

**The geographic distribution of *Senecio
glastifolius* in New Zealand: past,
current and climatic potential**

By

Josef Rehua Beautrais

A thesis

**submitted to Victoria University of Wellington
in partial fulfilment of the requirements for the degree of
Master of Science in Conservation Biology**

Victoria University of Wellington

2013

Abstract

Senecio glastifolius (Asteraceae) is an invasive species in New Zealand, where it threatens rare and vulnerable coastal floristic communities. It has expanded its range dramatically over recent years and continues to spread. It is subject to control programs in parts of its distribution. Uncertainty over its future distribution and invasive impacts in New Zealand contribute to the difficulty of its management. To address this knowledge gap, the potential distribution of *S. glastifolius* in New Zealand was predicted, based on its bioclimatic niche.

Existing information on its current distribution and historic spread is incomplete, stored in disparate sources, and is often imprecise or inaccurate. In this study, available information on its distribution and spread was synthesised, processed, and augmented with new data collected in the field by the author. This data set was optimised for use in species distribution modelling.

The distribution of *S. glastifolius* is described in its native range of South Africa, plus invaded regions in Australia, the British Isles and New Zealand. The data set describing its distribution is of higher quality than any known previous data set, is more extensive, and more suitable for use in species distribution modelling. The historic spread of *S. glastifolius* in New Zealand is presented, illustrating its expansion from sites of introduction in Wellington, Gisborne, plus several subsequent sites, to its now considerable range throughout much of central New Zealand.

A predictive model of the potential distribution of *S. glastifolius* was created based on the three main climatic variables observed to limit its distribution: mean annual temperature range, aridity, and minimum temperature of the coldest month. MaxEnt models were trained on data from all regions for which georeferenced records of the species were available; South Africa, Australia, New Zealand and the Isles of Scilly. Predictions were evaluated using methods appropriate to the special case of range-expanding species. Models performed well during validation, suggesting good predictive ability when applied to new areas.

Analysis of the realised niche space of *S. glastifolius* in the two climatic dimensions most influencing its distribution: Annual Temperature Range and Aridity, indicated that it is exploiting almost totally disjunct niche spaces in New Zealand and South

Africa. Of the climate space occupied in New Zealand, almost none is available to the species in its native range of South Africa.

Predictions of *S. glastifolius*'s potential distribution in New Zealand reveal significant areas of suitable habitat yet to be invaded. Much of this suitable habitat is contiguous with the current range and active dispersal front of *S. glastifolius*, suggesting that invasion is highly likely under a scenario of no management intervention. Specifically, it is suggested that control and surveillance in coastal Taranaki are required to prevent invasion of an area covering most of the northern third of the North Island.

Acknowledgements

This research was made possible by generous assistance and support from a large number of people, to whom I am sincerely grateful. In particular I would like to thank: My supervisors, Stephen Hartley and Mairead de Roiste, for their academic and moral support.

My parents Keith and Margie for their generosity in providing moral and financial support. Keith provided constructive comment on drafts.

My sisters Airini and Emily, who gave great encouragement. Special thanks, Emily, for bringing me food when I was working consecutive 14 hour days in the lab, and Airini for helping me when I was ill.

Constructive input was offered by Jane Elith and colleagues at University of Melbourne, and by Victoria University of Wellington academic groups 'Bug Club' and 'Space'. Generous technical assistance was offered by VUW's Baron Andrew Rae.

A big thank you to everyone who generously offered their knowledge on *Senecio glastifolius*; Bob Brockie, Colin Ogle, Mike Dodd, Jan Vlok, Rosemary Parslow, Diane Evers, David McNamara, Nick Singers, Jill Rapson, Don Ravine, Mick Parsons, Peter Williams, Phil Karaitiana, Kim Wright, Dave Edge, Graeme LaCock, Craig Davey, Melanie Newfield, Peter Daniel and John Scott. Thanks also to the staff at the herbaria listed in the main text. Further thanks to all of the herbarium staff and botanists whom I pestered, but who were unable to offer information.

Financial support was given by Victoria University's Centre for Biodiversity and Restoration Ecology (CBRE), and the Wellington Botanical Society.

Thanks, also, to all of my friends, flatmates, confidants, and partners-in-crime.

To anyone who I have inadvertently omitted from this list, I apologise.

Contents

Abstract	2
Acknowledgements	4
List of Figures	10
List of Tables	12
Chapter 1: Predicting the potential distributions of invasive species	13
1.1 Research context	13
1.2 Invasive species	13
1.3 Modelling invasive species' distributions	14
1.4 What is Species Distribution Modelling?	15
1.4.1 Uses of SDM	15
1.4.2 Types of model: mechanistic versus correlative	16
1.5 Ecological theory	17
1.5.1 Species range limits determined by environment	17
1.5.2 Space duality: geographic versus environmental space	18
1.5.3 Types of niche	18
1.5.3 Niche conservatism	19
1.6 Data used for SDM	19
1.6.1 Species presence data	19
1.6.2 Environmental data	20
1.6.3 Choice of predictor: direct predictors versus indirect	21
1.6.4 Limitation of resolution and scale of functional influence	22
1.6.5 Absence, Background, and Pseudo-absence data	22
1.6.6 Target groups absences and background bias matching	24
1.6.7 Screening pseudo-absence data	24
1.6.8 Delineation of the background region	25

1.6.9 Choice of geographic range for model training.....	26
1.7 Models used	27
1.7.1 Presence only	28
1.7.2 Presence – Absence/pseudo-absence/background	28
1.7.3 MaxEnt.....	29
1.7.4 Comment on uses – complex functions for geographic interpolation; simple functions for geographic extrapolation	29
1.8 Output.....	30
1.8.1 What is being mapped?	30
1.8.2 Thresholds for transformation	31
1.8.3 Minimum training presence	31
1.8.4 Response curves	32
1.9 Model Evaluation	32
1.9.1 AUC	33
1.9.2 Significance tests	34
1.9.3 Which evaluation measures are best for SDM of invasive species?	34
1.9.4 Obtaining test data	36
1.10 Research aims.....	36
1.11 Thesis structure	37
1.12 References	37
Chapter 2: Describing the current distribution and historical spread of <i>Senecio glastifolius</i>	47
Abstract	47
2.1 Introduction	47
2.1 Methods.....	48
2.1.1 Species occurrence data.....	48
2.1.2 Data cleaning	49
2.1.3 Data generalisation	51

2.3 Results.....	51
2.3.1 Native Range: South Africa	51
2.3.2 Australia	54
2.3.3 British Isles.....	56
2.3.4 Madeira	58
2.3.5 New Zealand	59
2.3.6 New Zealand invasion timeline:.....	61
2.4 Discussion and conclusions.....	63
2.5 References	64
Chapter 3: Expanding menace or invasional has-been? The potential distribution of <i>Senecio glastifolius</i> in New Zealand.....	67
Abstract	67
3. 1. <i>Senecio glastifolius</i> in New Zealand: history, ecology and threats.....	67
3.1.1 Ecological requirements	69
3.1.2 Legal Status	70
3.1.3 Control effort	70
3.1.4 Benefit of modelling.....	70
3.2 Aims.....	71
3.3 Methods.....	71
3.3.1 Species presence data.....	71
3.3.2 Environmental data	72
3.3.3 Transformation of Environmental Data.....	72
3.3.4 Choice of predictor variable and removal of correlated predictors	73
3.3.5 Background data	75
3.3.6 Modelling methods	77
3.3.7 Transformation of suitability score.....	79
3.3.8 Model evaluation	79

3.3.9 Madeira	80
3.4 Results.....	80
3.4.1 New Zealand	80
3.4.2 Australia	82
3.4.3 Scilly.....	83
3.4.4 South Africa.....	83
3.4.5 Species response to environmental variables	84
3.4.6 Results of Cross-validation.....	86
3.4.7 Ordinary cross-validation of globally fitted model	87
3.5 Discussion	87
3.5.1 Caveats	88
3.5.2 Sample bias	88
3.5.3 A note on minimum training presence	88
3.5.4 Background delineation.....	90
3.5.5 A comment on the extrapolation involved in predictions	90
3.6 Conclusion	91
3.7 References	91
Chapter 4: Discussion and conclusions.....	95
4.1 Summary of key findings	95
4.2 Challenges to modelling	96
4.3 Future research	96
4.3.1 Designed surveys of <i>S. glastifolius</i>	97
4.3.2 Better utilising information from imprecise species location data	97
4.4 Concluding remarks	97
4.5 References	97
Appendix 2.1 Recorded historical spread of <i>S. glastifolius</i> in New Zealand.	99
Appendix 2.2 Herbaria cited	111

Appendix 3.1 Comparison of MaxEnt models using raw and transformed data for MinTemp	112
Methods.....	112
Results.....	112
Appendix 3.2: Climatic suitability predictions for <i>Senecio glastifolius</i> using BIOCLIM.	116
Comparison to MaxEnt	116
References	119
Appendix 3.3: Correlation structure of predictor variables.....	120
References	121
Appendix 3.4: Predicted Suitability of Madeira Archipelago for <i>Senecio glastifolius</i>	122
Appendix 3.5: Invasion to New Zealand by <i>Senecio glastifolius</i> reveals niche space not available in native range of South Africa	124

List of Figures

Chapter	Figure	Description	Page
Two	2.1	Location of Cape Region	34
	2.2	Native distribution of <i>S. glastifolius</i>	35
	2.3	Distribution of <i>S. glastifolius</i> in Western Australia	
	2.4	Distribution of <i>S. glastifolius</i> in New South Wales, Australia	
	2.5	Location of the Isles of Scilly and Guernsey	
	2.6	Distribution of <i>S. glastifolius</i> in Scilly	
	2.7	Location of Madeira	
	2.8	Distribution of <i>S. glastifolius</i> in New Zealand	
Three	3.1	Winter temperature in the Cape Region	78
	3.2	Background region for New Zealand	
	3.3	Predicted climatic suitability for <i>S. glastifolius</i> in New Zealand	
	3.4	Predicted climatic suitability for <i>S. glastifolius</i> in Australia	
	3.5	Predicted climatic suitability for <i>S. glastifolius</i> in Scilly	
	3.6	Predicted climatic suitability for <i>S. glastifolius</i> in South Africa	
	3.7	Response curves	
Appendix 2.1	A2.1	Historical spread of <i>S. glastifolius</i> in New Zealand	
Appendix 3.1	A3.1.1	MaxEnt predictions using transformed data	
	A3.1.2	MaxEnt predictions using raw data	
	A3.1.3	Difference between predictions using transformed and untransformed data	
Appendix 3.2	A3.2.1	BIOCLIM prediction for New Zealand, Australia, South	

Africa and Scilly

Appendix 3.3 A3.3.1 Correlation structure of key environmental variables

Appendix 3.4 A3.4.1 Predictions for *S. glastifolius* in Madeira

A3.4.2 Novel climates in Madeira

Appendix 3.5 A3.5.1 Apparent niche shift of *S. glastifolius* between native range and New Zealand

A3.5.2 Biotope overlap between South Africa and New Zealand

List of Tables

Chapter	Table	Description	Page
Two	2.1	Distribution record processing	34
Three	3.1	Results of regional cross-validation	78
	3.2	Results of global cross-validation	
Appendix 3.1	A3.1.1	Key differences between MaxEnt models	
	A3.1.2	Predictor variable contribution	
Appendix 3.2	A3.2.1	Results summary for BIOCLIM models	
Appendix 3.3	A3.3.1	Correlation structure of predictor variables	

Chapter 1: Predicting the potential distributions of invasive species

1.1 Research context

This thesis aims to predict the potential distribution of the invasive species *Senecio glastifolius* (Asteraceae) in New Zealand, and to describe its current range and historical spread. The primary motivation is to provide information useful to the management of this species. *Senecio glastifolius* currently threatens coastal ecosystems in New Zealand and is considered a nuisance in some agricultural contexts. It has expanded its range significantly in recent years and continues to spread. There is uncertainty as to its future distribution and impacts. This thesis describes the current distribution of *S. glastifolius* in its native and invaded ranges, within a particular focus on New Zealand. The potential distribution of *S. glastifolius* in New Zealand is predicted, using information on its current distribution to infer the climatic requirements that limit its distribution.

1. 2 Invasive species

Invasive species are widely recognised as a significant threat to global biodiversity (Mack et al., 2000) and as a significant part of global biotic change (Vitousek et al., 1997). There is a vast and rapidly expanding literature devoted to their study, not only because of the threat they pose to valued ecosystems, economic, cultural and aesthetic values, but also the unique opportunity to expand ecological knowledge that species living outside of their native range present (Richardson et al., 2011). Species invasions can provide unique insights into the ecological constraints that determine species distributions, an understanding long sought by ecologists (Andrewartha & Birch 1954, cited in Kearney & Porter 2009; Fitzpatrick et al., 2007). An area that contributes to both of these questions is that of predictive modelling of invasive species' potential distributions.

Human agency has massively increased the rate at which species are introduced to new environments (Lodge et al., 2003). Although most species introductions do not result in naturalisation, and most naturalised species are not considered problematic, a small proportion becomes 'invasive' (Mack et al., 2000). The concept of 'invasive' is a subjective construct, and definitions lack consistency (Richardson et al., 2000), but here I use the term to describe any human-introduced species which causes undesirable effects in its new environment.

Invasive species cause major harm to valued ecosystems via a range of processes. These include displacement of desirable species through competition; modification of nutrient cycles, fire regimes, hydrology or habitat structure (Mack et al., 2000). Invasive species, collectively, create massive economic costs. Invasive species are estimated to cause nearly US\$120 billion per year of economic damage in the United States alone (Pimentel et al., 2005). Alongside the degradation of existing ecosystems, invasive species pose major limits to efforts of ecological restoration (Norton, 2009). Invasive species can also have negative cultural impacts, including displacement of culturally important native species (Pfeiffer & Voeks, 2008). Invasive species also have potential to cause aesthetic impacts, such as the modification of the structure or colour of landscapes.

Species introduced to new environments, including invasive species, can be thought of as constituting a "global experiment in biogeography", and offer opportunities to investigate a number of evolutionary and ecological questions (Richardson et al., 2011). The deliberate transplanting of species into natural environments is seldom considered ethical.

1.3 Modelling invasive species' distributions

The ability to anticipate the extent to which a species will naturalise in a given region is of value to environmental managers, as well as to industries vulnerable to invasive species. Reliable knowledge of the suitability of a region to pests can inform surveillance efforts and management strategies, with potential both to save money wasted on putative pests which are unlikely to naturalise in a region, and to identify

invaders whose threat is in fact greater than might be assumed from intuitive approaches or non-systematic data. Predictive models can be useful in identifying areas where invasive species are likely to first establish (Broennimann et al., 2007), facilitating rapid detection of founder populations which can be critical to their successful eradication or control (Williams et al., 2008 and references therein). Predictive distribution models, when projected onto regions where a species is not yet found, can complement non-spatial risk assessments (e.g. Pheloung et al., 1999) by offering a spatial assessment of invasion risk.

This chapter provides an introduction to the concepts underlying predictive distribution modelling, offers a brief overview of the mechanics involved, and explores some of the challenges encountered in species distribution modelling. An opportunity for research is introduced.

1.4 What is Species Distribution Modelling?

Species distribution models (SDMs: Franklin, 2009; Elith & Leathwick, 2009) predict the geographic distribution of species, based on measured or inferred aspects of the species' ecology. The key premise upon which SDMs rest is that species possess stable ecological niches, describable in terms of the environmental factors that limit their fitness and therefore their existence across associated geographic space.

1.4.1 Uses of SDM

Species distribution modelling is a well-entrenched and steadily expanding field of research (Franklin, 2009). Species distribution models are used in a wide variety of ecological and evolutionary applications. These applications can be loosely divided into four main categories: estimating current (realised) distributions, predicting potential distributions, predicting changes to distributions under environmental change, and inference of species' physiological tolerances (non-spatial).

Because information on species' current distributions is often limited, SDMs are commonly employed as a surrogate. Examples of when this is useful include informing decisions on conservation prioritisation (e.g. Mota-Vargas & Rojas-Soto, 2012), or planning field surveys of rare species (e.g. Guisan et al., 2006).

SDMs are also useful in estimating species' potential distributions – hypothetical distributions possible under current environmental conditions, but with the removal of dispersal constraints and biotic effects. This can be useful for reserve design (Wilson et al., 2005), planning species reintroductions (Hirzel et al., 2004), predicting species invasions (Peterson & Vieglais, 2001; Peterson, 2003), studying disequilibrium with environment (Leathwick, 1998), assessing feasibility of biological control agents before introduction to a region (Mukherjee et al., 2012), or narrowing search areas for collecting suitable bio-control agents (Mukherjee et al., 2011), and mapping potential distributions of disease vectors (Kitron 1998).

Another common SDM application is predicting how species' ranges will respond to environmental change, especially climate change (Pearson & Dawson, 2003, Iverson et al., 2008, Yates et al., 2010). Such predictions are made for invasive species (Kearney et al., 2008), native species (Yates et al., 2010), and economically important species such as crops (e.g. Watt et al., 2011). The knowledge gained from these SDM studies can assist adaptation to climate change, as well as strengthen arguments for minimising environmental change.

Alongside spatial predictions, SDMs can sometimes offer information on the physiological tolerances of species (Mac Nally, 2000). Inference can often be made as to which environmental variables most limit species' distributions. The nature of the relationship between a species and a certain environmental variable can also sometimes be inferred.

1.4.2 Types of model: mechanistic versus correlative

Species distribution models are commonly divided into two classes, mechanistic or correlative, based on the source of information they use to describe a species' niche.

Mechanistic models (Kearney & Porter, 2004; Kearney & Porter, 2009) use known ecological responses to predict distributions; they do not require species location data as input (Kearney et al., 2008). Input includes known physiological tolerances, often generated from controlled experiments, or other documented information such as habitat use. Correlative models rely on the observed distribution of species to elucidate the species' niche. Correlative models are more widely used than mechanistic models, because the requisite knowledge of species' ecological requirements is often not known *a priori*. Indeed, investigating the environmental tolerances of species is often a primary objective of correlative modelling (section 1.4.1). While mechanistic and correlative approaches are not always mutually exclusive (Dormann et al., 2012), the conceptual distinction between mechanistic and correlative prediction is still useful. This study focuses on correlative SDMs. Subsequent discussion of SDMs refers to the correlative approach, unless otherwise stated.

1.5 Ecological theory

Species distribution modelling rests upon an ecological model (*sensu* Austin, 2002). In this section I examine some of the concepts and assumptions underlying SDMs.

1.5.1 Species range limits determined by environment

The field of SDM rests on the tenet that species' distributions are determined by their environments (Gaston, 2009). By describing relationship between species and environment, especially the limits to distribution imposed by environment, the distribution of species may be predicted for areas where the environmental conditions are known. The theory that environment, especially climate, governs species distributions is very well supported. Evidence is found in the fossil record (Woodward, 1987; Huntley, 1999; Davis & Shaw, 2001, cited in Pearson & Dawson, 2003).

1.5.2 Space duality: geographic versus environmental space

A fundamental concept in SDM is the duality between geographic and environmental space (Hutchinson 1957). At any given point in geographic space (as given by a Cartesian coordinate), one can measure a set of environmental variables. The vector of environmental variables associated with the geographic point can be mapped in environmental space. Environmental space is an abstract entity, and can incorporate any number of environmental axes.

1.5.3 Types of niche

Ecological niche theory used in SDM draws on the work of Hutchinson (1957) and subsequent advances (e.g. Soberon & Nakamura, 2009). Species can be thought to possess a fundamental niche (F), defined as the n -dimensional hypervolume in environmental space within which the species can survive and reproduce (Hutchinson, 1957). The actual geographic space occupied by a species, or realised niche (R), is a subset of its fundamental niche. Parts of the fundamental niche will not be occupied by the species, due chiefly to biotic interaction and geographic availability of environmental space. This concept of environmental space availability is captured in the concept of the biotope: the full set of environmental conditions which exist in the geographic region in which the species has its range. The realised niche is further constrained by dispersal limitations.

As outlined above, estimations of a species' niche are drawn from the species' recorded distributions. Hence, accurate characterisation of a species' niche is complicated by the fact that species are often to be found outside of their fundamental niches. This is true for both vagile species, which might pass through unsuitable habitat (Manning et al., 2005), or for sessile species which can exist as sink populations (Hanski, 1999). This invites an addition to the set theory outlined above. The fundamental periniche (Fp) is the space outside the fundamental niche that permits existence but not self-perpetuation. This is the home of sink populations.

1.5.3 Niche conservatism

Because SDM predictions of species' distributions are based on the assumption that distribution is determined by niche, they also rely on the assumption that species' niches are stable within the spatial and temporal scales used (Martinez-Meyer et al., 2004).

There is significant evidence for realised niche shifts between native and invasive populations (Gallagher et al., 2010), and various hypotheses have been put forward to explain this including biotic release (Guisan & Thuiller, 2005), phenotypic plasticity (Sexton et al., 2002), differences in climatic space availability between regions (Fitzpatrick et al., 2007), rapid evolution (Buswell et al., 2011), and climate space anisotropy (Soberon & Peterson, 2011).

In a study examining 37 animal species, Peterson et al., (1999) showed high levels of niche conservation during introduction to new regions. Other studies suggest rapid evolution of invasive species is possible, with consequent extension or alteration of niche in the new environment (Maron et al., 2004). Niche conservatism is a reasonable assumption in most applications, but differences in realised niches can be expected between regions.

1.6 Data used for SDM

Correlative SDMs require georeferenced species data, plus mapped environmental data to bridge geographic and environmental space.

1.6.1 Species presence data

Species' location data for SDM are drawn from a wide range of sources. Usually, data are collected opportunistically from existing sources such as herbarium and museum databases, species atlases, published location data, etc. Designed field

surveys are carried out only occasionally, mostly because of constraints of funding or logistics.

Opportunistically collected presence data used for SDM comes with limitations, both inherent and source-dependent. This data source most often lacks explicit information on species absence. The limitations of presence-only data are widely acknowledged in the literature (Li et al., 2011) and much research has been devoted to overcoming them or working within the constraints they pose.

An inherent limitation is that prevalence is unknowable (Elith et al., 2011).

Prevalence is the proportion of sites within a region, at a defined scale, that are occupied by a species. Species presence data also tend to contain bias resulting from the methods of their collection (Graham et al., 2004). Sampling intensity is likely to be relatively higher in areas more frequently visited by people, such as near roads (see, for example, Kadmon et al., 2004).

1.6.2 Environmental data

Climatic variables (e.g. mean annual rainfall) are the most commonly used environmental predictors of plant distributions (Franklin, 2009). Other variables including soil properties (Coudun et al., 2006), and land cover (Pearson et al., 2004) are important determinants of plant distributions but are usually not included in plant SDM exercises.

Environmental data is typically matched to species location data by extracting values from spatially continuous data layers in a GIS (geographic information system). Data may be interpolated from weather station records (e.g. Hijmans et al., 2005) or derived from remote sensing (e.g. Andrew & Ustin, 2009).

While it is vastly more common to draw environmental information from interpolated climate surfaces, some studies have instead used point data from the climate station closest to training data locations (e.g. Mgidi et al., 2007).

The suite of variables included in a model is necessarily a finite subset of the total suite of environmental variables experienced by a species, which may be measured on an infinite number of axes.

1.6.3 Choice of predictor: direct predictors versus indirect

Predictor variables used for SDM do not always have a direct, functional ecological influence on the focal species (Austin 2002). Instead, variables used are often ones which correlate with true functional drivers of species range. An example is altitude (indirect) in place of a measure of temperature (direct). Indirect variables vary greatly in how closely they relate to direct variables. For example, latitude (indirect) is more indirectly related to aspects of temperature (direct) than to day length seasonality (direct).

