MaxEnt AUC values are all good, while most occupancy models pseudo $r^2$ values are above 0.7, indicating these models fit the training data well. However, when model outputs are related to the count data, differences between them become apparent. One explanation could be that fitness is not directly related to temperature experienced in this particular group. Temperature may indirectly mediate fitness via a link with resource availability or competitive ability of congeners (Austin 2002).

# Appendix S6: Key Abbreviations and Definitions

**accuCLIM surfaces**..... A set of 30 year average climate surfaces describing temperature and precipitation derived by statistically downscaling and then aggregating broad-scale weather surfaces (Storlie *et al*. 2013, Chapter 2)

**ANUCLIM surfaces**..... Australian National University Climatological Model, a set of 30 year average climate surfaces describing temperature and precipitation regimes at 250m resolution (McMahon *et al*. 1995, Xu & Hutchinson 2011).

**AWAP surfaces** ..... Australian Water Availability Project, a set of weather surfaces describing daily maximum temperatures, daily minimum temperatures, and daily rainfall at 5km resolution (Jones *et al*. 2009).

**Climate**..... Temporally aggregated description of temperature or precipitation regimes, usually a 30 year average.

**cSDM**..... Correlative species distribution model, a statistical technique to predict species environmental relationships and distributional characteristics based on spatial layers of the environment and species occurrence data.

**mSDM** ..... Mechanistic species distribution model, a biophysical technique to estimate a species fundamental niche and distributional characters based on spatial layers of the environment and species specific physiological rates.

**Weather**..... Temporally discrete description of temperature or precipitation regimes (e.g. daily maximum temperature).