While preference should be had for directly relevant variables (Mac Nally, 2000), the unavailability of mapped variables means that correlated variables are commonly used in their place. Unfortunately, the highest quality and most functionally relevant predictors are often only mapped for localised areas and do not always span the entire area examined by a study (Austin & Meyers, 1996). One example is the Land Environments of New Zealand (LENZ) data, which includes fifteen environmental variables of direct ecological relevance to vegetation, at extremely high resolution (25 m grid cells), but limited to the New Zealand extent.

Predictors with direct ecological influence on a species give the best transferability to new times or regions. Indirect variables serve to explain species distributions only to the extent that their correlations with direct variables remain stable. These correlations cannot be expected to always remain constant in space or time, making predictions to new environments unreliable (Elith et al., 2010).

Choosing direct variables relies on prior knowledge of a species' ecology. In many cases, no or insufficient prior knowledge exists as to the true limiting factors of a species' distribution (Kilroy et al., 2008).

Variable selection in the absence of comprehensive ecological knowledge of the focal species can be complicated. For example, an external variable can exert a causative correlation on two or more variables unrelated to each other (e.g. topographic variables can influence both temperature and precipitation), leading to complicating correlations among candidate predictors.

1.6.4 Limitation of resolution and scale of functional influence

Choice of predictor variable is also often limited by practical aspects of scale. Different variables function at different scales to influence species distributions. In cases where environmental data is measured at a significantly coarser scale than at which it exerts ecological influence, its signal will be diminished or lost to the noise imparted by its generalisation.

1.6.5 Absence, Background, and Pseudo-absence data

Species absence data can be highly informative in delimiting the species distribution and correlating this to a developing understanding of the species' niche. The usefulness of absence data for SDM is widely accepted, and has been proven experimentally (e.g. Wisz & Guisan, 2009).

In most cases, reliable information on species absence is unavailable (Lobo et al., 2010). For most species, rigorous surveys of geographic distribution do not exist. Even with good survey design, the reliability of observed absences is limited by a host of complicating factors which can lead to 'false absences'. Such contingent factors include past disturbance, dispersal limitation and biotic interaction (Lobo et al., 2010). Dispersal limitation is particularly relevant to invasive species undergoing range expansion, which have not been present in a region for long enough to have reached an equilibrium distribution. Hence, absences from within the geographic area of current expansion are difficult or impossible to equate with lack of suitability. This is less of a problem in areas of stabilised distribution, such as species' native ranges, or areas of long establishment. Major disequilibrium of distribution may,

however, exist even in species' native ranges (Leathwick, 1998). Even long-established invasive species can have expanding or in-filling distributions (e.g. Beans et al., 2012).

Absence data are useful because they provide information on the environmental conditions that are unsuitable to a species (Lobo et al., 2010). This cannot be inferred from presence data alone. When absence data is not available to be contrasted against presence data, comparison can instead be made between the distribution of presences and the space available to a species. Hence pseudo-absence or background data are used; samples drawn from the study region/s (usually randomly), which offer some suggestion of space not occupied by a species.

An important distinction exists between background and pseudo-absence data, although they are often treated as being the same thing. The distinction is that pseudo-absences serve to inform a model of *unsuitable* space – the space which has been sampled by the species and remains unoccupied. Background data, in contrast, inform the model of the space *available* to a species. Background data are assumed to include both unsuitable and suitable habitat. Background data are used by models such as MaxEnt and ENFA, while pseudo-absence data is conceptually suited to models such as GLMs (e.g. binary logistic regression) which deal with presence versus absence.

Whether to use pseudo-absence or background depends on the specific algorithm, and also on the task. Using non-random pseudo-absences is probably good for describing a species actual distribution while probably bad for describing potential distribution. The difference here is that MaxEnt's output variable (which relates to presence versus absence) is not the same as its input variable (which relates presence to available space). If you train a GLM with presence versus background, then the output will be presence versus background and not presence versus absence.

1.6.6 Target groups absences and background bias matching

The problem of spatially biased presence data can be reduced by selecting background data with approximately the same bias (Phillips et al., 2009).

Theoretically, if the bias present in the background data is the same as the bias in the presence data then the model will not be compromised by bias (Elith et al., 2011).

The exact nature of the bias in presence records is usually unknown, necessitating an approximation. This is simplest when presence data come from a single source. A proposed method (Phillips et al., 2009), is for background data to be comprised of presences for 'target group' species; those species likely to have been observed by the same methods as the study species. However, when multiple methods have been used to collect study species presences the bias is less easily replicated.

Absence can also be inferred where samples have been collected for other taxa, where the presence of the focal species can be expected to have been recorded were it present (e.g. Mateo et al., 2010).

1.6.7 Screening pseudo-absence data

Models which, conceptually, treat zeros as a mutually exclusive case to ones, can be subject to limited fit when trained on pseudo-absence data that contains many presences. This is likely to occur when pseudo-absences are a random sample of the study region (or 'background' region, section 1.6.5). Several studies have put forward possible solutions, including removing recorded presence from pseudo-absence samples, or biasing the ~random selection of pseudo-absences by weightings proportional to the environmental dissimilarity to presence records (Zaniewski et al., 2002). Biasing pseudo-absence samples to increase contrast with presences can involve circular logic, and has been discouraged (Stokland et al., 2011). Simply removing known presences from pseudo-absence samples, however, is not inappropriate.

1.6.8 Delineation of the background region

When absence, pseudo-absence or background data are to be used, the modeller must decide the geographic region/s from which they are drawn. This region, which I will refer to as the background region, has often been chosen arbitrarily. The importance of background region delineation has recently received increasing attention (VanDerWal et al., 2009; Barve et al., 2011).

When choosing background data, consideration must be given to what they represent ecologically. Ideally, they provide information on the environmental conditions *available* to a species. The distribution of presences within the available space then offers an estimate of which conditions are favourable to a species and which are not. This is conceptually similar to the idea of ecological niche factor analysis (Hirzel et al., 2002).

The idea that the geographic area from which samples are drawn should not be larger than the species' range is at least as old as Austin & Meyers (1996). They note that spurious absences, 'naughty noughts' in their terminology, distort observed response curves and increase model commissiveness.

Background size used in various studies has ranged in extent from the minimum convex polygon encompassing the presence points (Rodda et al., 2011), to global (Capinha et al., 2011). The area of environmental space represented in the background can be expected to increase with greater geographic area. Too small a background is undesirable for predicting potential distributions (Lobo et al., 2010; VanDerWal et al., 2009), because little information will be present about the accessible environmental space that the species does not inhabit. Too large an area will tend to lead to over-prediction (high commission) (Lobo et al., 2010; VanDerWal et al., 2009). Variable selection is also affected, with models utilising a very large background tending to be dominated by fewer variables than models trained using smaller backgrounds (VanDerWal et al., 2009).

The size of the background in relation to the presence data governs the ecological question that the resulting model answers (Elith et al., 2011). A highly local

background creates a model which explains local habitat selection, comparing occupied space to intervening or closely neighbouring unoccupied space. In contrast, using a background spanning many eco-climatic regions asks why a species is found in one eco-climatic zone and not others (Elith et al., 2011, *cf.* Webber et al., 2011). Clearly the size of background should be informed by the research question. Lobo et al., 2010 found that a greater contrast between ~absence and presence data is needed to predict potential distributions than realised distributions.

One consideration in choosing a region from which to draw absences highlighted by Elith et al., 2010 is the extent of extrapolation then involved in projecting to new regions; a larger training region will usually contain a wider variety of environmental space, thus reducing extrapolation. Thus it may seem useful to increase the background region, since extrapolation introduces uncertainty into predictions; however, more careful conceptual analysis is required. The reason that extrapolation is undesirable is that it involves fitting values to data values or combinations which the model has no information for (i.e. data which were not represented in the training data). However, including training data from regions where rigorous sampling has not occurred (including regions in which equilibrium assumptions are unreasonable) does not in fact offer any useful information to the model: it is unknowable whether values represent 'true' presences or absences. In short, there exists an unavoidable choice between a) using reliable data but extrapolating in order to make predictions, and b) reducing the apparent level of extrapolation at the expense of data reliability. In this study a preference has been given to using a reliable model and to measure and state the uncertainty (level of extrapolation), than to fit an unreliable model with spurious accuracy.

1.6.9 Choice of geographic range for model training

Invasive species, by definition, exist in more than one region. Usually these regions are non-contiguous, though this need not be the case (e.g. *Acacia spp.* in Australia (Webber et al., 2011) and *Corynocarpus laevigatus* in New Zealand (Costall et al., 2006)).

When species presence data exists for more than one region, the modeller must choose which region or combination of regions to use for model-training. The literature includes examples of using only the invaded region for which predictions are to be made (e.g. Elith et al., 2010); the native region only (e.g. Chen et al., 2007); or a combination of both (Fitzpatrick et al., 2007), sometimes including other invaded regions elsewhere in the world (Webber et al., 2011).

Training models using all available regions has important advantages. The different geographic regions occupied by a species will often have different climates, each giving a different window into the species niche (Soberon & Peterson 2011). This can be very pronounced. Different variables will limit species' distributions in different parts of its total range (Grinnell, 1917). In addition to these benefits, using the full set of occupied ranges yields a larger data set. Model performance tends to improve with larger samples, but the effect is asymptotic (Stockwell & Peterson, 2002).

One mechanism by which the inclusion of the full range of available data tends to improve prediction is the reduction of extrapolation required. The union of multiple geographic regions usually spans a greater volume of environmental hyperspace. Webber et al., (2011) found that when predicting invasive acacia species in South Africa, use of the full set of geographic ranges improved model sensitivity (defined as the proportion of presence data correctly classified by the model). Sensitivity is an especially important quality of invasive species SDMs.

1.7 Models used

Once species location data has been compiled, and a suitable suite of predictor variables selected, a model must be constructed to describe the relationship between the two. A vast number of algorithms have been used in SDM, and new approaches are continually emerging (e.g. Li et al., 2011). Examples include statistical methods: generalised linear models (GLM, Venables & Ripley, 1994), generalised additive models (GAM, Hastie & Tibshirani, 1990), multivariate adaptive regression splines (MARS, Moisen & Frescino, 2002), Bayesian models (Latimer et al., 2006), spatial autoregressive models (Dormann, 2007), and machine learning

methods: tree based methods such as classification and regression trees (CART, Breiman et al., 1984), random forests (RF, Prasad et al., 2006) and boosted regression trees (BRT, Elith et al., 2008, also known as stochastic gradient boosting or generalised boosted models), artificial neural networks (ANN, Olden et al., 2008), genetic algorithms (e.g. GARP, Stockwell & Peters 1999), maximum entropy models (e.g. MaxEnt, Phillips et al., 2006), support vector machines (SVM, Drake et al., 2006); envelope models: BIOCLIM (Busby, 1991); distance metrics: DOMAIN (Carpenter et al., 1993), Mahalanobis distance (Farber & Kadmon, 2003); and ecological niche factor analysis (ENFA, Hirzel et al., 2002). This list is far from exhaustive.

Choice of algorithm often reflects common usage in a particular sub-discipline (Elith & Leathwick, 2009), and the choice of algorithm may be less critical to the modelling process than the practitioner's familiarity with the chosen algorithm (Austin et al., 2006).

Algorithms can be categorised by several means, including conceptual and mechanical approach, interpretability of process (i.e. black box to clear box continuum), and data requirements.

1.7.1 Presence only

Certain algorithms examine presence data only, with no consideration of absence. These include rectilinear envelopes (e.g. BIOCLIM, Busby 1991); distance metrics such as the Gower metric or the Mahalanobis distance (Farber & Kadmon, 2003).

1.7.2 Presence – Absence/pseudo-absence/background

Approaches using pseudo-absence or background data in place of true absence have occasionally been referred to as presence-only models, but a conceptual distinction should be made between these and the purely presence-only models (section 1.7.1).

Presence only techniques make no use of the environmental characteristics of a species range except that contained in the presence data. Examining not only presence locations, but the intervening and surrounding geographic space allows some inference of habitat selection preferences.

Some algorithms are formulated specifically to examine the background from which a species is selecting, such as ENFA (Hirzel et al., 2002) or MaxEnt (Phillips et al., 2006).

1.7.3 MaxEnt

Maximum entropy modelling has seen increasing usage in SDM applications with the advent of the freely available Java program 'MaxEnt' (Phillips et al., 2006).

Complementing its increased usage, a number of studies have been published focussing on its strengths and limitations for predicting species distributions (Rodda et al., 2011; Elith et al., 2011; Baldwin 2009; Warren & Seifert 2011; Phillips & Dudik 2008; Dudik et al., 2007; Phillips et al., 2006).

1.7.4 Comment on uses – complex functions for geographic interpolation; simple functions for geographic extrapolation

Some models fit only very simple relationships to the data (e.g. rectilinear envelopes in environmental space), while at the other extreme methods such as boosted regression trees or MaxEnt are capable of fitting very complex relationships.

Choosing how closely the model fits the environmental data is important. Overly simplistic models tend to provide comparatively poor predictions of species' current distributions (Elith et al., 2006). Limitations of very simple modelled functions include the oversimplification of a species' response to a given environmental variable. For example, the Mahalanobis distance assumes a Gaussian response. Asymmetric responses to environmental variables are not uncommon (Austin, 1980), and even multimodal responses can be observed (Le Maitre et al., 2008). Judicious transformation of variables can allow asymmetric responses to fit more closely the shapes accommodated by the model (Franklin, 2009).

In a recent study Heikkinen et al. (2012), showed that interpolative accuracy does not always come at the expense of transferability to new regions, and that some algorithms, such as MaxEnt, demonstrate particularly good capability of both.

1.8 Output

1.8.1 What is being mapped?

The chief output of SDMs is mapped values, usually on a continuous scale from 0 to 1. What this value actually represents is often poorly defined (Kearney, 2006).

Mapped output is, broadly, the geographic realisation of the species' estimated relationship with environmental space. This has often been taken to be the probability of species presence (Guisan & Zimmermann, 2000), which follows naturally from the use of techniques such as generalised linear models, which were developed prior to SDM. If certain assumptions are met, then probability of presence is a fair interpretation of mapped output. These assumptions include input data from well-designed surveys containing both species presence and absence, the absence of bias, and equilibrium between species and environment.

MaxEnt's default 'logistic' output, scaled from 0 to 1, can be interpreted as a probability of presence, subject to assumptions of sampling design such as plot size (Elith et al., 2011). However, where these assumptions cannot be met, the true probability of presence is unknowable. However, mapped output still provides a useful measure of relative environmental suitability.

Predicted values should have a monotone relationship with environmental suitability. While the relationship is unlikely to be perfectly linear, higher predicted values can usually be expected to be more suitable than lower predicted values.

Monotonicity depends on the algorithm and the data. For example, the Mahalanobis distance treats the training mean as the species optimum for each gradient. When the training mean differs from the true niche optimum, the highest assigned values

will not correspond with the highest true suitability (Knick & Rotenberry, 1998). This is a likely scenario, given data limitations and niche availability discussed above.

From the above we can see that mapped output from SDMs offer, at least, the spatial pattern of relative environmental suitability. Rather than simply knowing relative suitability, it is often desirable to differentiate between suitable and unsuitable geographic space. The continuous output variable of most SDMs therefore requires transformation via a threshold.

1.8.2 Thresholds for transformation

A number of approaches for choosing a threshold exist (Liu et al., 2005). Thresholds are often chosen to optimise a certain measure of performance, such as true positive rate, true negative rate, or some combination thereof. The choice of threshold should be governed by the particular application, taking into account the relative costs of different types of error, as well as properties of the training and test data. For example, when predicting the potential distributions of invasive species, omission is likely to be more consequential than commission, and is also more able to be determined (see section 1.9.3).

1.8.3 Minimum training presence

Minimum training presence is the lowest value of predicted suitability when the model is applied to the same data for which it is trained. An interpretation of this is that all environmental space which receives a higher suitability score is closer to the species' niche optimum. Assuming that the occurrence record which received the minimum training presence is reliable (i.e. was recorded correctly), then all space with an equal or higher suitability score can be expected to be suitable to the existence of the species, or, at least, that the combination of environmental variables included in the model does not preclude the existence of the species at associated sites. Factors not included in the model might render some of the associated geographic space unsuitable (e.g. biotic competition). Because species are sometimes recorded in space outside of their niche (e.g. growing as a sink population), it cannot always be assumed that the part of environmental space

associated with the minimum training presence represents part of the species' fundamental niche (section 1.5.3 niche theory).

1.8.4 Response curves

In addition to mapped predictions, a common output of SDMs is a graphical representation of a species' 'response curve' to an environmental variable. The response curve shows how the outcome (e.g. habitat suitability or probability of presence), plotted on the y axis, responds to changing values of a predictor variable. They are also referred to as partial dependence plots (e.g. Zurrel et al., 2012). Such plots are useful for identifying implied ecological relationships, such as environmental optima. They may also be examined to reveal evidence of undesirable model behaviour. An example is overfitting, which can be evidenced by unrealistically complex response curves (Elith et al., 2010). Unexpected shapes can, however, be due to the effect of other elements of a species' environment, including biotic influence. For example, where a Gaussian response is expected to a variable 'a', apparent bimodality can be observed where a limiting variable 'b' coincides with favourable values of 'a' (Austin, 1980).

Many modelling methods have built-in functionality for producing response curves. For those that do not, methods exist to create them manually (Elith et al., 2005). Caution is required in their interpretation, though, because a species responds to all dimensions of its environment simultaneously. This is difficult or impossible to represent graphically, and typically only one variable is examined per response curve plot (but see Hartley et al., 2006; Zurrel et al., 2012).

1.9 Model Evaluation

Validation of SDM predictions is inherently conceptually difficult. The very absence of information that justifies the use of a predictive model precludes its absolute validation at the time of modelling. Therefore, tests of predictive model performance

are only ever estimates. There exists a great diversity of metrics and methods to evaluate and validate SDMs, each with strengths and weaknesses that lend them best to different applications.

Perhaps the simplest measures of model performance are those relating to the number of correctly predicted cases. These include the number of true positives, false positives, true negatives and false negatives. This information is often presented as a 2x2 table, referred to as a 'confusion matrix' (Fielding & Bell, 1997). This information is typically expressed as a proportion, e.g. the true positive rate, which measures the proportion of correctly predicted positives. The false negative rate and the false positive rate are commonly termed omission and commission, respectively. Related concepts are sensitivity and specificity. Sensitivity measures the proportion of presences correctly predicted by the model; specificity measures the proportion of absences correctly predicted. All of these measures require data to be classified as presence or absence, and thus require continuous model output to be transformed via a threshold. For a given model, choice of threshold determines the ratio of sensitivity and specificity. A lower threshold will, *ceteris paribus*, increase sensitivity and the expense of specificity, while a higher threshold will increase specificity at the expense of sensitivity.

1.9.1 AUC

One measure which has gained popularity is the AUC, or, the Area Under the [receiver operating characteristic] Curve. AUC measures discrimination. It reports the proportion of correctly predicted cases when examining whether a randomly selected presence scores higher than a randomly selected absence. Where absence data are unavailable, background data may be substituted (Phillips et al., 2006), with a corresponding change in what the AUC represents. Because AUC examines discrimination over the entire range of predicted values, no threshold for binary classification is used. This aspect has helped increase its popularity compared to some other discrimination measures such as TSS (True Skill Statistic, (Allouche et al. 2006)), or Cohen's Kappa (Fielding & Bell, 1997), which require a threshold to be specified. Measures which require a threshold give differing judgements of

performance depending on the threshold selection, which can make it harder to interpret overall model performance (Fielding & Bell, 1997).

AUC is scaled from 0 to 1. A value of 1 represents perfect discrimination; 0.5 shows discrimination no better than a random prediction; and a value less than 0.5 shows worse discrimination than a random prediction.

1.9.2 Significance tests

Another means of evaluation is the testing for statistical significance of spatial predictions. The null hypothesis is typically that the model predicts the class membership of test data no better than a random model. This requires model predictions to be classed as presence or absence. When traditional species absence data is not used (such as in presence-background models), a modification is required. The program MaxEnt performs significance tests on models, using the null hypothesis that the model predicts the test data (presence only) no better than a random model with the same fractional predicted area (Phillips et al., 2006). Fractional predicted area is, for a finite geographic area, the proportion of grid cells predicted as presences.

1.9.3 Which evaluation measures are best for SDM of invasive species?

A caveat of using AUC for models of range expanding species is that ranking the greatest proportion of presences higher than 'absences' (or background) is no longer desirable. Due to the inherent dispersal limitations faced by range expanding species, some of the background which constitutes the 'absences' can be expected to be suitable. This is compounded by imperfect detection of species presence. A very high AUC score for a species which has not yet expanded to fill its entire potential range, and/or which has not been sampled perfectly, could indicate an overfitted model (except when obtained from truly independent test data; see section 1.9.4 below). AUC is still a useful measure, but must be interpreted in the context.

A number of techniques and metrics exist for the evaluation of model performance. There has been much discussion in the literature on the relative merits of each (e.g. Fielding & Bell, 1997; Allouche et al., 2006; Lobo et al., 2008), and the applicability of each to different SDM applications. In particular it has been noted that certain metrics are less suited to the special case of invading species (Jimenez-Valverde et al., 2011). This stems from the fact that an invading species is unlikely to be in equilibrium with its environment. The usual case is of an expanding species, which is yet to disperse into its entire suitable range (though note that imperfect dispersal and historic influences such as disturbance can cause suitable habitat to be unoccupied even for long-established or native species (e.g. Leathwick, 1998)). Disequilibrium during invasion can also take the converse case, where in some stages of invasion the distribution is greater than the eventual equilibrium distribution. An example is the apparent colony collapse of Argentine ants (*Linepithema humile*) in New Zealand (Cooling et al., 2011).

Because of disequilibrium, measures incorporating specificity are, at best, useful only with cautious interpretation. For plants, which are generally less likely to occupy unsuitable habitat than vagile animals, sensitivity scores can generally be used without great risk of misinterpretation. Specificity is more problematic. Ostensibly a measure of a model's ability to predict unsuitable habitat, it is directly influenced by the proportion of suitable habitat that remains un-colonised.

Sensitivity is often of heightened importance in iSDM (Invasive Species Distribution Modelling), since costs are likely to be higher when suitable habitat is predicted to be unsuitable. When there is a lack of accurate knowledge of where a species is absent due to unsuitable environment, such as the case of invading species, there is a risk that high sensitivity results from over-predicted prevalence (Franklin, 2009). A model predicting a prevalence of 100% would have perfect sensitivity, but zero specificity. A solution is to calculate the statistical significance of the number of correctly predicted presences, given the proportion of the area predicted as suitable (Phillips et al., 2006).

1.9.4 Obtaining test data

Model evaluation requires data against which to test predictions. Test data may be obtained by re-substitution, data splitting, or an independent source (Fielding & Bell, 1997; Jeschke & Strayer, 2008). Re-substitution involves testing the model using the same data with which it was trained, and does not constitute an adequate test of predictive performance. There are several methods by which data can be withheld, including re-substitution, *k*-fold cross validation, 'leave-one-out' (*sensu* Hengl et al., 2009), and bootstrapping (see Fielding & Bell, 1997).

Ideally, test data should be independent from training data. Test data taken from the same geographic region/s as training data are compromised in their independence by spatial autocorrelation (Legendre, 1993). This is true of data from 'independent' sources (e.g. field collection) as well as data drawn randomly from the same pool as training data. An often cited solution is to use geographically separate regions as the folds, a process referred to as *k*-fold regional cross-validation (Hartley et al., 2006; Jimenez-Valverde et al., 2011). For each iteration, one region is withheld as a test region and the rest are pooled for training. Not only does this approach ensure independence of test data, it gives a more direct measure of the model's ability to predict to new geographic regions (Jimenez-Valverde et al., 2011), which is the purpose of this study.

A rarely used approach in predictive modelling is to experimentally plant seeds of the study species across a broad range of the environmental variables used to build the model and to measure plant performance (Wright et al., 2006). This technique has obvious ethical limitations for use with invasive species.

1.10 Research aims

This study offers a predicted potential distribution of *S. glastifolius* in New Zealand. *Senecio glastifolius* is native to South Africa and was first recorded in New Zealand in 1963. It is currently regarded as an environmental weed, and is the subject of considerable control efforts.

Conservation managers can presently only speculate upon the future impacts of *S. glastifolius* in New Zealand. For example, the wide shingle banks of the South Island's many braided rivers have been suspected to be potentially vulnerable to invasion (S. Timmins, *pers. comm.*).

1.11 Thesis structure

The remainder of thesis is organised into two major sections. In the first section, chapter two describes the current and historical distribution of *S. glastifolius*, with emphasis on New Zealand. The result of the work described in this chapter is a new data set appropriate for use in SDM. In the second section, chapter three predicts the potential distribution of *S. glastifolius* in New Zealand, with implications for its management. The SDM algorithms MaxEnt and BIOCLIM are used on the New Zealand and global data. Throughout the process of model training, parameterisation and testing, special attention is given to the conceptual challenges posed by range-expanding species. Key messages of the research, plus possible avenues of further enquiry are discussed in chapter four.

1.12 References

- Allouche, O., A. Tsoar, et al. (2006). "Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS)." Journal of Applied Ecology **43**(6): 1223-1232.
- Andrew, M. E. and S. L. Ustin (2009). "Effects of microtopography and hydrology on phenology of an invasive herb." Ecography **32**(5): 860-870.
- Austin, M. P. (1980). "Searching for a model for use in vegetation analysis." Vegetatio **42**: 11-21.
- Austin, M. P. (2002). "Spatial prediction of species distribution: an interface between ecological theory and statistical modelling." Ecological Modelling **157**(2-3): 101-118.
- Austin, M. P., L. Belbin, et al. (2006). "Evaluation of statistical models used for predicting plant species distributions: Role of artificial data and theory." Ecological Modelling **199**(2): 197-216.

- Austin, M. P. and J. A. Meyers (1996). "Current approaches to modelling the environmental niche of eucalypts: Implication for management of forest biodiversity." Forest Ecology and Management **85**(1-3): 95-106.
- Baldwin, R. A. (2009). "Use of Maximum Entropy Modeling in Wildlife Research." Entropy **11**(4): 854-866.
- Barve, N., V. Barve, et al. (2011). "The crucial role of the accessible area in ecological niche modeling and species distribution modeling." Ecological Modelling **222**(11): 1810-1819.
- Beans, C. M., F. F. Kilkenny, et al. (2012). "Climate suitability and human influences combined explain the range expansion of an invasive horticultural plant." Biological Invasions.
- Breiman, L., et al. (1984). Classification and regression trees. Chapman and Hall.
- Broennimann, O., U. A. Treier, et al. (2007). "Evidence of climatic niche shift during biological invasion". Ecology Letters **10**(8): 701-709.
- Busby, J. R. (1991). BIOCLIM - A Bioclimatic Analysis and Prediction System. Nature Conservation: Cost Effective Biological Surveys and Data Analysis. C. R. Margules and M. P. Austin. Canberra, CSIRO: 64-68.
- Buswell, J. M., A. T. Moles, et al. (2011). "Is rapid evolution common in introduced plant species?" Journal of Ecology **99**(1): 214-224.
- Capinha, C., B. Leung, et al. (2011). "Predicting worldwide invasiveness for four major problematic decapods: an evaluation of using different calibration sets." Ecography **34**(3): 448-459.
- Carpenter, G., A. N. Gillison, et al. (1993). "DOMAIN — A flexible modeling procedure for mapping potential distributions of plants and animals". Biodiversity and Conservation **2**(667–680).
- Chen, H., L. J. Chen, et al. (2007). "Predicting the potential distribution of invasive exotic species using GIS and information-theoretic approaches: A case of ragweed (*Ambrosia artemisiifolia* L.) distribution in China." Chinese Science Bulletin **52**(9): 1223-1230.
- Cooling, M. et al. (2011). "The widespread collapse of an invasive species: Argentine ants (*Linepithema humile*) in New Zealand". Biology Letters.rsbl.2011.1014v1-rsbl20111014.
- Costall, J. A., R. J. Carter, et al. (2006). "The endemic tree *Corynocarpus laevigatus* (karaka) as a weedy invader in forest remnants of southern North Island, New Zealand." New Zealand Journal of Botany**44**: 5-22.

- Coudun, C., J. C. Gegout, et al. (2006). "Soil nutritional factors improve models of plant species distribution: an illustration with *Acer campestre* (L.) in France." Journal of Biogeography**33**(10): 1750-1763.
- Davis, M.B. & Shaw, R.G. (2001). "Range shifts and adaptive responses to Quaternary climate change". Science**292**: 673–679.
- Dormann, C. F. (2007). "Assessing the validity of autologistic regression." Ecological Modelling**207**(2-4): 234-242.
- Dormann, C. F., S. J. Schymanski, et al. (2012). "Correlation and process in species distribution models: bridging a dichotomy." Journal of Biogeography.
- Drake, J. M., C. Randin, et al. (2006). "Modelling ecological niches with support vector machines." Journal of Applied Ecology**43**(3): 424-432.
- Dudik, M., S. J. Phillips, et al. (2007). "Maximum entropy density estimation with generalized regularization and an application to species distribution modeling." Journal of Machine Learning Research**8**: 1217-1260.
- Elith, J., S. Ferrier, et al. (2005). "The evaluation strip: A new and robust method for plotting predicted responses from species distribution models." Ecological Modelling**186**(3): 280-289.
- Elith, J., C. H. Graham, et al. (2006). "Novel methods improve prediction of species' distributions from occurrence data." Ecography**29**(2): 129-151.
- Elith, J., M. Kearney, et al. (2010). "The art of modelling range-shifting species". Methods in Ecology & Evolution.
- 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**: 677-697.
- Elith, J., J. R. Leathwick, et al. (2008). "A working guide to boosted regression trees." Journal of Animal Ecology**77**(4): 802-813.
- Elith, J., S. J. Phillips, et al. (2011). "A statistical explanation of MaxEnt for ecologists." Diversity and Distributions**17**(1): 43-57.
- Farber, O. and R. Kadmon (2003). "Assessment of alternative approaches for bioclimatic modeling with special emphasis on the Mahalanobis distance." Ecological Modelling**160**(1-2): 115-130.
- Fielding, A. H. and J. F. Bell (1997). "A review of methods for the assessment of prediction errors in conservation presence/absence models." Environmental Conservation**24**(1): 38-49.

- Fitzpatrick, M. C., J. F. Weltzin, et al. (2007). "The biogeography of prediction error: why does the introduced range of the fire ant over-predict its native range?" Global Ecology and Biogeography**16**(1): 24-33.
- Franklin, J. (2009). Mapping species distributions : spatial inference and prediction. Cambridge; New York, Cambridge University Press.
- Gallagher, R. V., L. J. Beaumont, et al. (2010). "Evidence for climatic niche and biome shifts between native and novel ranges in plant species introduced to Australia." Journal of Ecology**98**(4): 790-799.
- Gaston, K. J. (2009). "Geographic range limits of species." Proceedings of the Royal Society B-Biological Sciences**276**(1661): 1391-1393.
- Graham, C. H., S. Ferrier, et al. (2004). "New developments in museum-based informatics and applications in biodiversity analysis." Trends in Ecology & Evolution**19**(9): 497-503.
- Grinnell, J. (1917). "Field Tests of Theories Concerning Distributional Control." American Naturalist**51**(602): 115-128.
- Guisan, A. and W. Thuiller (2005). "Predicting species distribution: offering more than simple habitat models." Ecology Letters**8**(9): 993-1009.
- Guisan, A. and N. E. Zimmermann (2000). "Predictive habitat distribution models in ecology." Ecological Modelling**135**(2-3): 147-186.
- Guisan, A., O. Broennimann, et al. (2006). "Using niche-based models to improve the sampling of rare species." Conservation Biology**20**(2): 501-511.
- Hanski, I. (1999). Metapopulation Ecology. Oxford, UK, Oxford University Press.
- Hartley, S., R. Harris, et al. (2006). "Quantifying uncertainty in the potential distribution of an invasive species: climate and the Argentine ant." Ecology Letters**9**(9): 1068-1079.
- Hastie, T. J., and R. Tibshirani. (1990). Generalized additive models, Chapman and Hall.
- Heikkinen, R. K., M. Marmion, et al. (2012). "Does the interpolation accuracy of species distribution models come at the expense of transferability?" Ecography**35**(3): 276-288.
- Hengl, T., H. Sierdsema, et al. (2009). "Spatial prediction of species' distributions from occurrence-only records: combining point pattern analysis, ENFA and regression-kriging." Ecological Modelling**220**: 3499–3511.
- Hijmans, R. J., S. E. Cameron, et al. (2005). "Very high resolution interpolated climate surfaces for global land areas." International Journal of Climatology**25**(15): 1965-1978.

- Hirzel, A., J. Hausser, et al. (2002). "Ecological-niche factor analysis: how to compute habitat-suitability maps without absence data?" Ecology**83**(7): 2027–2036.
- Hirzel, A. H., B. Posse, et al. (2004). "Ecological requirements of reintroduced species and the implications for release policy: the case of the bearded vulture." Journal of Applied Ecology**41**(6): 1103-1116.
- Hutchinson, G. E. (1957). Concluding remarks. Cold Spring Harbor Symposia on Quantitative Biology, Volume 22, Cold Spring Harbor (NY), Cold Spring Harbor Laboratory Press.
- Huntley, B. (1999). "Species distribution and environmental change: considerations from the site to the landscape scale". Ecosystem management: questions for science and society (ed. by E. Maltby, M. Holdgate, M. Acreman and A. Weir), pp. 115–130. Royal Holloway Institute for Environmental Research, Virginia Water, UK.
- Iverson, L. R., A. M. Prasad, et al. (2008). "Estimating potential habitat for 134 eastern US tree species under six climate scenarios." Forest Ecology and Management**254**(3): 390-406.
- Jeschke, J. M. and D. L. Strayer (2008). Usefulness of bioclimatic models for studying climate change and invasive species. Year in Ecology and Conservation Biology 2008. **1134**: 1-24.
- Jimenez-Valverde, A., A. T. Peterson, et al. (2011). "Use of niche models in invasive species risk assessments." Biological Invasions**13**(12): 2785-2797.
- Kadmon, R., O. Farber, et al. (2004). "Effect of roadside bias on the accuracy of predictive maps produced by bioclimatic models." Ecological Applications**14**(2): 401-413.
- Kearney, M. (2006). "Habitat, environment and niche: what are we modelling?" Oikos**115**(1): 186-191.
- Kearney, M., B. L. Phillips, et al. (2008). "Modelling species distributions without using species distributions: the cane toad in Australia under current and future climates." Ecography**31**(4): 423-434.
- Kearney, M. and W. P. Porter (2004). "Mapping the fundamental niche: Physiology, climate, and the distribution of a nocturnal lizard." Ecology**85**(11): 3119-3131.
- Kearney, M. and W. Porter (2009). "Mechanistic niche modelling: combining physiological and spatial data to predict species' ranges." Ecology Letters**12**(4): 334-350.

- Kilroy, C., T. H. Snelder, et al. (2008). "A rapid technique for assessing the suitability of areas for invasive species applied to New Zealand's rivers." Diversity and Distributions**14**(2): 262-272.
- Kitron, U. (1998). "Landscape Ecology and Epidemiology of Vector-Borne Diseases: Tools for Spatial Analysis." Journal of Medical Entomology**35**(4): 435-445.
- Knick, S. T. and J. T. Rotenberry (1998). "Limitations to Mapping Habitat Use Areas in Changing Landscapes Using the Mahalanobis Distance Statistic." Journal of Agricultural, Biological, and Environmental Statistics, **3**(3): 311-322
- Latimer, A. M., S. S. Wu, et al. (2006). "Building statistical models to analyze species distributions." Ecological Applications**16**(1): 33-50.
- Le Maitre, D. C., W. Thuiller, et al. (2008). "Developing an approach to defining the potential distributions of invasive plant species: a case study of *Hakea* species in South Africa." Global Ecology and Biogeography**17**(5): 569-584.
- Leathwick, J. (1998). "Are New Zealand's *Nothofagus* Species in Equilibrium with Their Environment?" Journal of Vegetation Science**9**(5): 719-732.
- Lee, J. E., Janion, C., Marais, E., van Vuuren, B. J. & Chown, S. L. (2009). "Physiological tolerances account for range limits and abundance structure in an invasive slug". Proc. R. Soc. B 276, 1459–1468. (doi:10.1098/rspb.2008.1240)
- Legendre, P. (1993). "Spatial autocorrelation - trouble or new paradigm." Ecology**74**(6): 1659-1673.
- Li, W., Guo, Q. and Elkan, C. (2011). "Can we model the probability of presence of species without absence data?" Ecography**34**: 1096–1105. doi: 10.1111/j.1600-0587.2011.06888.x
- Liu, C. R., P. M. Berry, et al. (2005). "Selecting thresholds of occurrence in the prediction of species distributions." Ecography**28**(3): 385-393.
- Lobo, J. M., A. Jimenez-Valverde, et al. (2008). "AUC: a misleading measure of the performance of predictive distribution models." Global Ecology and Biogeography**17**(2): 145-151.
- Lobo, J. M., A. Jimenez-Valverde, et al. (2010). "The uncertain nature of absences and their importance in species distribution modelling." Ecography**33**(1): 103-114.
- Lodge, D. M. and K. Shrader-Frechette (2003). "Nonindigenous Species: Ecological Explanation, Environmental Ethics, and Public Policy *Especies No Nativas: Explicación Ecológica, Ética Ambiental y Política Pública*." **17**: 31-37.
- Mack, R. N., D. Simberloff, et al. (2000). "Biotic invasions: causes, epidemiology, global consequences, and control." Ecological Applications**10**: 689–710.

- Mac Nally, R. (2000). "Regression and model-building in conservation biology, biogeography and ecology: The distinction between – and reconciliation of – ‘predictive’ and ‘explanatory’ models." Biodiversity and Conservation**9**: 655–671.
- Manning, A.D. et al. (2005). "A bioclimatic analysis for the highly mobile Superb Parrot of south-eastern Australia". Emu**105**: 193–201.
- Maron, J. L., M. Vila, et al. (2004). "Rapid evolution of an invasive plant." Ecological Monographs**74**(2): 261-280.
- Martinez-Meyer, E., A. Townsend Peterson, et al. (2004). "Ecological niches as stable distributional constraints on mammal species, with implications for Pleistocene extinctions and climate change projections for biodiversity." Global Ecology and Biogeography**13**(4): 305-314.
- Mateo, R. G., T. B. Croat, et al. (2010). "Profile or group discriminative techniques? Generating reliable species distribution models using pseudo-absences and target-group absences from natural history collections." Diversity and Distributions**16**(1): 84-94.
- Mgidi, T. N., D. C. Le Maitre, et al. (2007). "Alien plant invasions – incorporating emerging invaders in regional prioritization: a pragmatic approach for Southern Africa." Journal of Environmental Management**84**: 173–187.
- Moisen, G. G. and T. S. Frescino (2002). "Comparing five modelling techniques for predicting forest characteristics." Ecological Modelling**157**(2-3): 209-225.
- Mota-Vargas, C. and O. R. Rojas-Soto (2012). "The importance of defining the geographic distribution of species for conservation: The case of the Bearded Wood-Partridge." Journal for Nature Conservation**20**(1): 10-17.
- Mukherjee, A., M. C. Christman, et al. (2011). "Prioritizing areas in the native range of hygrophila for surveys to collect biological control agents." Biological Control**56**(3): 254-262.
- Mukherjee, A., R. Diaz, et al. (2012). "Niche-based prediction of establishment of biocontrol agents: an example with *Gratiana boliviana* and tropical soda apple." Biocontrol Science and Technology**22**(4): 447-461.
- Norton, D. A. (2009). "Species Invasions and the Limits to Restoration: Learning from the New Zealand Experience." Science**325**(5940): 569-571
- Olden JD, Lawler JJ, Poff NL. (2008). "Machine learning methods without tears: a primer for ecologists". Q. Rev. Biol. 83:171–93
- Pearson, R. G. and T. P. Dawson (2003). "Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful?" Global Ecology and Biogeography**12**(5): 361-371.

- Pearson, R. G., T. P. Dawson, et al. (2004). "Modelling species distributions in Britain: a hierarchical integration of climate and land-cover data." Ecography**27**(3): 285-298.
- Peterson, A. T. (2003). "Predicting the geography of species' invasions via ecological niche modeling." Quarterly Review of Biology**78**(4): 419-433.
- Peterson, A. T., J. Soberon, et al. (1999). "Conservatism of ecological niches in evolutionary time." Science**285**(5431): 1265-1267.
- Peterson, A. T. and D. A. Vieglais (2001). "Predicting species invasions using ecological niche modeling: New approaches from bioinformatics attack a pressing problem." Bioscience**51**(5): 363-371.
- Pfeiffer, J. M. and R. A. Voeks (2008). "Biological invasions and biocultural diversity: linking ecological and cultural systems." Environmental Conservation**35**(4): 281–293.
- Pheloung, P. C., P. A. Williams, et al. (1999). "A weed risk assessment model for use as a biosecurity tool evaluating plant introductions." Journal of Environmental Management**57**(4): 239-251.
- Phillips, S. J., R. P. Anderson, et al. (2006). "Maximum entropy modeling of species geographic distributions." Ecological Modelling**190**(3-4): 231-259.
- Phillips, S. J. and M. Dudik (2008). "Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation." Ecography**31**(2): 161-175.
- Phillips, S. J., M. Dudik, et al. (2009). "Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data." Ecological Applications**19**(1): 181-197.
- Pimentel, D., R. Zuniga, et al. (2005). "Update on the environmental and economic costs associated with alien-invasive species in the United States." Ecological Economics**52**: 273– 288.
- Prasad AM, Iverson LR, Liaw A. (2006). "Newer classification and regression tree techniques: bagging and random forests for ecological prediction". Ecosystems**9**:181–99
- Richardson, D. M., J. Carruthers, et al. (2011). "Human-mediated introductions of Australian acacias – a global experiment in biogeography." Diversity and Distributions**17**(5): 771-787.
- Richardson, D. M., P. Pysek, et al. (2000). "Naturalization and invasion of alien plants: concepts and definitions." Diversity and Distributions**6**(2): 93–107.
- Rodda, G. H., C. S. Jarnevich, et al. (2011). "Challenges in Identifying Sites Climatically Matched to the Native Ranges of Animal Invaders." Plos One**6**(2): e14670.

- Sexton, J.P., McKay, J.K. & Sala, A. (2002). "Plasticity and genetic diversity may allow saltcedar to invade cold climates in North America". Ecol. Appl., 12, 1652–1660.
- Soberon, J. (2007). "Grinnellian and Eltonian niches and geographic distributions of species." Ecology Letters**10**(12): 1115-1123.
- Soberon, J. and A. T. Peterson (2011). "Ecological niche shifts, and environmental space anisotropy: a cautionary note." Revista Mexicana De Biodiversidad**82**(4): 1348-1355.
- Soberon, J. and M. Nakamura (2009). "Niches and distributional areas: Concepts, methods, and assumptions." Proceedings of the National Academy of Sciences of the United States of America**106**: 19644-19650.
- Stockwell D, Peters D. (1999). "The GARP modelling system: problems and solutions to automated spatial prediction". Int. J. Geogr. Inf. Sci. **13**:143–58
- Stockwell, D. R. B. and A. T. Peterson (2002). "Effects of sample size on accuracy of species distribution models." Ecological Modelling**148**(1): 1-13.
- Stokland, J. N., R. Halvorsen, et al. (2011). "Species distribution modelling-Effect of design and sample size of pseudo-absence observations." Ecological Modelling**222**(11): 1800-1809.
- VanDerWal, J., L. P. Shoo, et al. (2009). "Selecting pseudo-absence data for presence-only distribution modeling: How far should you stray from what you know?" Ecological Modelling**220**(4): 589-594.
- Venables, W. M. and B. D. Ripley (1994). Modern Applied Statistics with S-Plus. New York, Springer-Verlag.
- Vitousek, P. M., D'Antonio, C. M., Loope, L. L., Rejmanek, M. and Westbrooks, R. 1997. Introduced species: a significant component of human-caused global change. New Zealand Journal of Ecology, 21(1):1-16.
- Warren, D. L. and S. N. Seifert (2011). "Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria." Ecological Applications**21**(2): 335-342.
- Watt, M. S., J. K. Stone, et al. (2011). "Using a climatic niche model to predict the direct and indirect impacts of climate change on the distribution of Douglas-fir in New Zealand." Global Change Biology**17**: 3608–3619.
- Webber, B. L., C. J. Yates, et al. (2011). "Modelling horses for novel climate courses: insights from projecting potential distributions of native and alien Australian acacias with correlative and mechanistic models." Diversity and Distributions**17**: 978-1000.

- Williams, N. S. G., A. K. Hahs, et al. (2008). "A dispersal-constrained habitat suitability model for predicting invasion of alpine vegetation." Ecological Applications**18**(2): 347-359.
- Wilson, K. A., M. I. Westphal, et al. (2005). "Sensitivity of conservation planning to different approaches to using predicted species distribution data." Biological Conservation**122**(1): 99-112.
- Wisz, M. S. and A. Guisan (2009). "Do pseudo-absence selection strategies influence species distribution models and their predictions? An information-theoretic approach based on simulated data." BMC Ecology**9**(8).
- Woodward, F.I. (1987) Climate and Plant Distribution. Cambridge University Press, Cambridge.
- Wright, J. W., K. F. Davies, et al. (2006). "Experimental verification of ecological niche modeling in a heterogeneous environment." Ecology**87**(10): 2433-2439.
- Yates, C. J., A. McNeill, et al. (2010). "Assessing the impacts of climate change and land transformation on Banksia in the South West Australian Floristic Region." Diversity and Distributions**16**(1): 187-201.
- Zaniewski, A. E., A. Lehmann, et al. (2002). "Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns." Ecological Modelling**157**(2-3): 261-280.
- Zurell, D., J. Elith, et al. (2012). "Predicting to new environments: tools for visualizing model behaviour and impacts on mapped distributions." Diversity and Distributions: 1-7.

Chapter 2: Describing the current distribution and historical spread of *Senecio glastifolius*

Abstract

Senecio glastifolius is a tall perennial herb native to part of South Africa's Cape region. It is naturalised in New Zealand, Australia, Great Britain and Madeira, and is considered invasive in most of its naturalised range. As is typically the case for many invasive species, the range of *S. glastifolius* is, so far, poorly described, with information incomplete and often imprecise. Details of its historical spread are similarly limited. This paucity of information limits quantitative ecological analyses drawing on distribution data, such as species distribution modelling. To address this knowledge gap, a new, more comprehensive dataset was compiled. A substantial body of expert knowledge was drawn upon, and extensive field observations made, to supplement information from herbaria and other databases. Information was collected and processed for issues of accuracy and reliability. Methods employed reflect the requirements of species distribution modelling (SDM), which is the primary intended use of the dataset. The result is a more complete and higher quality dataset describing the distribution and historical spread of *S. glastifolius*.

2.1 Introduction

Data describing the geographic distribution of species is an essential requirement of many studies in biogeography and ecology, including species distribution modelling. The data used for SDM are typically drawn opportunistically from existing sources (Newbold, 2010). The quality of the existing data, in reference to the research questions they are used to answer, is often inadequate, necessitating significant filtering and cleaning (Feeley & Silman, 2010; Webber et al., 2011). Existing data for *S. glastifolius* show many of the typical issues, including incomplete coverage of the actual geographic distribution, a lack of synthesis of existing data, and issues of imprecision and inaccuracy of georeferencing. To allow robust models of *S. glastifolius*'s potential distribution to be trained and validated, a new data set was developed describing its current distribution and historic spread.

2.1 Methods

2.1.1 Species occurrence data

Data on the observed distribution of *S. glastifolius* was drawn from its recorded global range. This includes its native South Africa plus naturalised populations in New Zealand, Australia and Great Britain's Isles of Scilly, but excludes Madeira and Guernsey for which no georeferenced records were available. Data were obtained through a number of channels as described in the following paragraphs.

For South Africa, data were taken from the PRECIS (SANBI, 2007) and ACKDAT (Rutherford et al., 2003) databases and cleaned extensively (see section 2.1.2 for methods of data cleaning). These were augmented by data provided by local ecologists Dave Edge (Edge, 2005) and Jan Vlok (pers. comm.), plus records obtained from a website containing georeferenced photographs of South Africa's flora and fauna (Anon., 2011a; Anon., 2011b).

For Great Britain, detailed data were provided by local botanist Rosemary Parslow (R. Parslow, *unpub. data*). One additional record was obtained from a photograph published online, accompanied by a detailed description of location (Fenwick, 2011).

For Australia, data were provided by David McNamara of Greenskills, Albany. This dataset constitutes many years of detailed observation from the Albany area. Data were also taken from the Australian Virtual Herbarium (AVH) and cleaned.

For New Zealand, data were sourced from various herbaria (AK, WELT, MPN, CHR, and NZFRI). Records were also taken from the Department of Conservation's unpublished database 'BIOWEB'.

Several ecologists, botanists and weeds officers from around New Zealand were contacted, all of whom offered their knowledge on the distribution of *S. glastifolius*. This list of experts consulted included Bob Brockie (formerly Victoria University of Wellington), Mike Dodd (AgResearch), Colin Ogle (Wanganui Museum Botanical Group), Mick Parsons (Wellington Botanical Society), Don Ravine (Ecologist), Craig Davey (Horizons Regional Council), Peter Williams (Landcare Research), Nick Singers (Ecologist), Kim Wright (Tasman District Council), Jill Rapson (Manawatu Botanical Society), Phil Karaitiana (Gisborne District Council), Darin Underhill

(Hawke's Bay Regional Council), Harvey Phillips (Greater Wellington Regional Council), and Catherine Law (Taranaki Regional Council).

The author also provided extensive personal observations from around the lower North Island, representing the period 1997 to 2012.

2.1.2 Data cleaning

Data from herbaria required extensive 'cleaning' (screening for detectable inaccuracies, and either removing or amending records as appropriate), with a large proportion of the records either too spatially imprecise or incorrectly georeferenced. Judging the level of spatial precision required is not an exact science, but it should be assessed relative to both the resolution of environmental data used for modelling and, perhaps more importantly, the spatial scale over which the environmental variables vary.

Previous studies have examined the effect of spatial imprecision in species occurrence data (Graham et al., 2008), with a general conclusion that commonly used modelling algorithms (e.g. MaxEnt, GAMs etc.) are fairly robust to some imprecision. No definitive guidelines exist as to what level of imprecision is acceptable. Indeed it would be hard to establish such guidelines, as spatial imprecision functions in conjunction with environmental heterogeneity to impart noise or bias into model predictions. This idea has recently been expressed in terms of spatial autocorrelation of environmental variables (Naimi et al., 2011). The extent to which a given level of imprecision affects model predictions depends on the environmental heterogeneity of the study region. Accordingly, no threshold of imprecision could apply equally to all study areas.

Coordinates of recorded species occurrences were compared to the accompanying location descriptions for congruency. Data points were corrected where location descriptions were sufficiently precise to allow this, by updating the coordinates to those of the location described. This was done only for a very small proportion of records. This sort of 'data tampering' requires keen circumspection on the part of the modeller.

Data imprecision is frequently encountered in herbarium data sets, resulting from both the purpose for which data were collected and the methods used for that collection. Specimens are often sent to herbaria with the chief purpose being to provide a sample of species' physical attributes. Herbaria certainly aim to provide a source of information of floristic distributions at a regional scale, but this scale is generally coarser than the often very fine resolution of SDM exercises (Franklin, 2009; McPherson et al., 2006). The method of data collection often limits precision. The personal, hand-held GPS, which is nowadays frequently used to provide location information for herbarium specimens, has had its advent very recently in relation to the period over which most herbarium collections have been collated. Records for *S. glastifolius*, for example, date back at least as far as 1881 (PRE0722577). Prior to the GPS, the greatest detail of a specimen's source location was often simply the name of the nearest town or significant geographic feature such as a river or mountain (*pers. obs.*).

Some data points fell outside of the extent of land represented in the gridded environmental data layers. In most cases the coordinates of these occurrence data were correct, but gridded data (in this case 30 second resolution) inevitably gives an imprecise, 'pixelated' representation of a coastline. Such presence data, missing environmental information due to this imprecision, can be either omitted or artificially shifted onshore. Omission will impart a bias in environmental space (penalising space associated with lower elevation and more maritime climates) and result in fewer training data, which can erode model accuracy (e.g. Wisz et al., 2008; Guisan et al., 2007). Artificial shifting will also bias data, by matching shifted presence records to the environments found immediately inland of their true locations. The severity of this bias depends on the steepness of environmental gradients from the shore inland, and knowledge of the environmental geography of the study area should be a necessary prerequisite to artificial data shifting. Several presence locations for *S. glastifolius* in New Zealand were artificially shifted onshore, subject to these cautions.

2.1.3 Data generalisation

After collation, species location data were generalised to remove duplicates from within grid cells used for the species distribution modelling exercise presented in the following chapter. The grids used had a resolution of 30 arc seconds, and used the WGS84 (World Geodetic System) geographic coordinate system. Technically, this could have imparted a slight latitudinal bias in occurrence records, since data were unprojected. In the case of the data set used it was judged that the bias would be so small as to be almost certainly inconsequential, since co-occurrence of records within grid cells was rare. But even small biases compound uncertainty and attention to this problem is certainly warranted in the general case of SDM.

2.3 Results

The compilation of data from all sources yielded a total of 309 geographically referenced records. A significant proportion was removed during processing (see Table 2.1).

Table 2.1 Distribution record processing

Region	Raw	Cleaned	Shifted	Generalised	Final
New Zealand	181	43	10	4	134
South Africa	67	31	3	0	36
Australia	46	6	1	0	40
Scilly	15	0	1	11	4
TOTAL	309	80	15	15	214

2.3.1 Native Range: South Africa

The native range of *S. glastifolius* is very small, spanning the part of South Africa's Cape Region from near Mossel Bay to Port Elizabeth. Its range comprises the seaward-most mountain ranges (Outeniqua, Tsitsikama and Kareedouw), extending to the coast. It is found up to an altitude of 1,090m (Germishuizen & Meyer 2003). It is not recorded as growing farther inland than the seaward ranges (see Figure 2.2). PRECIS contains five records of *S. glastifolius* from the Cape Peninsula, dated from

1881 to 1940. None of these records were precise enough for inclusion in the data set to be used for species distribution modelling. It is unknown whether these records represent part of the native distribution or whether they exist due to human transportation. The gap between the Cape Peninsula and the main population to the East is over 300km.



Figure 2.1.Location of the recorded native distribution of *Senecio glastifolius* in South Africa's Cape Region.

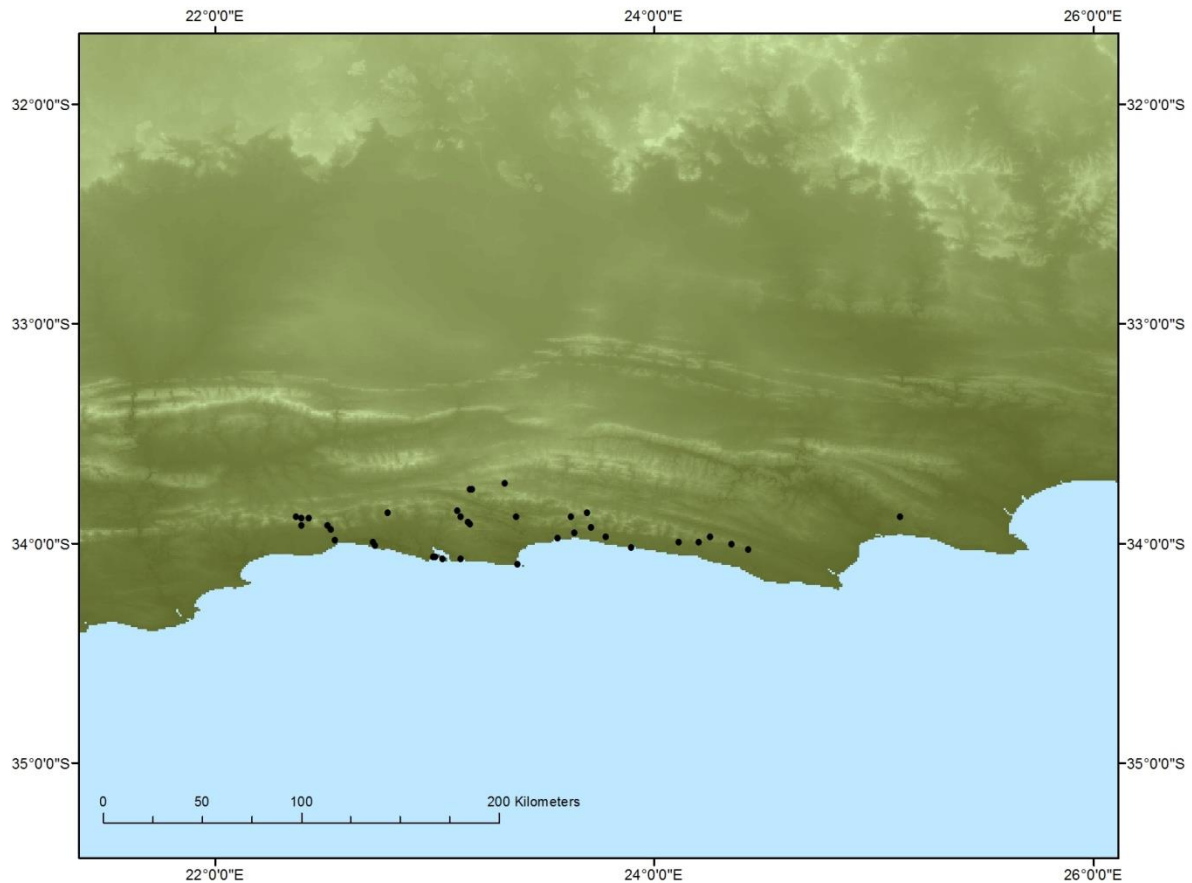


Figure 2.2. Recorded native distribution of *S. glastifolius* within South Africa's Cape region. Kilometre scale is indicative only.

2.3.2 Australia

Herbarium records suggest that *S. glastifolius* was first recorded in Western Australia in 1953, (PERTH 00544965) and in New South Wales in 1957 (NSW 615875).

Herbarium notes suggest, however, that true naturalisation in Western Australia was not evidenced until 1986 (PERTH 00790818). In Western Australia, it is found mostly around the city of Albany, with one previous outlier at Manjimup where it appears to no longer exist (Diane Evers, *pers. comm.*). Few records exist for New South Wales, where it appears to be sparse and limited to Royal National Park in the vicinity of Bundeena.

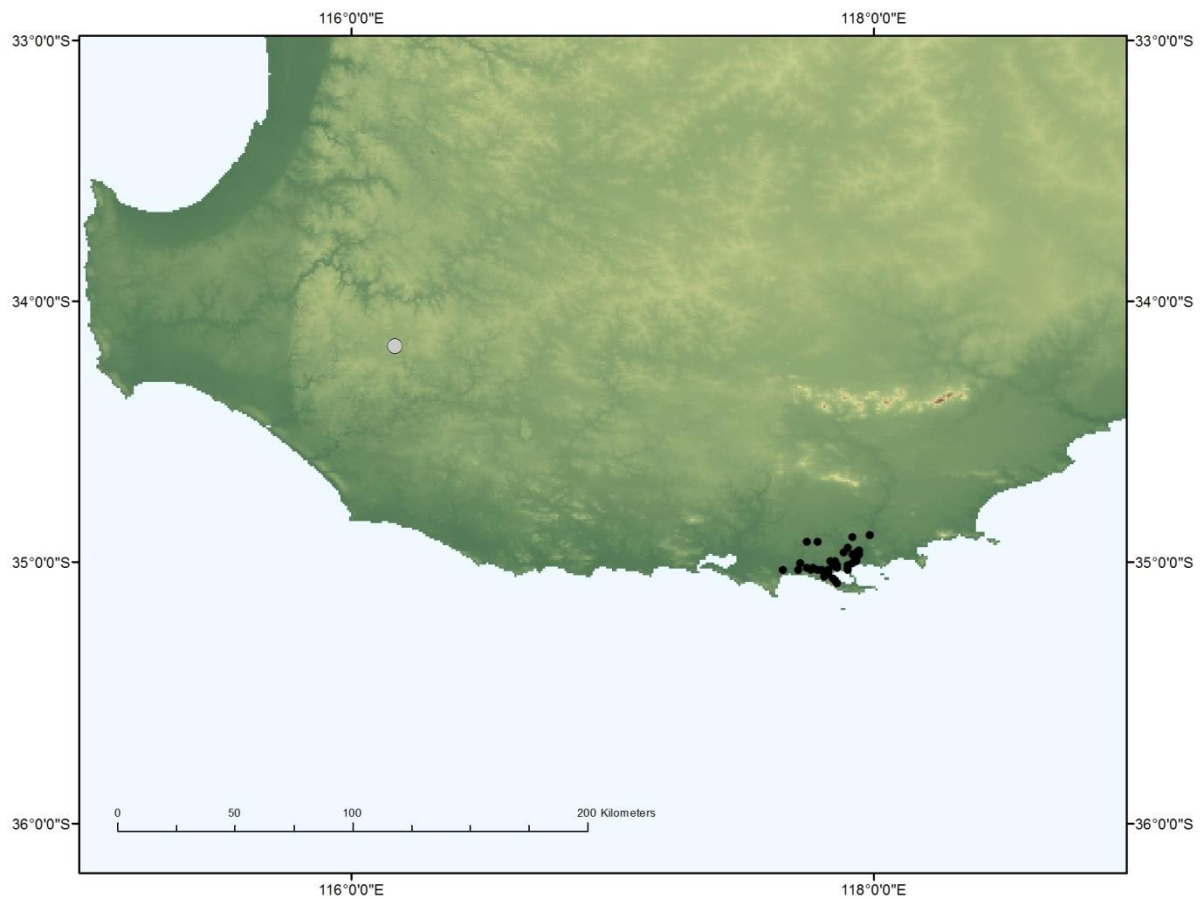


Figure 2.3.Recorded distribution of *S. glastifolius* in Western Australia. Grey circle represents a historic record (1963) where the species is believed to no longer occur. Kilometre scale is indicative only.

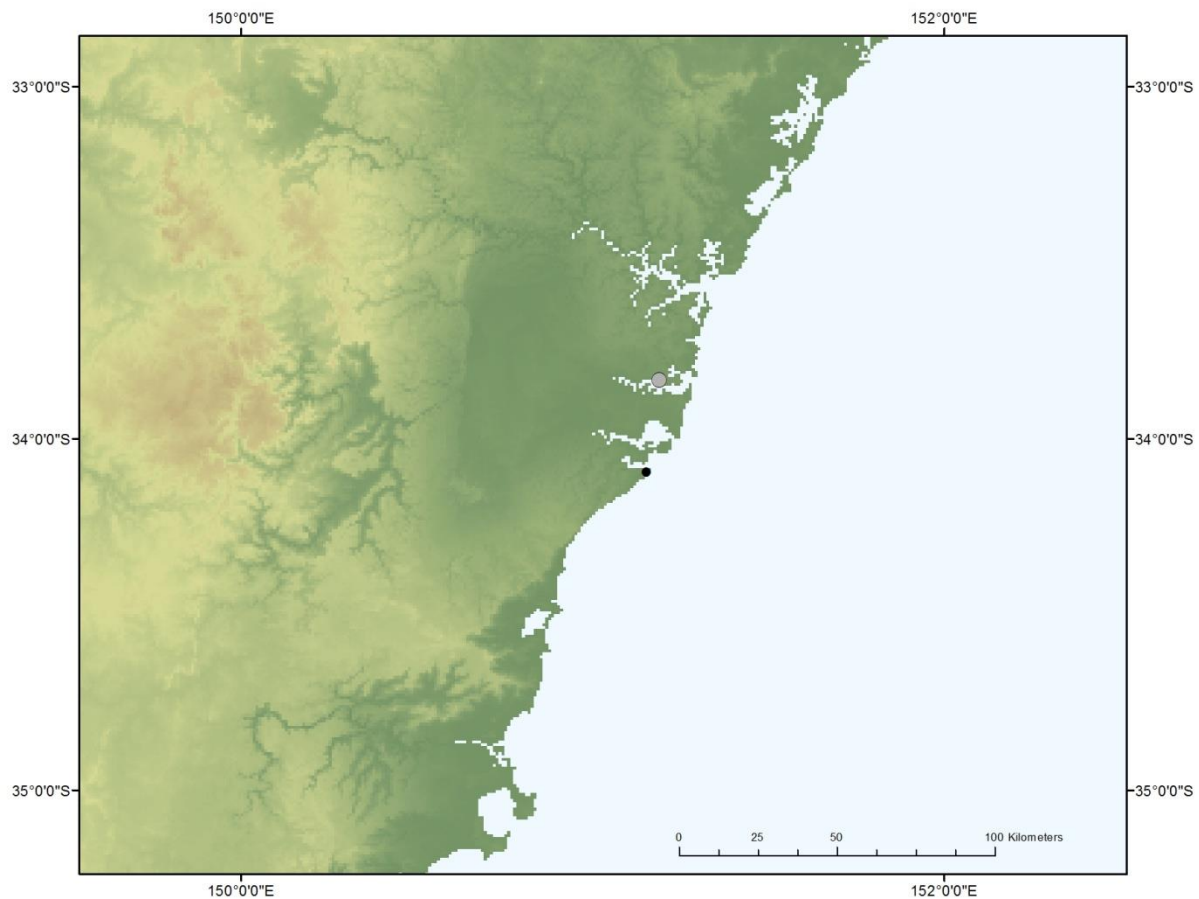


Figure 2.4. Recorded distribution of *S. glastifolius* in New South Wales. Grey circle represents a historic record (1957), where the current status of the plant is unknown. Kilometre scale is indicative only.

2.3.3 British Isles

Senecio glastifolius is naturalised and invasive in the Isles of Scilly, near Cornwall, United Kingdom. The first record is for 1971 (Preston et al., 2002) but it appears to have remained sparse until the 1990s. A separate source lists it as first naturalised in 1993 on the island of Tresco (Scilly), where it was cultivated at the Tresco Abbey Gardens (Rosemary Parslow, *pers. comm.*). It has since invaded the islands of St Mary's and St Martin's. It is found mostly in sand dunes in these isles.

It is also listed by Preston, Pearman & Dines (2002) as naturalised on rough ground in Guernsey. No further information on this population was found.

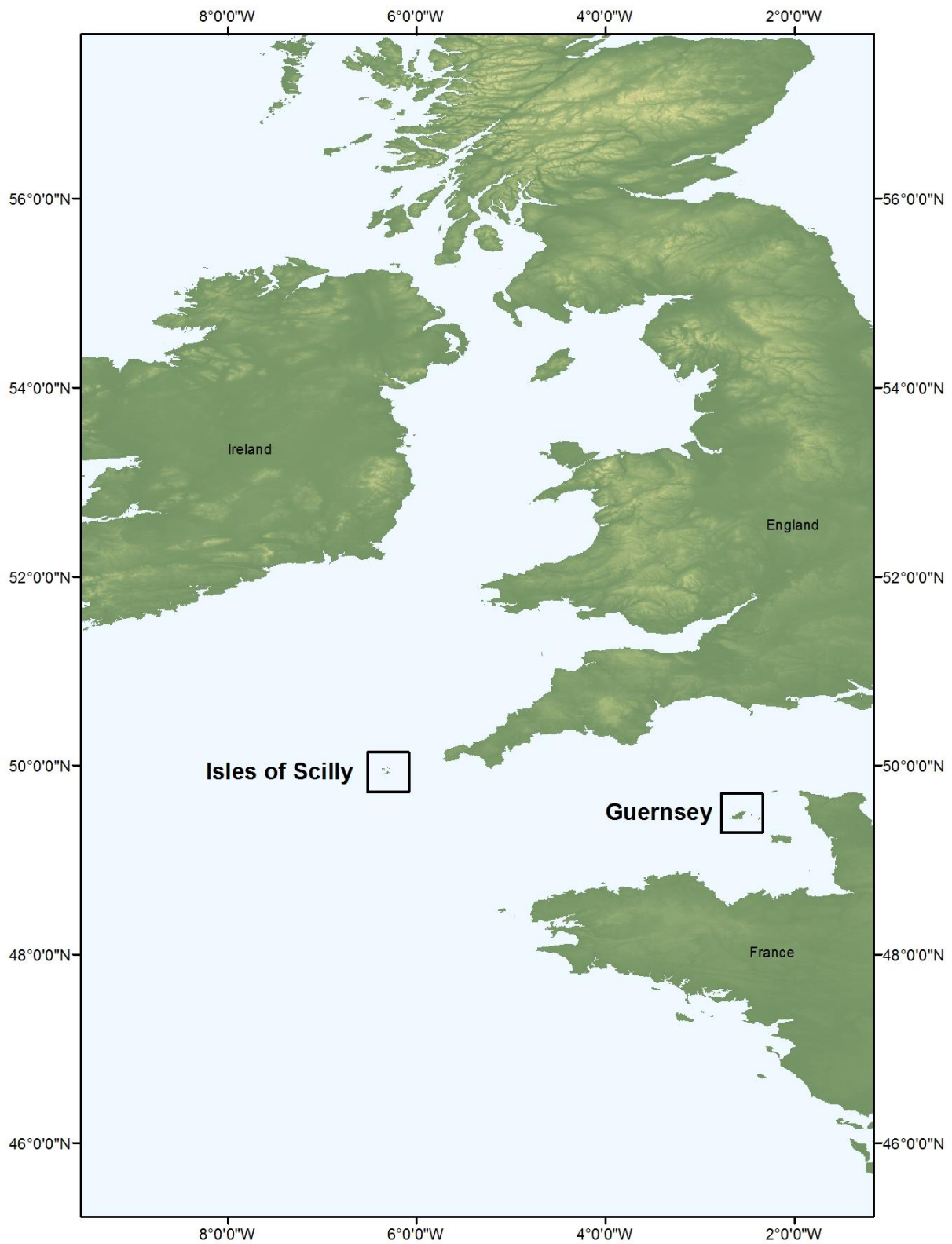


Figure 2.5. Location of the Isles of Scilly and Guernsey.

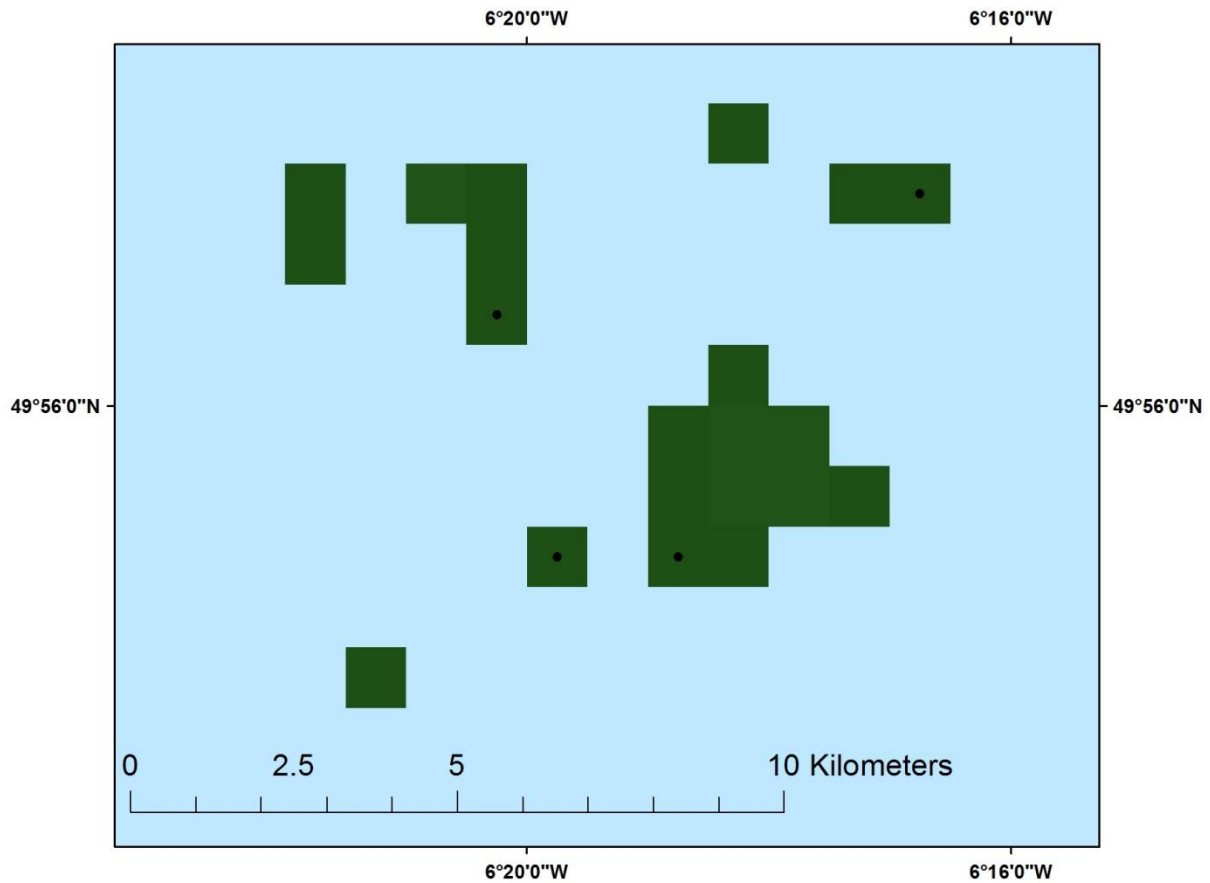


Figure 2.6. Map of the Isles of Scilly as represented by 30 arc second grid cells. Black dots represent occurrences of *S. glastifolius*. Kilometre scale is indicative only.

2.3.4 Madeira

Senecio glastifolius is recorded as naturalised in Madeira by several sources (Hansen & Sunding, 1993; Vieira 2002; Jardim & Sequeira, 2008), but no information was found regarding the exact locations, nor any invasive impacts there.



Figure 2.7.Location of Madeira archipelago.

2.3.5 New Zealand

Senecio glastifolius exists in several regions throughout central New Zealand (see Figure 2.8). Its range has expanded from separate points of naturalisation. Details are given in the following section (2.5.1). Accompanying maps can be found in Appendix 2.1.

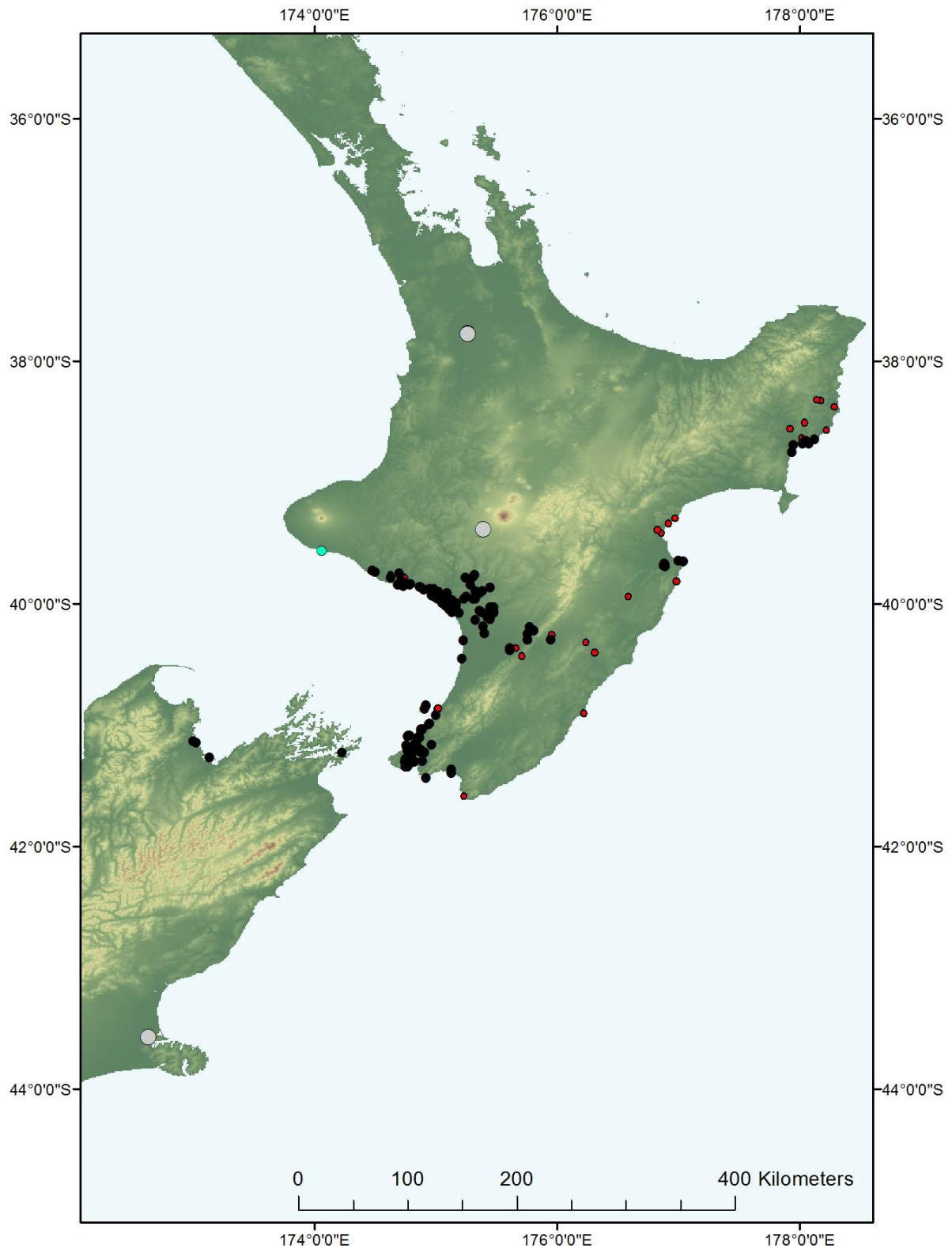


Figure 2.8.Recorded distribution of *S. glastifolius* in New Zealand. Red dots differentiate anecdotal records of imprecise location (up to ~10 km). Grey circles indicate historic occurrences believed to no longer exist (Christchurch, 1977; Hamilton, 1992; Hapuawhenua, 2008). Green dot represents a record found in April 2012 (not included in modelling – see following chapter). Kilometre scale bar is indicative only.

2.3.6 New Zealand invasion timeline:

Senecio glastifolius was first recorded as naturalised in New Zealand in **1963** as a garden escapee near Gisborne (CHR 143793). In the same year it was present at several locations around Wellington including Paekakariki, Paremata and Tawa (Bob Brockie, pers. comm.). It is likely that it had been present before 1963 in the Wellington region (B. Brockie, pers. comm.).

Side note: William Colenso erroneously recorded *S. glastifolius* as present in New Zealand in his 1865 work "Essay on the Botany, of the North Island of New Zealand" (Colenso, 1865). The plant he was probably referring to was *Brachyglottis kirkii* (Kirk) C.J. Webb, which has variously been known as *Senecio glastifolius* Hook. f. [illegitimate], *Senecio kirkii* Hook. f. ex Kirk, *Urostemon kirkii* (Hook. f. ex Kirk) B.Nord., *Senecio neo-zealandicus* Druce and *Solidago arborescens* A.Cunn. [illegitimate] (TPL, 2010).

By the **late 1960s** it was widespread in the Wellington Region from Island Bay to Paekakariki with significant populations around Tawa, as well as Paremata, Thorndon, Northland and Wadestown (B. Brockie, pers. comm.).

A large but highly localised population appeared in **1970** on a causeway near Motueka (CHR 220819). Occasional plants have since been found in the Tasman Bay area, on the Mapua Estuary and on Rabbit and Jacket Islands (Peter Williams; Kim Wright, pers. comm.).

In **1972** it was recorded at Havelock North (CHR 231945), the first record for Hawke's Bay.

By **1975** it had established as a 'local weed in waste areas' in Palmerston North (MPN 25437).

A single population of unspecified size, recorded in **1977**, existed briefly at one site in Cashmere, Christchurch (CHR 322943).

In **1984** the first plant appeared on Mana Island (WELT SP070458). Although this did not represent a major expansion in *S. glastifolius*'s range, it does offer some information on the plant's dispersal capability. It is likely the plant grew from seed blown from the mainland, at least 3.5km away (Colin Ogle, pers. comm.). It has also been recorded on Matiu/Somes Island in Wellington Harbour, over 2.5km from the mainland (DoC, unpublished data).

An important jump in its distribution was its invasion of the Whanganui coast, probably at some time in the **1980s** (Colin Ogle, *pers. comm.*). The first official record for the area is dated **1990** from Corliss Island (CHR 471405). From Whanganui, it was able to spread rapidly along the coast toward Wellington with the prevailing Westerly to North-West wind. The origin of this new, outlier population is not known. With no nearby populations at the time, the most likely hypothesis is human facilitation, e.g. deliberate planting as an ornamental garden plant.

By **1992** it was well established on the western cliffs of Kapiti Island (B Brockie, *pers. comm.*), some ~6km from the mainland.

Also in **1992**, it was recorded opposite Fairfield Bridge in urban Hamilton (CHR 480733). Despite the presence of several apparently naturalised specimens at the site (CHR), there have been no reports of it from the area since and it is unlikely that it established (Williams et al., 1999).

In **1994** it was found at Tawhiriho near the mouth of the Rangitikei River. It appeared to have arrived in contaminated gravel, growing only next to a gravel track alongside another newly arrived invasive plant *Eqisetum arvense* (Nick Singers, *pers. comm.*).

In **1998** it was recorded at Hawken's Lagoon conservation area (now called Tapuarau) (DoC, unpub. data), 30km north-west of the first record for the Whanganui-Taranaki population. It is unknown whether *S. glastifolius* dispersed naturally toward Taranaki from Whanganui or whether there were human mediated jumps in this expansion.

In **1999** it was found at Marton (CHR 532923). This was the farthest inland it had been recorded for the Whanganui-Rangitikei population (assuming this population spread with the prevailing wind and not against it from Palmerston North). Since then it has expanded over a wide area inland from Whanganui to Manawatu (see Figure 2.8).

In **2007** it was found at Arapawa Island, Marlborough sounds (DoC, unpub data). The dispersal origin of this plant is unknown. The record's location is over 30km from the North Island, and over 40km from the closest recorded population of the species. No information was found as to whether the species established successfully in this area.

In **2008** a single plant was found beneath the Hapuawhenua railway viaduct near Ohakune (Nick Singers, *pers comm.*). It is assumed that the plant grew from seed

transported in railway metal or by train. It is unlikely to be established in the area, due both to comparatively harsh local environmental conditions (altitude >600m) and because the only plant found was uprooted for accession to a herbarium (CHR 595304).

In **2009** two separate occurrences were found near Patea (Mick Parsons, pers. comm.). The species remains sparse in the area and is, in 2012, absent or undetected at significant areas of suitable ground cover (e.g. at the Patea River mouth).

Reports by local pest officers and personal field observation show that by **2010** it was widespread in the Gisborne area, Hawke's Bay, South Taranaki, Manawatu-Tararua area (see Figure 2.8).

In **2012** *S. glastifolius* was found by the author at the Kaupokonui River mouth, Taranaki. This population represents the leading edge of the known distribution moving north along this coast. The site is approximately 40 km further into Taranaki than the next known population at Patea. *Senecio glastifolius* is likely to occur between the two sites, as well as further north along the coast.

Maps illustrating the spread of *S. glastifolius* in New Zealand are presented in Appendix 2.1.

2.4 Discussion and conclusions

This data set shows that *S. glastifolius* has continued its rapid spread since its distribution was last described in detail by Williams et al. (1999).

The information presented here represents the most comprehensive and up-to-date single data set for the distribution of *S. glastifolius*. This creates opportunities for research, and represents a valuable data source not only for the following chapter, but for other researchers.

No data set provides a full census for a species, nor is it likely that all existing information will be synthesised into a single work. No illusion of completeness is implied for this data set.

A limitation of the data set is spatially uneven sampling effort. Observer effort is highest in and around areas of high population. An information gap exists for much of the southern Manawatu coast. No records were found for that area between the Manawatu River mouth and Waikanae (a distance of around 50 kilometres). No

absence information exists for this area either, except that *S. glastifolius* is absent or undetected at Otaki and Waikanae (pers. obs.; Bob Brockie, pers. comm.). Most of this section of coast is sparsely populated.

The anecdotal reports of ecologists, botanists and weeds officers represent a large body of useful information. However, the anecdotal nature of this body requires a judicious approach to its use and inclusion in modelling. Reports are best cross-referenced, as information is often conflicting, especially reports regarding the timing of invasion of new areas.

When interpreting these data and accompanying maps (see Appendix 2.1), attention should be given to the fact that the dates for each location are the date of recording and not explicitly the date of colonisation. For many of the records the species had probably been present for some time before the year in which it was recorded. This is particularly relevant to data for 2010. The apparent explosion in range is due not to rapid dispersal by *S. glastifolius* in that year, but is due to a significant expansion of the data set in that year. In some cases there is reasonable certainty that the records closely reflect the time of colonisation (e.g. Mana Island, 1984).

2.5 References

- Anonymous. (2011a). "*Senecio glastifolius*".
<http://za.ispot.org.uk/node/141714?nav=search> accessed 23 November 2011.
- Anonymous. (2011b). "*Senecio glastifolius*".
<http://za.ispot.org.uk/node/134474?nav=search> accessed 23 November 2011.
- Colenso, W. (1865). "Essay on the botany of the North Island of New Zealand".
- Edge, D.A. 2005. "Ecological factors influencing the survival of the Brenton Blue butterfly, *Orachrysops niobe* (Trimen) (Lepidoptera: Lycaenidae)". Potchefstroom: North-West University. (Thesis - D.Phil.)
- Feeley, K. J. and M. R. Silman (2010). "Modelling the responses of Andean and Amazonian plant species to climate change: the effects of georeferencing errors and the importance of data filtering." *Journal of Biogeography* **37**(4): 733-740.
- Fenwick, D. (2011). "*Senecio glastifolius* - Woad-leaved Ragwort".
http://www.aphotoflora.com/d_senecio_glastifolius_woad_leaved_ragwort.html accessed 23 November 2011.

- Franklin, J. (2009). Mapping species distributions: spatial inference and prediction. Cambridge; New York, Cambridge University Press.
- Germishuizen, G. and Meyer, N.L. (2003). "Plants of southern Africa: an annotated checklist". *Strelitzia*, 14, 1–1231. No 53.
- Graham, C. H., J. Elith, et al. (2008). "The influence of spatial errors in species occurrence data used in distribution models." Journal of Applied Ecology**45**(1): 239-247.
- Guisan, A., N. E. Zimmermann, et al. (2007). "What matters for predicting the occurrences of trees: Techniques, data, or species' characteristics?" Ecological Monographs**77**(4): 615-630.
- Hansen, A. & Sunding, P. (1993) - Flora of Macaronesia. Checklist of vascular plants. 4th revised edition - Sommerfeltia Oslo 17.
- Jardim, R. & Sequeira, M.M. (2008). "List of Vascular Plants (Pteridophyta and Spermatophyta)". In: Borges, P.A.V., Abreu, C., Aguiar, A.M.F., Carvalho, P., Jardim, R., Melo, I., Oliveira, P., Sérgio, C., Serrano, A.R.M. & Vieira, P. (eds.). *A list of the terrestrial fungi, flora and fauna of Madeira and Selvagens archipelagos*. pp.13-25, Direcção Regional do Ambiente da Madeira and Universidade dos Açores, Funchal and Angra do Heroísmo.
- McPherson, J. M., W. Jetz, et al. (2006). "Using coarse-grained occurrence data to predict species distributions at finer spatial resolutions-possibilities and limitations." Ecological Modelling**192**(3-4): 499-522.
- Naimi, B., A. K. Skidmore, et al. (2011). "Spatial autocorrelation in predictors reduces the impact of positional uncertainty in occurrence data on species distribution modelling." Journal of Biogeography**38**(8): 1497-1509.
- [SANBI] South African National Biodiversity Institute (2007). "National Herbarium Pretoria Computerised Information System (PRECIS)". Pretoria South Africa. Data acquired 04/08/2010.
- Newbold, T. (2010). "Applications and limitations of museum data for conservation and ecology, with particular attention to species distribution models." Progress in Physical Geography**34**(1): 3-22.
- Preston, C.D., Pearman, D.A. & Dines, T.D. (2002) *New Atlas of the British and Irish Flora*, Oxford, Oxford University Press.
- [TPL] The Plant List (2010). Version 1. Published on the Internet; <http://www.theplantlist.org/> (accessed 1st January, 2012).
- Rutherford, M. C., L. W. Powrie, et al. (2003). "ACKDAT: a digital spatial database of distributions of South African plant species and species assemblages." South African Journal of Botany**69**(1): 99-104.

- Vieira, R. M. da S. (2002) – “Flora da Madeira. Plantas vasculares naturalizadas no arquipélago da Madeira”. Boletim do Museu Municipal do Funchal (Hist. Nat.) Supl. 8.
- Webber, B. L., C. J. Yates, et al. (2011). "Modelling horses for novel climate courses: insights from projecting potential distributions of native and alien Australian acacias with correlative and mechanistic models." Diversity and Distributions**17**: 978-1000.
- Williams, P. A., C. C. Ogle, et al. (1999). "Biology and ecology of *Senecio glastifolius* and its spread and impacts in New Zealand". Science for Conservation **112**: 22.
- Wisz, M. S., R. J. Hijmans, et al. (2008). "Effects of sample size on the performance of species distribution models." Diversity and Distributions**14**(5): 763-773.

Chapter 3: Expanding menace or invasional has-been? The potential distribution of *Senecio glastifolius* in New Zealand

Abstract

Senecio glastifolius is an invasive species in New Zealand where it threatens rare and vulnerable coastal floristic communities. It has expanded its range dramatically in recent years. Its future spread and impacts are largely uncertain. To aid its management in New Zealand, a predictive model of its potential distribution was created based on its bioclimatic niche. A new, expanded data set of its global range was matched to climate data using MaxEnt. Challenges associated with predictive modelling of invading species are discussed. Recommendations for management are given, based on significant areas of suitable habitat identified by the model which have yet to be invaded. Specifically, control is called for in coastal Taranaki to halt the current northward spread along the West Coast. This would be well complemented by active surveillance in the region.

3. 1. *Senecio glastifolius* in New Zealand: history, ecology and threats

Senecio glastifolius (Asteraceae) is a perennial herb native to South Africa, and invasive in New Zealand, Australia, the British Isles and Madeira. It was first recorded as naturalised in New Zealand in 1963 (CHR 143793). Interest was mostly limited to a few botanists, until its range expanded dramatically in the 1990s, attracting attention from the conservation community (Williams et al., 1999). New Zealand is considered a hotspot of global biodiversity (Myers et al., 2000), with such a unique biota that it has been described by evolutionary biologists as “as close as we will get to studying life on another planet” (Diamond, 1990: p3). New Zealand’s evolutionary isolation and recentness of colonisation by humans contribute to the intensity of on-going invasion impacts. Especially endangered in New Zealand, are coastal dune ecosystems (Jamieson, 2010), which have been heavily modified by forestry and agriculture and drastically reduced in their extent.

Senecio glastifolius owes part of its reputation as a potentially deleterious invader to its conspicuous presence at high-value conservation sites. It has reached high densities, for example, at Whitiua Scientific Reserve near Whanganui and is the subject of on-going weed control there (Colin Ogle, pers. comm.). Whitiua is home to several threatened native plant taxa including *Sebaea ovata*, *Selliera rotundifolia*, *Mazus novaezeelandiae*, *Isolepis basilaris*, and local forms of *Libertia peregrinans* and *Pimelia arenaria* (Williams et al., 1999 p11). Whitiua is one of only two New Zealand sites known to contain *Sebaea ovata*, a low-growing herb (Champion et al., 2003). The other site, at the Waitotara River mouth also contains *S. glastifolius*. *Pimelea actea* (formerly *Pimelea "turakina"*), a critically endangered endemic coastal herb endemic to the Manawatu dunelands on the North Island's West Coast also has its entire range encompassed by that of *S. glastifolius*. Further details of *S. glastifolius*'s association with native plants, and possible negative interactions, are given by Williams et al. (1999).

Senecio glastifolius is capable of forming dense stands, often of considerable area. This can lead not only to ecosystem modification, but also to the total visual transformation of landscapes, with entire hillsides turning purple during flowering in Spring (Williams et al., 1999; pers. obs.; B Brockie, pers. comm.). This phenomenon is common in the years after disturbance such as fire or harvesting of plantation forest, but also occurs regularly in habitats such as stabilised sand dunes, cliff faces and roadside embankments.

In New Zealand, *S. glastifolius* inhabits a range of habitats, chiefly: coastal areas of open or low-growing vegetation; disturbed or ruderal sites; roadsides; rocky banks; open understory of plantation forest; riparian sediment; and a wide range of agricultural settings where mowing/grazing is low (pers. obs.). Of this list, *S. glastifolius*'s invasiveness in coastal habitats is of the greatest concern. The integrity and abundance of New Zealand coastal dune floras has been heavily impacted by human development, and species introductions both accidental and deliberate (Jamieson, 2010). Very little of New Zealand's original coastal ecosystems remain intact, so preserving what is left is a high conservation priority.

Senecio glastifolius has also attracted the ire of farmers in New Zealand (Maslin, 2009), although opinions vary as to its seriousness as an agricultural pest. It is seldom found growing in intensive agricultural pasture, but does grow where

stocking rates are low or where other weeds such as gorse are present and restrict grazing. It seems to be reduced in its palatability to stock once the stem becomes woody.

Senecio glastifolius has many common names including holly-leaved Senecio, woad-leaved ragwort (chiefly UK), pink ragwort (esp. NZ), purple ragwort (less common), large Senecio (chiefly South Africa), waterdissel (Afrikaans), water thistle (South Africa, less common) and holly-leaved groundsel (UK, less common). *Senecio glastifolius*, commonly grows to around 1m tall in New Zealand, but has been reported up to 2.5m tall in Australia (CRC, 2003). It is a short lived perennial, typically living for 1-4 years (Williams et al., 1999).

3.1.1 Ecological requirements

The native range of *S. glastifolius* is very small, spanning the part of South Africa's Cape Region between Mossel Bay and Port Elizabeth (~300km). Its range comprises the seaward-most mountain ranges, extending to the coast. It is not recorded as growing farther inland than the seaward ranges (see chapter 2). Within this narrow range it is further restricted to damp sites, chiefly: wet, south-facing slopes (Jan Vlok, pers. comm.), around rivers, or in seeps (Bean & Johns, (year unknown)). Hence, it appears that within South Africa, water availability is a strong limiting factor in its distribution. Its association with wetter sites gives it its name 'waterdissel' or 'water thistle' (Williams et al., 1999). *Senecio glastifolius*'s association with damp areas is also observed in Australia (CRC, 2003). The climate in its narrow range is temperate, with more equable temperature than surrounding areas.

Its restriction to the poleward, maritime terminus of a predominantly dry, hot and climatically inequable region invites speculation that its fundamental ecological niche might not be fully available in its native range (see Appendix 3.5 for analysis of niche availability).

Despite being found at altitudes of up to 1,090m in South Africa (Germishuizen & Meyer 2003), the area is not cold (see discussion below, section 3.3.4). In New Zealand it grows at much lower altitudes than in South Africa. The highest it has been reliably recorded is 290m (± 10 m) in Wellington (pers. obs.).

Senecio glastifolius is an early successional species in its native range, where it is most frequent on disturbed sites, especially burns. It appears within a year following disturbance and persists for about 2-3 years (Jan Vlok; Dave Edge, pers. comm.). Its

habitat in Australia includes *Banksia* woodland and coastal shrubland (Hussey et al., 2007).

3.1.2 Legal Status

In Australia *S. glastifolius* is listed as a National Environmental Alert weed (CRC, 2003). In New Zealand it is banned for sale and propagation in Taranaki and Northland, and is managed as a site-led pest in Wanganui-Manawatu.

3.1.3 Control effort

Senecio glastifolius has been the subject of considerable control efforts in parts of Australia and New Zealand. Control is mostly in localised sites, where local biodiversity values are threatened by high densities. Examples include Whitiau Scientific Reserve; Manawatu coast (Jill Rapson, pers. comm.); Tawhirihoe (DoC, 2011); Mana Island, (Colin Ogle, pers. comm.); Kapiti Island (Diane Batchelor pers. comm.). Control is also undertaken where near or total local eradication is thought to be possible, such as Te Mata Peak, New Zealand (TMPTP, 2011); Motueka Estuary, New Zealand (Kim Wright, pers. comm.); and Albany, Western Australia. In addition to countless hours of volunteer labour, significant financial capital has been directed to the control or eradication of *S. glastifolius* (e.g. AU\$30,000 to Green Skills Inc. for the *Senecio glastifolius* eradication project (DEC, 2011)). The eradication programme led by Greenskills Albany has been in place since 2005. Success appears to be good so far (Diane Evers, pers. comm.).

3.1.4 Benefit of modelling

Predictions of invasive species' potential distributions are of value to environmental managers (Higgins et al., 1999; Peterson, 2003; Jimenez-Valverde et al., 2011; Mgidi et al., 2007; Venette et al., 2010). For example, predictive maps offer information on the likely spread of invasive species under natural dispersal. Continuous areas of suitability can be expected to act as dispersal corridors, while isolated areas of high suitability are less likely to be colonised without human mediation. Predictions can also help identify areas where monitoring for new incursions should be focussed, or where control is most beneficial in preventing further spread.

Currently there is little information on the potential distribution of *S. glastifolius* in New Zealand. Predictions do exist; Thuiller et al. (2005) presented a coarse scale (quarter-degree grid resolution, $\sim 25 \times 25 \text{ km}^2$) global prediction, and other unpublished predictions have been made (Scott et al., 2008).

This study presents an updated prediction, at a much higher resolution, and benefitting from a much improved species location data set (presented in chapter 2). Since previous studies, *S. glastifolius* has spread significantly in New Zealand (see chapter 2), including further inland than previously recorded, presumably into climatic space less represented in previous data sets.

3.2 Aims

This study aims to predict the potential distribution of *S. glastifolius* in New Zealand, as governed by climatic requirements. Specifically, the prediction aims to identify currently uninvaded areas which are suitable to habitation by *S. glastifolius*.

3.3 Methods

The potential distribution of *S. glastifolius* in New Zealand was predicted using SDM methods deemed appropriate for a range-expanding species, and within the constraints of the available data. Details are presented in the following sections, outlining the species presence data used; environmental predictor variables considered, plus details of their selection and transformation; choice of geographic extent for model training; algorithms used; and methods for evaluating the validity of predictions.

3.3.1 Species presence data

Models were trained using the global occurrence data set for *S. glastifolius* described in chapter 2. This data set incorporates 214 presence records from New Zealand, Australia and the Isles of Scilly, generalised to 30 arc second resolution. No published ecophysiological data exist for *S. glastifolius*, although some unpublished data do exist (John Scott, pers. comm.). For New Zealand, one record in the described data set (Kaupokonui, Taranaki) was not included in modelling because it was located shortly prior to the completion of this research study, after predictive models had been finalised.

3.3.2 Environmental data

Environmental predictor data layers were chosen from a candidate set of the 19 'Bioclim' data layers from Worldclim (Hijmans et al., 2005) at 30 arc second resolution, plus PET (potential evapotranspiration) and aridity, available from CGIAR (Trabucco & Zomer, 2009). Trabucco & Zomer's (2009) Aridity Index is a function of mean annual precipitation divided by mean annual potential evapotranspiration; values increase for more humid conditions and decrease for more arid conditions. Worldclim's 'Bioclim' variables are derived bioclimatic indices based variously on monthly values of temperature and precipitation.

PET data was sourced as monthly averages, plus an annual average. Monthly PET data was generalised to four annual quarters (e.g. December to February) roughly corresponding with the four seasons in the regions from which species data were drawn. These data layers were arranged by season rather than by months to account for the fact that the Northern Hemisphere experiences its seasons in different months from the Southern Hemisphere.

Senecio glastifolius has a strong anecdotal association with 'poor' substrates, such as sand, and with certain vegetation covers (it is excluded by very dense vegetation of more than about 1-2m height). These variables were not included in modelling. These variables vary on a very fine scale over much of *S. glastifolius*'s range. The requisite data sets, covering the entire range that the model examines, and at a resolution fine enough to match the variation of the variable, were not available. These variables are also comparatively dynamic compared to climate, especially in the case of vegetation cover; *S. glastifolius* often exploits the sparse cover of disturbed sites before being replaced by gorse and later, native scrub.

3.3.3 Transformation of Environmental Data

Gridded environmental data for the minimum temperature of the coldest month were transformed from the Celsius scale to the Kelvin scale. Untransformed data spanned zero and into the negative number range, which can produce undesirable mathematical artefacts during model training. This is true for the creation of product or quadratic responses in MaxEnt (see Elith et al., 2011 for an explanation of MaxEnt

features), because of the inherent behaviour of zero and negative numbers when multiplied, or when raised to an exponent.

The Kelvin scale was chosen arbitrarily as a way to avoid values spanning zero.

Transformation to positive-only values using some other arbitrary constant would produce an essentially identical model, the only and quite meaningless difference being that the absolute figures for model coefficients would differ accordingly.

However the predictions of the models and the shapes of the response curves would be the same.

The Kelvin scale, while less familiar than Celsius, is a more logical scale for measuring temperature in quantitative ecological studies, when the properties of temperature are considered. It is true that the freezing of water is of general ecological relevance, but a ratio scale of measurement (one that has an absolute zero (see Stevens, 1946)) is a better logical match to temperature, which has an absolute zero. Temperature cannot have a negative quantity, unlike other ecologically relevant variables such as soil water deficit; species do not experience an absence or negative quantity of temperature.

The use of untransformed data not only gives different model coefficients, but changes the relative contribution of each variable to the model (see Appendix 3.1).

3.3.4 Choice of predictor variable and removal of correlated predictors

Predictor variables were chosen primarily for postulated ecological relevance to the species. Dry stress seems to be the most limiting factor in the species' native distribution, and was represented in modelling by Trabucco & Zomer's (2009) Aridity Index. Several alternative measure of dry stress were available in the Worldclim dataset (Hijmans et al., 2005), plus the monthly, quarterly and annual PET data. Hence, some screening for the best candidate predictor was necessary.

Environmental predictor variables were screened for correlation (see Appendix 3.3). Where two or more variables were highly correlated ($r > 0.85$), all were discarded save the one deemed to be the most ecologically relevant based on knowledge of the plant's ecology. The variables included in the final model were Aridity Index, and Annual Temperature Range.

Aridity Index, which incorporates both precipitation and PET, arguably provides a closer ecological representation of water availability to a plant than other available measures of dry stress such as precipitation of the driest month. It is desirable to use

predictor variables with more direct physiological influence on a species, rather than those which act as proxies (Austin, 2002). Aridity Index performed better in preliminary trials than alternative measures of dry stress [results not shown].

Annual Temperature Range was included as a measure of temperateness, and consistently gave the greatest contribution to MaxEnt models used, as indicated by drop in test gain when omitted (Phillips, 2012).

Acute cold stress (i.e. frost severity) seems to have a limiting effect on *S. glastifolius*'s distribution, especially in New Zealand. In the few places where it exists well inland it tends to be limited to river valleys, which can be expected to offer some thermal buffering. The variable most closely approximating this ecological stress from the candidate set is Worldclim's minimum temperature of the coldest month (MinTemp). Note though, that this is not the absolute minimum temperature recorded, but the average minimum daily air temperature for the coldest month. During regional cross validation using MaxEnt (see section 'evaluation' below) this variable performed poorly. The apparent reason is that it is only limiting to *S. glastifolius* over a certain range of its spectrum: between approximately 0°C and 3°C. It acts as a limiting factor below a threshold within this range. This sort of predictor is best modelled using a multiplicative model rather than additive. Hence the additive MaxEnt produced ecologically unrealistic responses to MinTemp when trained in regions where the values were outside the relevant limiting range (South Africa, Australia and Scilly study areas, all of which are too warm for cold stress to limit *S. glastifolius*'s distribution – see subsection below).

To include MinTemp in the final model without undesirable effects, it was included via a manually enforced threshold rule (binary multiplicative term). A MaxEnt model was created using only Aridity Index and Annual Temperature Range as predictors, and the resulting predictions processed to include MinTemp. Grid cells with a MinTemp value below the lowest training value of 2.3°C (Rabbit Island, Nelson, New Zealand) were assigned a suitability value of 0. Grid cells with MinTemp values equal to or greater than 2.3°C retained the predicted suitability assigned by the MaxEnt model. Note that the observed threshold of 2.3°C is an interpolated value for a 30 arc second grid cell, and should not be regarded as a precise physiological value.

Of the training regions other than New Zealand, the lowest values for MinTemp are found in South Africa. Despite the very high altitudes at which *S. glastifolius* grows in

South Africa (up to 1090m, Germishuizen & Meyer, 2003), the minimum value for MinTemp for any recorded occurrence of *S. glastifolius* is 3°C. Very little of *S. glastifolius*'s biotope in South Africa has low values for MinTemp (see Figure 3.1 below). If a minimum convex polygon were drawn around recorded presences of *S. glastifolius*, only four grid cells within that polygon would have a MinTemp value below 2.3°C.

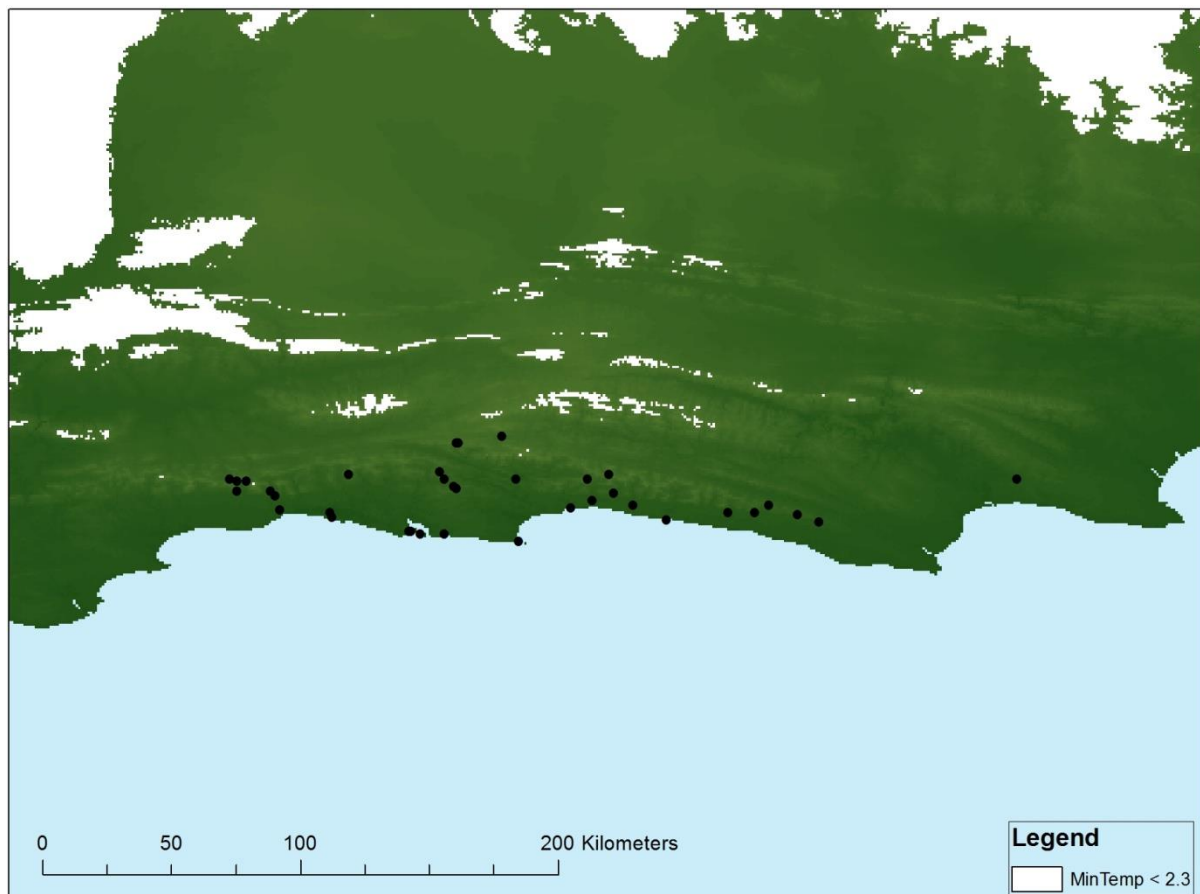


Figure 3.1. Areas of the Cape region with MinTemp values below the observed approximate threshold of 2.3°C. Kilometre scale is indicative only.

3.3.5 Background data

Species absence data is difficult to define for range-expanding invasive species, due to their limited dispersal opportunities in their invaded ranges. This problem is further aggravated for *S. glastifolius* which, as an often ephemeral, early-successional species in its native range, is often absent from climatically suitable habitat due to factors of vegetation and [lack of] disturbance.

In lieu of reliable species absence data, background data were used in model training (see Chapter 1, section 1.6.5). Background data were drawn from random

points within specified regions, with a correction made for latitudinal differences in the size of unprojected raster grid cells. Separate background regions were delineated for each of the four global regions from where *S. glastifolius* presence data were taken (South Africa, New Zealand, Australia and Scilly). In all cases, the concept of 'accessible area' was used to inform the delineation of the background (Elith et al., 2010; see also Barve et al., 2011).

The background region for New Zealand was created based on the four most significant regional populations for *S. glastifolius*: Whanganui, Wellington, Hawke's Bay and Poverty Bay. Not all recorded sites were included: Christchurch, Hamilton and Ohakune records were excluded because they consisted of only single observations of populations which appear to no longer exist. It is questionable that these constitute true naturalisation.

For each of the four populations, an initial site of invasion was identified based on the dates of the earliest records. A circular 'background' was created using this point as its centre, and the distance to the farthest record for that population as the radius. This assumes an equal opportunity of propagule spread in all directions such that the entire area has received sufficient propagule pressure that an absence of records can suggest unsuitable environmental conditions for growth. Mapped data were visually inspected in time-series to allow judgement of the validity of this approach in approximating spread of propagules.

Background delineation for South Africa, the species' native range, could not rely on documented spread to identify reachable areas. Instead, a buffer of ~50km was created around the recorded distribution to cover an area that balanced environmental contrast with reachability.

For Australia, a similar approach was taken, buffering the main area of infestation at Albany by a distance reaching the approximate location of the outlying record at Manjimup. No geographic background region was created for the sole occupied grid cell at Bundeena, NSW, but MaxEnt was specified to add to the background distribution any presence sample which has a combination of environmental variables not already present in the background. For the Isles of Scilly, the background region was simply the land area of the group of islands.

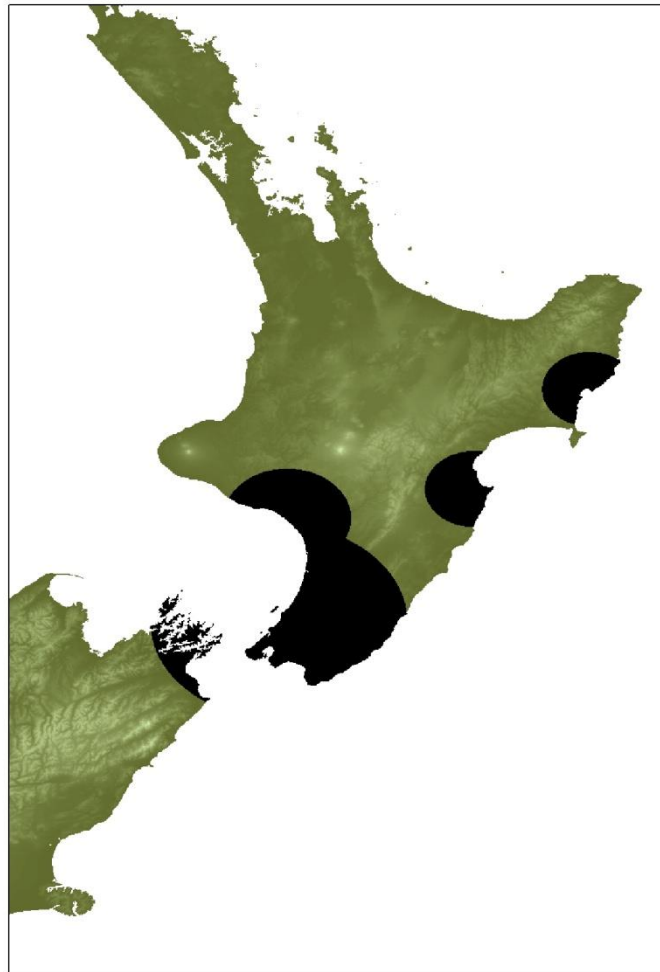


Figure 3.2 “Background” regions for New Zealand, shown in black.

3.3.6 Modelling methods

3.3.6.1 MaxEnt

Models were fitted using MaxEnt (version 3.3.3) (Phillips et al., 2006; Elith et al., 2011). MaxEnt is a machine learning algorithm which has been widely used in species distribution modelling, and compares favourably to alternative algorithms in a number of studies (Elith et al., 2006; Heikkinen et al., 2012).

MaxEnt’s default parameter settings were used, except for the following changes. To limit overfitting, regularisation was increased from the default values by a multiplier of 2 (Warren & Seifert 2011; Elith et al., 2011). Threshold features were excluded from model-fitting because they had led to overfitting in preliminary models (see Elith et al., 2011 for an explanation of MaxEnt’s feature classes). Overfitting was identified as

unrealistically detailed marginal response curves and stark fine-scale variation in mapped predicted suitability (results not shown).

To predict the potential distribution of *S. glastifolius* in New Zealand, a model was trained using presence records for all four study regions (see chapter 2 for description of the data set). Background data were drawn from the background regions described above at a ratio of 50 background points per presence. This ratio was held even across the separate regions.

MaxEnt models were created using predictor variables Aridity and Annual Temperature Range. Mapped predictions were further processed by manually including MinTemp as a binary multiplicative term, and transforming predictions by predicted suitability. Values below the threshold of minimum training presence were transformed to zero. Values equalling or exceeding this threshold were not adjusted. This is similar to the binary transformation often used in SDM studies (Liu et al., 2005), but retains the predicted relative suitability of suitable sites.

3.3.6.2 Bioclim

The argument has been made that modelling techniques placing zero or low emphasis on species absence or background (e.g. BIOCLIM (Busby, 1991)) may be the most suitable for predicting the *potential* distributions of species (Jimenez-Valverde et al., 2011). This is despite the fact models using absence, pseudo-absence or background, such as MaxEnt, are generally better for predicting species *actual* distributions. The thesis is that observed absence, including absence implied by gaps in the background region, is often due to contingent factors (e.g. biotic interactions, dispersal limitation). Thus, areas of environmental space which are within the study-species' fundamental niche may be wrongly penalised. Following this argument, rectilinear envelope models were fitted for comparison to MaxEnt models, using the R (R Development Core Team, 2012) package 'dismo's (Hijmans et al., 2012) implementation of the BIOCLIM algorithm. BIOCLIM fits rectilinear envelopes to training data in environmental space. Each environmental axis is considered separately, and the presence data arranged into quantiles.

Environmental space outside of the bounding limits for any variable (i.e. the maximum and minimum observed values for species presence) is assigned a value of zero. For variables which are ecologically limiting in only one direction (e.g. cold stress), bounding limits may be removed from one tail of the distribution. This

increases the ecological realism of the algorithm considerably, allowing the modeller to supervise limiting factors. Space within the observed range (inside the rectilinear envelope) is assigned a suitability score based on its value relative to the ranked presence data.

3.3.7 Transformation of suitability score

Model output is a continuous variable from 0 to 1, which can be interpreted as correlating with climatic suitability (chapter 1, section 1.8). To distinguish suitable geographic space from unsuitable, output was transformed via a threshold equal to the minimum training presence. Minimum training presence indicates the lowest value of predicted climatic suitability at which the species can demonstrably occur. Grid cells within the predicted regions which received a predicted suitability less than this threshold had their scores transformed to zero. Grid cells scoring equal to or higher than the threshold retained their original predicted suitability values.

3.3.8 Model evaluation

Models were evaluated using 4-fold regional cross-validation, treating each geographically distinct study region as a single fold (New Zealand, South Africa, Australia and Scilly). This yielded folds of greatly varying number of presences (New Zealand: $n = 134$ presence records; South Africa: $n = 36$; Australia: $n = 40$, Scilly: $n = 4$). Regional cross-validation ensures that test data are independent of training data, and demonstrates a model's ability to predict to new geographic regions (Hartley et al., 2006; Jimenez-Valverde et al., 2011).

Omission on test data was calculated, using minimum training presence as the threshold. Binomial tests of model significance were performed, with the null hypothesis that test presence points were predicted no better than by a random prediction with the same fractional predicted area. Again, the minimum training presence was used as the threshold. AUC is also reported.

One limitation of the regional cross-validation used here is that each region contributes different information on the species' niche, because of their different climates (see Figure A3.5.2, Appendix A3.5). None of the models trained on $k-1$ regional subsets has access to as much information as the full, global model.

Therefore, evaluation scores from cross-validation are likely to under-represent the performance of the full model. To complement regional cross-validation, the globally

fitted model was evaluated using ordinary *k*-fold cross-validation, splitting data randomly into five folds of equal size.

3.3.9 Madeira

Madeira was not included in model training or testing because no georeferenced records were available. The final model, described above, was projected onto Madeira as an additional, informal test of model transferability. Results are given in Appendix A3.4.

3.4 Results

The model predicted all currently occupied areas as suitable, using minimum training presence as the threshold for suitability. Additionally, for all three invaded regions projected to, as yet uninvaded areas were predicted to be suitable, suggesting continued spread of *S. glastifolius* is likely.

3.4.1 New Zealand

All of the currently invaded areas were predicted as suitable (Figure 3.3). Presently unoccupied areas were predicted as suitable, including most of the northern half of the North Island and associated offshore islands. The entire western coast of the North Island was predicted to be suitable, save for only a few very small gaps of four pixels or less between Taranaki and Auckland. Other large, unoccupied or sparsely populated areas of predicted suitability include parts of the Wairarapa and southern Hawke's Bay, plus inland Whanganui and Manawatu. In all of these latter regions, availability of suitable land cover is likely to limit, though not preclude, the expansion of *S. glastifolius*.

In the South Island, predicted climatic suitability is mostly limited to the coastal areas of the northern third of the island. Predicted suitability is high for the eastern coastal areas from the Marlborough Sounds south to Kaikoura and northern Canterbury. The Kaikoura Peninsula in particular is predicted as highly suitable. Isolated pockets of predicted suitability occur at Otago Peninsula, Banks Peninsula, Murchison, and the Karamea Bight. Despite the high predicted suitability of the Otago Peninsula, both this area and Banks Peninsula have only marginally suitable winter temperatures and so might be less habitable than mapped predictions suggest.

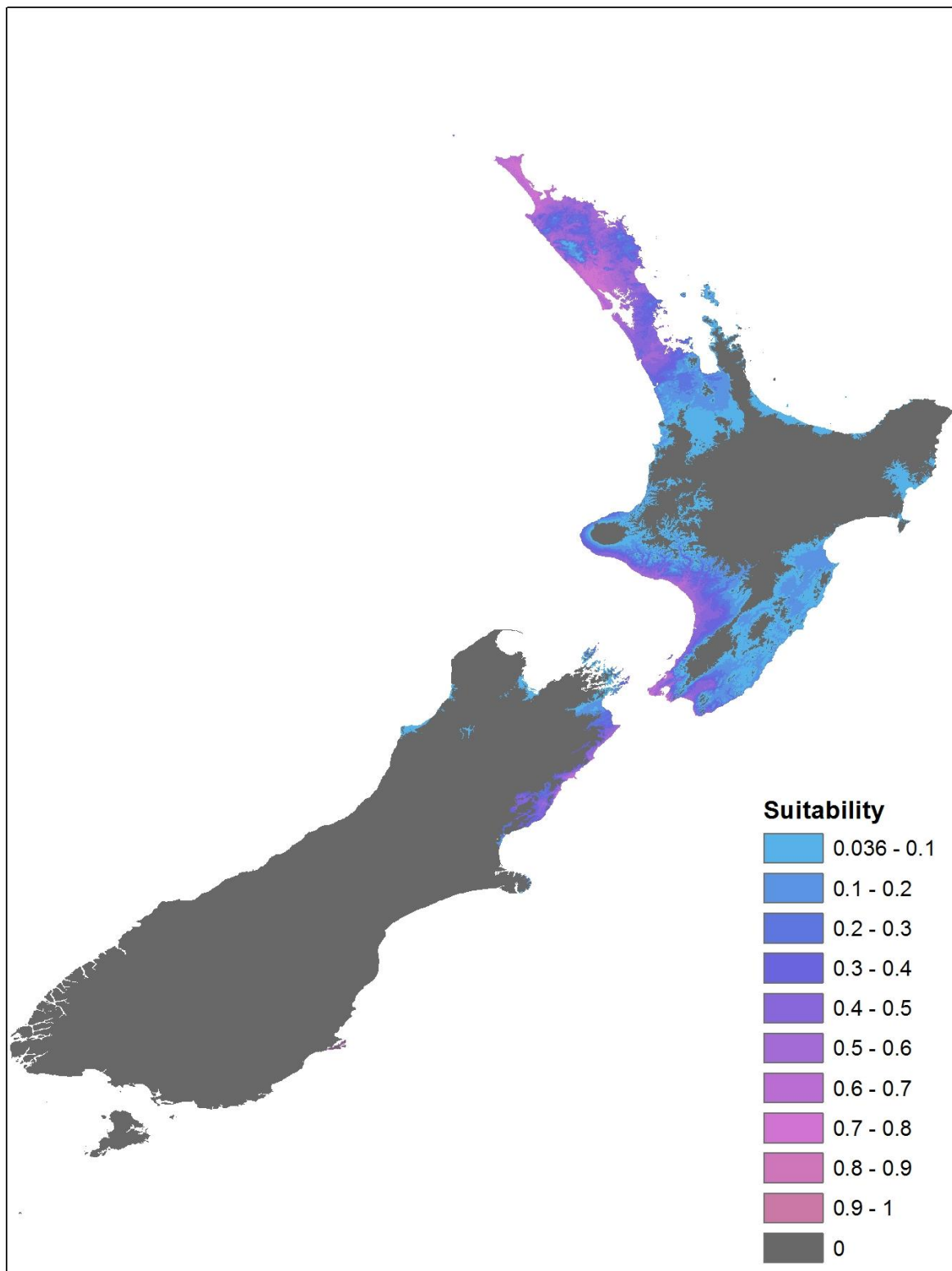


Figure 3.3. Predicted climatic suitability for *Senecio glastifolius* in New Zealand, based on the MaxEnt model trained on the global data set, with transformation to include the influence of MinTemp. Predicted suitability values are transformed by the minimum training presence threshold to distinguish suitable from unsuitable areas.

3.4.2 Australia

Mapped predictions for Western Australia are congruent with existing opinion of where the species is likely to invade, with a localised region of high predicted suitability from Busselton to Albany (Williams et al., 1999). Predictions also show high suitability for much of the mainland coast and Tasmania. Predicted suitability for northern regions is likely to be spurious, owing to environmental variables excluded from the model (chiefly, measures of heat stress). These variables were extraneous to model predictions for New Zealand, the purpose of this study.

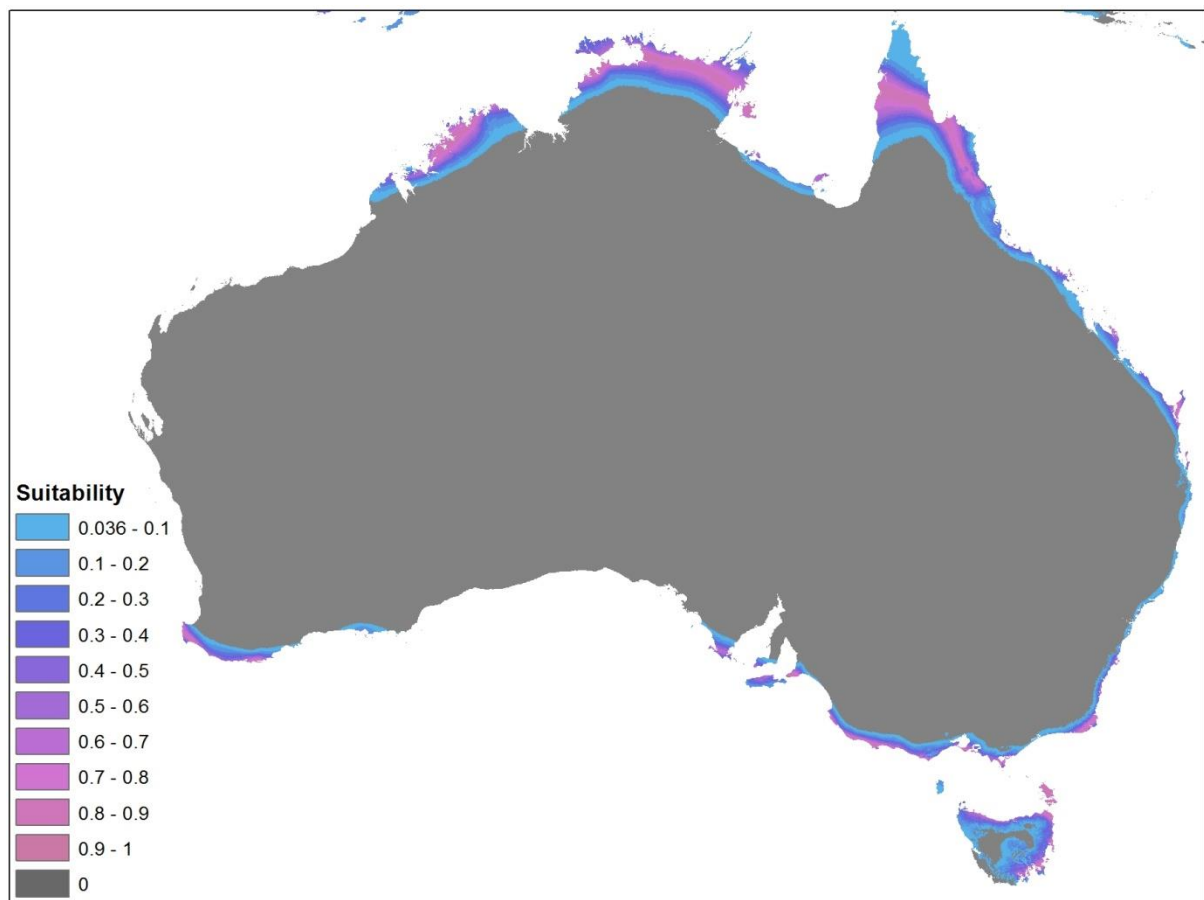


Figure 3.4. Predicted climatic suitability for *S. glastifolius* in Australia.

3.4.3 Scilly

Predicted climatic suitability was fairly uniform across the Isles of Scilly, showing all of the represented land area to be suitable.

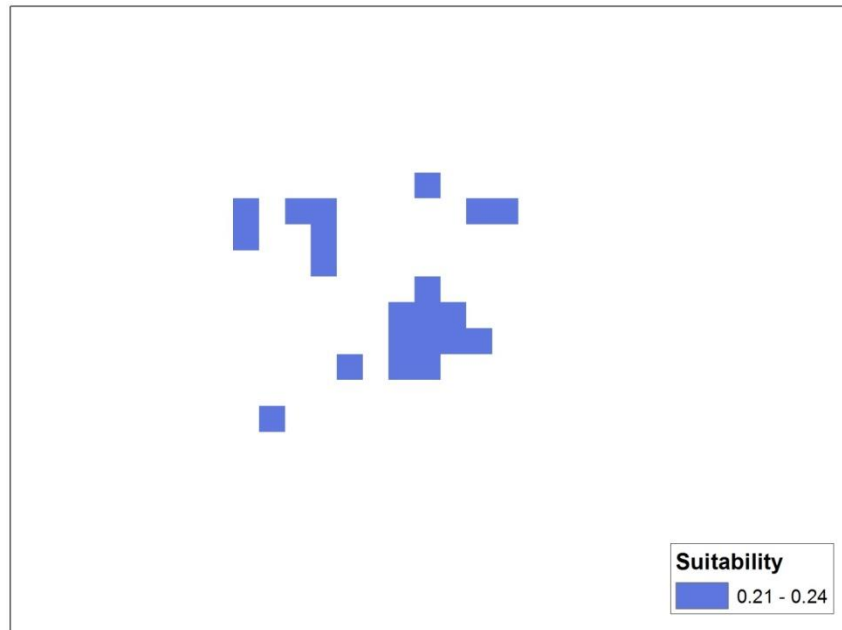


Figure 3.5 Predicted climatic suitability for *Senecio glastifolius* in the Isles of Scilly. Note that the colour is scaled by the same intervals as for the other predicted regions (Figures 3.2 – 3.4).

3.4.4 South Africa

The model predicts the whole of the occupied range in South Africa to be suitable, as well as fairly substantial areas in a coastal band to the east and west. The lack of recorded presence in these areas might be due to factors external to the model, including biotic interactions or missing climatic variables.

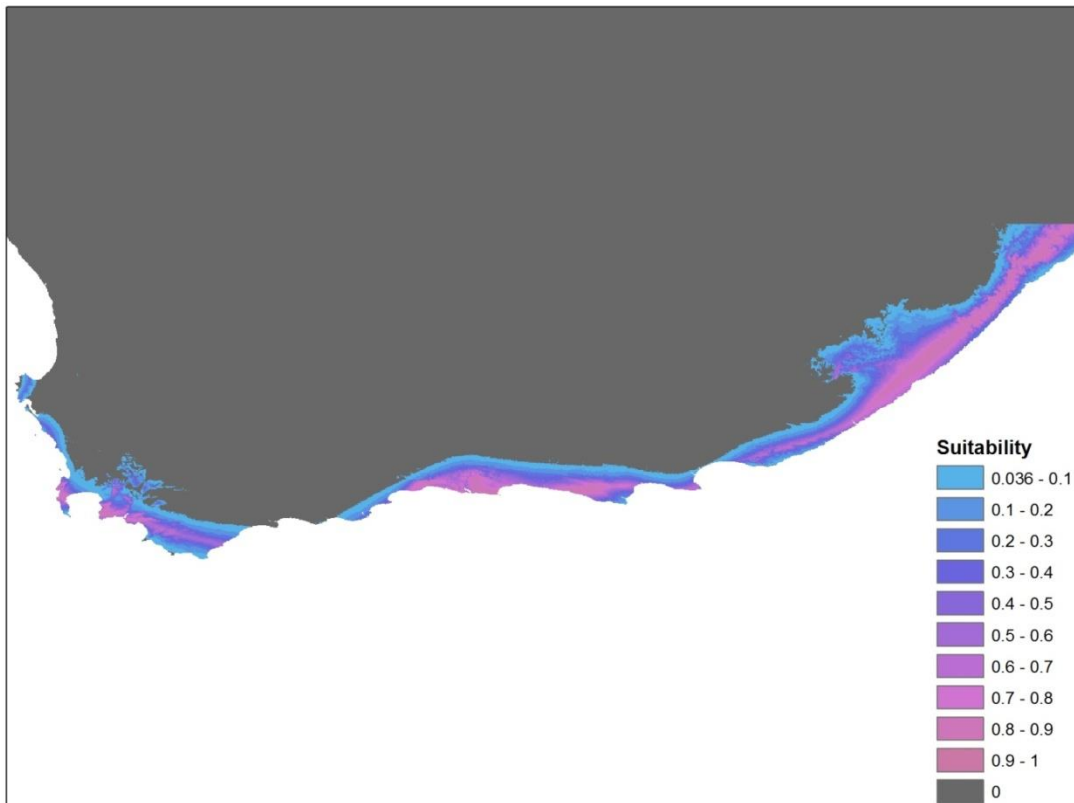
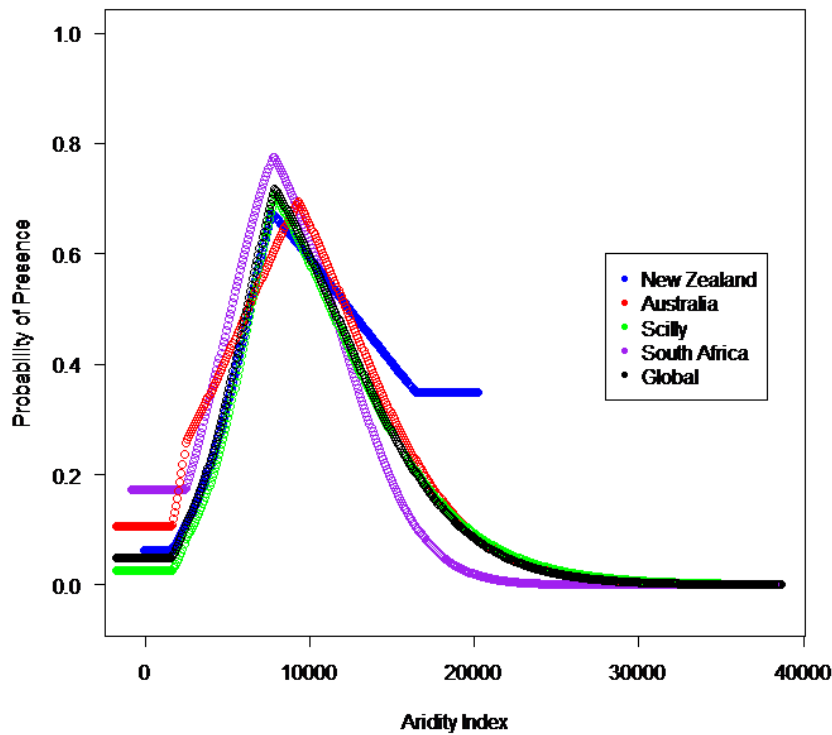


Figure 3.6 Predicted climatic suitability for *S. glastifolius* in South Africa's Cape region.

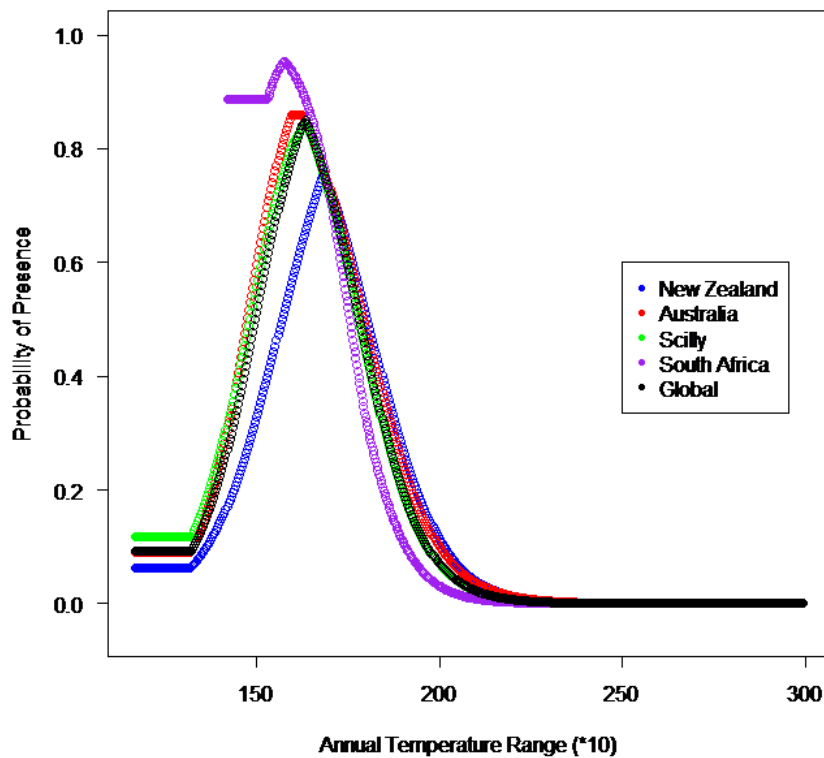
3.4.5 Species response to environmental variables

Response curves show relatively symmetric, unimodal responses to both predictor variables included in the MaxEnt model. Shapes of curves and positions of optima were fairly robust under regional cross-validation. The modelled responses are in agreement with prior ecological knowledge that *S. glastifolius* has a preference for reasonably temperate climates (Annual Temperature Range optimum at 16.3°C), and average water needs (Aridity Index optimum at ~0.8 [unit-less variable]), and does not grow in highly arid areas.

Note that Annual Temperature Range is not the difference between the absolute extreme temperatures, but for monthly averages of daily maximum and minimum temperatures for the hottest and coldest months respectively.



A



B

Fig 3.7. Response curves for *S. glastifolius* to aridity (A) and annual temperature range (B), from regional cross validation. The global model is shown in black; coloured curves are trained on $k-1$ regions. Colour indicates the test region for each model (e.g. Green curve tested on Scilly, trained on New Zealand, Australia and

South Africa). Flat sections of response curves are a result of MaxEnt's 'clamping' function. Note that the scale for Aridity Index is multiplied by 10,000 on the plot.

3.4.6 Results of Cross-validation

All test regions were fairly well predicted during 4-fold regional cross-validation. Test AUC ranged from medium (Scilly: 0.835) to good (Australia: 0.963), noting that AUC scores require a different interpretation for range-expanding species.

Table3.1. Summary of results from 4-fold regional cross-validation. Omission is calculated using minimum training presence.

Test region (withheld from training):	New Zealand	South Africa	Australia	Scilly
Omission	0	0	0	0
Test AUC	0.879	0.879	0.963	0.835
Min Training Presence	0.032	0.015	0.054	0.033
Binomial P value	<0.001	<0.001	<0.001	0.158

Omission was zero for all regional models, associated with extremely low minimum training presence scores in all cases. This lack of omission shows good ability of the model to predict to new geographic regions.

Binomial tests were significant for all regions except the Isles of Scilly. Considering the climatic geography of this tiny group of islands, this is neither surprising nor is it indicative of poor model performance. The climate varies very little across the islands, and so, accordingly, did predicted suitability (range 0.21 – 0.24); the entire land area of the islands was predicted as suitable. Hence the null hypothesis that test points were predicted no better than by a random prediction with the same fractional predicted area was impossible to reject; a random prediction with the same fractional area (100%) is necessarily identical to the modelled prediction.

3.4.7 Ordinary cross-validation of globally fitted model

The model performed equally well under ordinary, non-regional cross-validation. Average omission was less than 1%, AUC was medium to good and binomial tests were consistently significant at the 0.1% significance level.

Table 3.2. Summary of results from 5-fold cross-validation. Omission is calculated using minimum training presence. Values are rounded to three significant figures.

Repetition:	Average	1	2	3	4	5
Test	0.009	0	0	0.047	0	0
Omission						
MinTrainPres	0.036	0.034	0.036	0.043	0.027	0.036
Binomial P value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
AUC	0.885 (s.d. 0.017)	0.872	0.896	0.884	0.862	0.911

3.5 Discussion

The most salient information from the mapped prediction for New Zealand is that significant areas of potentially highly suitable habitat remain yet to be invaded; *S. glastifolius* should not be regarded as an ‘invasional has-been’.

Some of the suitable, but not yet invaded areas are unlikely to be invaded by *S. glastifolius* through natural dispersal. A clear example is Otago Peninsula. In contrast, mapped predictions suggest almost continuous suitable habitat from existing populations in Taranaki to the North of New Zealand. Gaps are smaller than observed dispersal distances. Without management intervention it seems likely that *S. glastifolius* will spread north from Taranaki into the King Country, Waikato, Auckland and Northland.

In the opinion of the author, *S. glastifolius* requires control and surveillance in Taranaki. The North Island’s West Coast experiences a prevailing wind from the Westerly quarter, which has probably slowed the expansion of *S. glastifolius* into low-lying Taranaki. If allowed to reach the northern coast of Taranaki it can be

expected that *S. glastifolius* will spread north-east comparatively rapidly, expanding its range significantly.

3.5.1 Caveats

The predicted potential distribution of *S. glastifolius* in New Zealand presented here represents a simplified model of the full range of factors controlling the species' distribution. Factors not included in the model will exert a strong influence on the actual distribution of the species. Notably, there are large areas where the predominant land cover would prohibit high densities. In these areas, which include areas of dense vegetation or intensive agriculture, *S. glastifolius* will be limited to refugia such as roadsides or banks. Biotic limitations will also constrain the distribution. Personal observation suggests that in areas where feral goats are common, such as inland from Whanganui, densities of *S. glastifolius* are lower than might be expected climatically, but this link has not been tested. No formal information on vertebrate herbivory of *S. glastifolius* was available, though studies are currently being conducted on the plant's alkaloid chemistry (Mikey Wilcox, pers. comm.).

3.5.2 Sample bias

Training data used in this study are likely to suffer from geographic biases, which could translate into bias in environmental space (Kadmon et al., 2004). Both MaxEnt and BIOCLIM are vulnerable to environmentally biased data. In the case of BIOCLIM, provided that the extremes of a species' niche are sufficiently sampled the model should still provide an accurate description of the limits to a species' distribution. Bias will simply distort the relative suitability assigned to grid cells.

3.5.3 A note on minimum training presence

The minimum training presence from the full model was used as a threshold to discriminate between suitable and unsuitable sites, yielding the predicted potential distribution for the species. This approach is susceptible to the possible inclusion of false-presences, especially records of plants growing outside of their niche (i.e. unable to maintain self-perpetuating populations). In the case of *S. glastifolius*, the seeds of which are wind-dispersed, it is likely that some recorded presences represent plants growing outside of their fundamental niche. Without sufficient

propagule supply from nearby source patches (in source-sink dynamics terminology, e.g. Hanski, 1999), sites associated with suitability equal to minimum training presence will not support the species long-term. Hence, the model predictions will be commissive.

In this study there is evidence that predicted suitability values very close to the minimum training presence fall within the fundamental niche of *S. glastifolius*. From the full data set, the record assigned the third lowest suitability (0.0576) is for a patch of plants known to have produced viable seed. The record is for a patch of plants growing on a causeway embankment near Motueka (CHR 220819). Not only its self-perpetuation implied by the existence of a highly localised group of many plants existing over several decades, but Williams et al. (1999) experimentally grew seed from this population. The experimental germination rate was 53% (Williams et al., 1999, p12). Additionally, it is reported that a member of the public collected seed from this population and sowed them successfully on nearby Jacket Island (Kim Wright, *pers. comm.*).

Regardless of demonstrated viability at sites of low predicted suitability, the minimum training presence provides an estimate of the predicted suitability at which a species can demonstrably exist. Even if the species cannot sustain its population at associated sites, invasion impacts (i.e. competition with native species) can still be anticipated where there is sufficient propagule input from source populations.

When interpreting minimum training presence and associated ideas, it must be considered that predicted values are just that – predictions – and not all sites with the same predicted value will have the same real-world suitability to the species.

Not too low a value for minimum training presence has been suggested as a useful check for models (Jimenez-Valverde et al., 2011). The spread of training data along each independent axis influences this value; if some presence points are outliers on any one axis then the algorithms used (especially BIOCLIM) will tend to assign them a low predicted suitability value, resulting in a low minimum training presence for the model. One possible cause of outliers is that the intervening environmental space between the outliers and the bulk of the data is scarcely available in reachable geographic space. This could be tested quantitatively.

3.5.4 Background delineation

Delineation of background regions for SDM is a poorly developed area of the literature, and a topic of active research (Barve et al., 2011). Choice of background is known to have a large influence on model predictions (Webber et al., 2011).

Approaches have been advanced to improve choice of background, but no definitive guidelines exist. Backgrounds used in this study were chosen to incorporate an appropriate amount of environmental contrast between presence and unoccupied background, while attempting to minimise inclusion of unreachable geographic space. The actual extent of geographic space sampled by the plant – i.e. that space which has received appreciable propagule pressure from the species – is unknown, and probably unknowable.

It is acknowledged that the backgrounds used in this study carry an arbitrary element. During model development, a number of alternative backgrounds were trialled ranging from more spatially restricted to more expansive. Results were not deemed to be grossly altered (results not presented).

Another issue with simply drawing a radius equal to maximum observed distance of spread from original invasion sites is that it does not take into account that sites toward the invasion front are likely to be less ‘in-filled’ than sites closer to the original source. One possible approach to reduce this issue is to weight absences closer to the source higher than those close to the invasion front.

3.5.5 A comment on the extrapolation involved in predictions

Projections of climatic suitability to the whole of New Zealand require extrapolation from the training data. This is visible in the climatic space scatterplots of Appendix 3.5, and in MESS maps produced by MaxEnt (not presented here). Inspection of the response curves (Figure 3.7 above) show that they are closed within the training range of the data. This is supported by ecological plausibility; *S. glastifolius* is likely to have a truly unimodal response to aridity and temperateness. Similarly, the effect of low temperatures (MinTemp) is extremely unlikely to have unexpected effects beyond the span of training data. Therefore, the extrapolation present in our predictions for New Zealand can be accepted as being ecologically reasonable, and not contributing much uncertainty to predictions.

However, the measures of extrapolation used examine only the variables present in the model. This is standard (e.g. MESS). Predictions to new areas involve

uncertainty not only due to the model's extrapolation of fitted functions but the unknown nature of which environmental variables will limit the species' distribution in the novel areas. This idea is illustrated well by the failure of models trained on *S. glastifolius*'s overseas range (South Africa, Scilly, Australia) to identify MinTemp as a limiting variable of *S. glastifolius*'s distribution. The anticipation of limiting factors acting in projection ranges, but not training ranges relies on ecological knowledge of the modeller.

3.6 Conclusion

Senecio glastifolius is likely to continue its expansion in New Zealand if left unchecked. Its distribution appears to be strongly influenced by climate, especially water availability, temperateness and cold winter temperatures. Predicted climatic requirements are met for *S. glastifolius* in significant areas of New Zealand that remain, for now, free of this species. If human-mediated spread can be avoided, and control and surveillance are carried out in just a few key areas, then the invasive impacts of this species can be minimised.

3.7 References

- Austin, M. P. (2002). "Spatial prediction of species distribution: an interface between ecological theory and statistical modelling." Ecological Modelling**157**(2-3): 101-118.
- Barve, N., V. Barve, et al. (2011). "The crucial role of the accessible area in ecological niche modeling and species distribution modeling." Ecological Modelling**222**(11): 1810-1819.
- Bean, A. & Johns, A. (Year Unknown). *Stellenbosch to Hermanus, South African Wild Flower guide No. 5*. Publication of the Botanical Society of South Africa.
- Busby, J. R. (1991). BIOCLIM - A Bioclimatic Analysis and Prediction System. Nature Conservation: Cost Effective Biological Surveys and Data Analysis. C. R. Margules and M. P. Austin. Canberra, CSIRO: 64-68.
- Champion, P. D., D. E. Hofstra, et al. (2003). "Growth and habitat of *Sebaea ovata* (*Gentianaceae*) in New Zealand and Australia." SCIENCE FOR CONSERVATION**229**: 32.
- CRC Cooperative Research Centre for Australian Weed Management. 2003. "Weed Management Guide: Holly leaved Senecio", available from:

<http://www.dpi.nsw.gov.au/agriculture/pests-weeds/weeds/profiles/holly-leaved-senecio>, retrieved 23/6/2011.

DEC Department of Environment and Conservation, 2011. "2011 ENVIRONMENTAL COMMUNITY GRANT RECIPIENTS", available from <http://www.dec.wa.gov.au/content/view/5135/2274/>

Diamond JM 1990. New Zealand as an archipelago: an international perspective. In: Towns DR, Daugherty CH, Atkinson IAE eds. Ecological restoration of New Zealand islands. Conservation Science Publication No. 2, Department of Conservation, Wellington, New Zealand, pp. 3–8.

[DoC] Department of Conservation. (2011). "Volunteer, join or start a project", <http://www.doc.govt.nz/getting-involved/volunteer-join-or-start-a-project/volunteer/volunteer-programme-by-region/manawatu-whanganui/manawatu/>, retrieved: 22/11/11.

Edge, D.A. 2005. Ecological factors influencing the survival of the Brenton Blue butterfly, *Orachrysops niobe* (Trimen) (Lepidoptera: Lycaenidae). Potchefstroom: North-West University. (Thesis - D.Phil.)

Elith, J., C. H. Graham, et al. (2006). "Novel methods improve prediction of species' distributions from occurrence data." *Ecography***29**(2): 129-151.

Elith, J., M. Kearney, et al. (2010). "The art of modelling range-shifting species. ." *Methods in Ecology & Evolution*.

Elith, J., S. J. Phillips, et al. (2011). "A statistical explanation of MaxEnt for ecologists." *Diversity and Distributions***17**(1): 43-57.

Germishuizen G, Meyer NL (2003) Plants of southern Africa: an annotated checklist. *Strelitzia*, 14, 1–1231. No 53.

Hanski, I. (1999). *Metapopulation Ecology*. Oxford, UK, Oxford University Press.

Hartley, S., R. Harris, et al. (2006). "Quantifying uncertainty in the potential distribution of an invasive species: climate and the Argentine ant." *Ecology Letters***9**(9): 1068-1079.

Heikkinen, R. K., M. Marmion, et al. (2012). "Does the interpolation accuracy of species distribution models come at the expense of transferability?" *Ecography***35**(3): 276-288.

Higgins, S. I., Richardson, D. M., Cowling, R. M. and Trinder-Smith, T. H. 1999. "Predicting the Landscape-Scale Distribution of Alien Plants and Their Threats to plant Diversity". *Conservation Ecology*. 13(2): 303-313.

Hijmans, R. J., S. E. Cameron, et al. (2005). "Very high resolution interpolated climate surfaces for global land areas." *International Journal of Climatology***25**(15): 1965-1978.

- Hijmans, R. J., S. Phillips, et al. (2012). *dismo: Species distribution modeling*. R package version 0.7-17. <http://CRAN.R-project.org/package=dismo>
- Hussey, B.M.J., Keighery, G.J., Dodd, J., Lloyd, S.G. & Cousens, R.D. (2007) *Western Weeds. A guide to the weeds of Western Australia*. 2nd Edition. The Plant Protection Society of Western Australia, Victoria Park
- Jamieson, S. L. 2010. *Sand dune restoration in New Zealand: Methods, Motives, and Monitoring*. Victoria University of Wellington. (Thesis – MSc).
- Jimenez-Valverde, A., A. T. Peterson, et al. (2011). "Use of niche models in invasive species risk assessments." *Biological Invasions***13**(12): 2785-2797.
- Kadmon, R., O. Farber, et al. (2004). "Effect of roadside bias on the accuracy of predictive maps produced by bioclimatic models." *Ecological Applications***14**(2): 401-413.
- Liu, C. R., P. M. Berry, et al. (2005). "Selecting thresholds of occurrence in the prediction of species distributions." *Ecography***28**(3): 385-393.
- Maslin, J., (2009). Farmer wild about spreading weed, *Wanganui Chronicle*, 5 Nov. p.3.
- Mgidi, T. N., D. C. Le Maitre, et al. (2007). "Alien plant invasions – incorporating emerging invaders in regional prioritization: a pragmatic approach for Southern Africa." *Journal of Environmental Management***84**: 173–187.
- Myers, N., R. A. Mittermeier, et al. (2000). "Biodiversity hotspots for conservation priorities." *Nature***403**(6772): 853-858.
- Peterson, A. T. (2003). "Predicting the geography of species' invasions via ecological niche modeling." *Quarterly Review of Biology***78**(4): 419-433.
- Phillips, S. J., R. P. Anderson, et al. (2006). "Maximum entropy modeling of species geographic distributions." *Ecological Modelling***190**(3-4): 231-259.
- Phillips, S.J. (2012) "A brief tutorial on Maxent". Available online: <http://www.cs.princeton.edu/~schapire/maxent/> (accessed 2012).
- 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/>.
- Rutherford, M. C., L. W. Powrie, et al. (2003). "ACKDAT: a digital spatial database of distributions of South African plant species and species assemblages." *South African Journal of Botany***69**(1): 99-104.

- Scott, J. K., K. L. Batchelor, et al. (2008). "Modelling Climate Change Impacts on Sleeper and Alert Weeds: Final Report." (CSIRO technical document).
- Stevens, S.S (1946). "On the Theory of Scales of Measurement". *Science* 103 (2684): 677–680. doi:[10.1126/science.103.2684.677](https://doi.org/10.1126/science.103.2684.677)
- Thuiller, W., D. M. Richardson et al., (2005). "Niche-based modelling as a tool for predicting the risk of alien plant invasions at a global scale". *Global Change Biology* **11**: 2234–2250. doi: 10.1111/j.1365-2486.2005.001018.x
- [TMPTP] Te Mata Trust Park (2011). "Friends of Te Mata Park", http://www.tematapark.co.nz/?page_id=295 retrieved 22/11/11.
- Trabucco, A., and Zomer, R.J. (2009) .*Global Aridity Index (Global-Aridity) and Global Potential Evapo-Transpiration (Global-PET) Geospatial Database*. CGIAR Consortium for Spatial Information. Published online, available from the CGIAR-CSI GeoPortal at: <http://www.csi.cgiar.org>.
- Venette, R. C., D. J. Kriticos, et al. (2010). "Pest Risk Maps for Invasive Alien Species: A Roadmap for Improvement." *Bioscience* **60**(5): 349-362.
- Warren, D. L. and S. N. Seifert (2011). "Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria." *Ecological Applications* **21**(2): 335-342.
- Webber, B. L., C. J. Yates, et al. (2011). "Modelling horses for novel climate courses: insights from projecting potential distributions of native and alien Australian acacias with correlative and mechanistic models." *Diversity and Distributions* **17**: 978-1000.
- Wells, M.J.; Balsinhas, A.A.; Joffe, H.; Engelbrecht, V.M.; Harding, G.; Stirton C.H. 1986: A catalogue of problem plants in southern Africa. Memoirs of the botanical survey of Southern Africa.
- Williams, P. A., C. C. Ogle, et al. (1999). "Biology and ecology of *Senecio glastifolius* and its spread and impacts in New Zealand." *Science for Conservation* **112**: 22.

Chapter 4: Discussion and conclusions

4.1 Summary of key findings

Senecio glastifolius is an invasive species in New Zealand that continues to spread rapidly, with significant potential to expand its range in New Zealand. This thesis presented a new description of its current distribution, demonstrating a much wider distribution than was previously available from any single source. *Senecio glastifolius* is spreading steadily into Taranaki, and appears likely to spread over vast areas if it is not halted by management intervention.

The distribution of *S. glastifolius* appears to be strongly controlled by the climatic variables of annual temperature range (temperateness), aridity, and winter cold stress. *Senecio glastifolius* shows a different response to these variables in the different geographic regions it inhabits. Most saliently, it inhabits much wetter conditions in New Zealand than in its native range. This is attributable to differences in the biotopes between the countries, with greater availability in New Zealand of the humid conditions preferred by *S. glastifolius*. The other important difference in response to climate between major geographic regions relates to winter temperatures. Extremes of winter temperature limit the distribution of *S. glastifolius* only in New Zealand. Again, this is due to differences in climates between regions; winter temperatures in the parts South Africa, Australia and Scilly inhabited by *S. glastifolius* are not low enough to limit persistence of the species.

The spread of *S. glastifolius* in New Zealand from sites of introduction to its current distribution show an interplay between climatic suitability and human-mediated dispersal. Populations which exist in geographically isolated areas of high climatic suitability have not undergone enormous expansion. This is exemplified at Poverty Bay where despite naturalisation in Gisborne as early as 1963, the species has not spread far. In contrast, a jump in dispersal to Whanganui in the late 1980s, presumed to be human-mediated, resulted in rapid and dramatic invasion of an expansive area. This was possible due to large, contiguous areas of climatic suitability surrounding Whanganui.

Contiguity of climatically suitable habitat extends from the current invasion front in coastal Taranaki, along the West coast of the North Island, to a large area including

Auckland and Northland. The hypothesised dispersal pathway along the coast in Taranaki is fairly narrow, making it a good candidate as an area for management control to prevent invasion. Field observation by the author revealed that *S. glastifolius* has already spread much further into Taranaki than was previously recorded, extending at least as far north as Kaupokonui.

4.2 Challenges to modelling

Making predictions of potential distribution for poorly studied and range-expanding species poses unique challenges to researchers, especially when predictions are needed for regions with novel climates. This study exemplified many of those challenges, necessitating careful methodological choice throughout the SDM process.

A major challenge to SDM of poorly studied species is adequate characterisation of their response to environment. Where ecophysiological data is unavailable, as was the case in this study, the species' niche is inferred from its distribution. Incomplete and imperfect data impose major limits to SDM. This study improved the data available for *S. glastifolius* through synthesising information from many and disparate sources, processing data, and creating new information through field observation. Limitations in the data remained, however, including probable spatial bias. Another challenge to SDM for invasive species is the difficulty of evaluating predictions, due to a lack of reliable information on species 'true' absence, and to ever present issues of spatial autocorrelation. These challenges can be addressed by careful choice of evaluation criteria, use of truly independent test data, and careful interpretation of results.

4.3 Future research

Based on research undertaken for this thesis, I recommend useful avenues of future research into two areas, relating first to *S. glastifolius*, and to the field of species distribution modelling.

4.3.1 Designed surveys of *S. glastifolius*

Data quality and completeness is a major limiting factor to studies into species' distributions (Lobo, 2008). Although this thesis made advances to the available data for *S. glastifolius*, limitations in the data remain. Designed field surveys of *S. glastifolius* could provide data free of spatial and environmental bias. Bias remains one of the most problematic aspects of available species data for SDM (Elith et al., 2011).

4.3.2 Better utilising information from imprecise species location data

The dataset of species presence records used for model training in this study, though larger than any known existing previously for *S. glastifolius*, excluded a large number of records. Many of these records were apparently accurate, but were deemed to spatially imprecise to use for SDM. A useful avenue of future research would be the development of methods to utilise imprecise records without the current problems they pose of introducing an un-quantified degree of uncertainty into predictions.

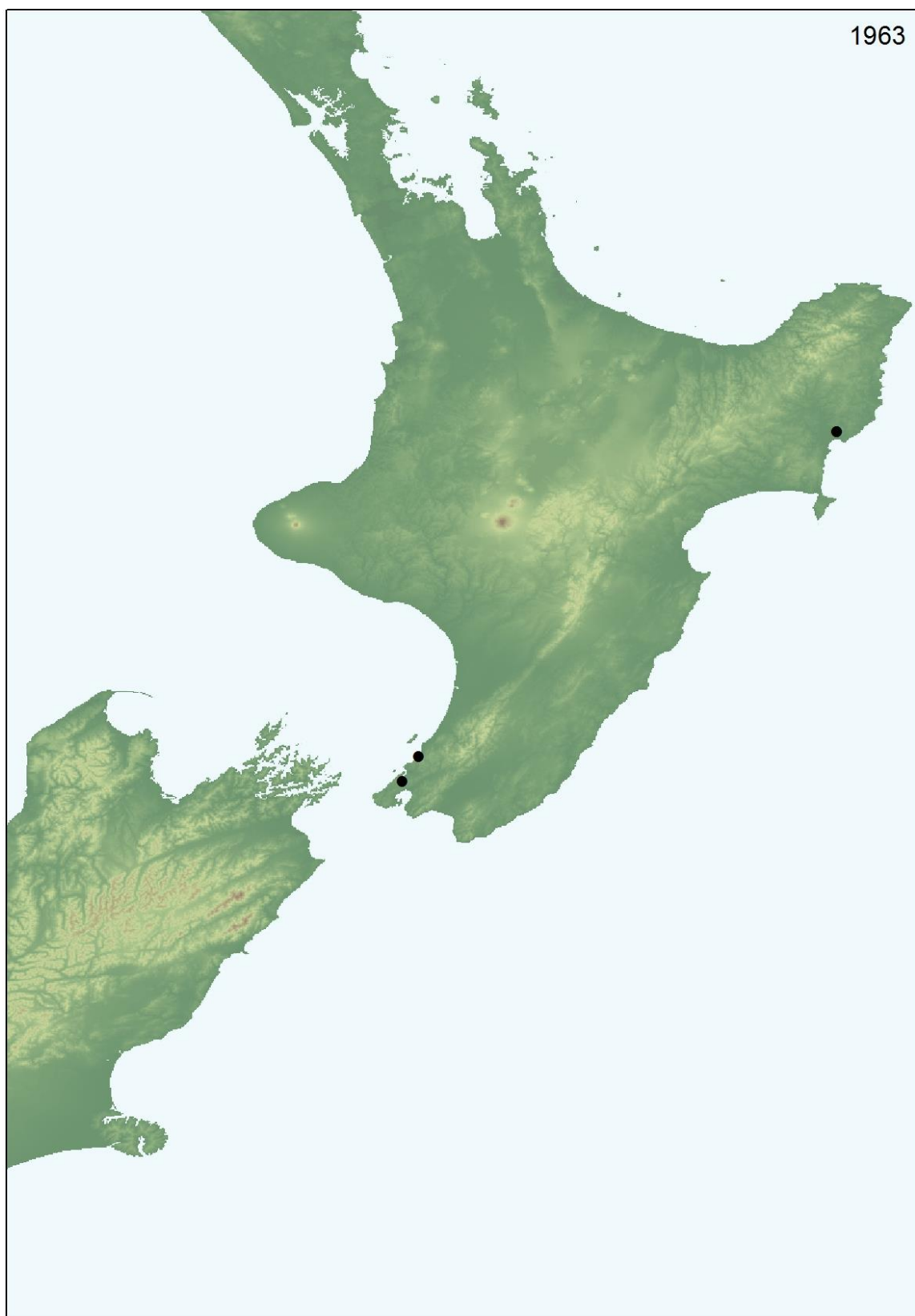
4.4 Concluding remarks

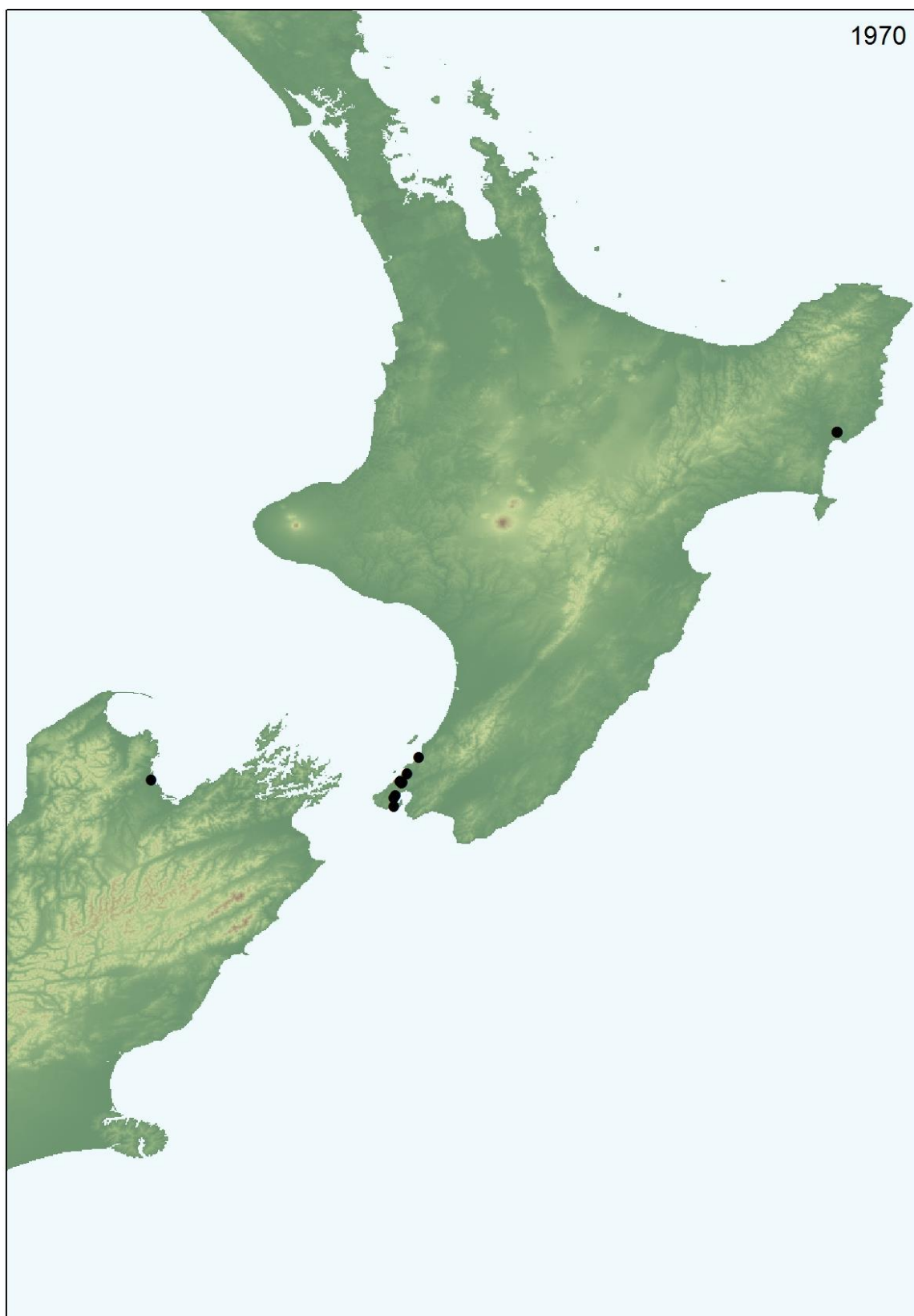
Conceptual challenges in formulating and validating such models remain, despite continual advances in the field of SDM. With careful attention to the limitations of the study systems, predictive models of invasive species distributions can offer useful information for the management of invasive species, as well as explore questions of ecological niche biogeography.

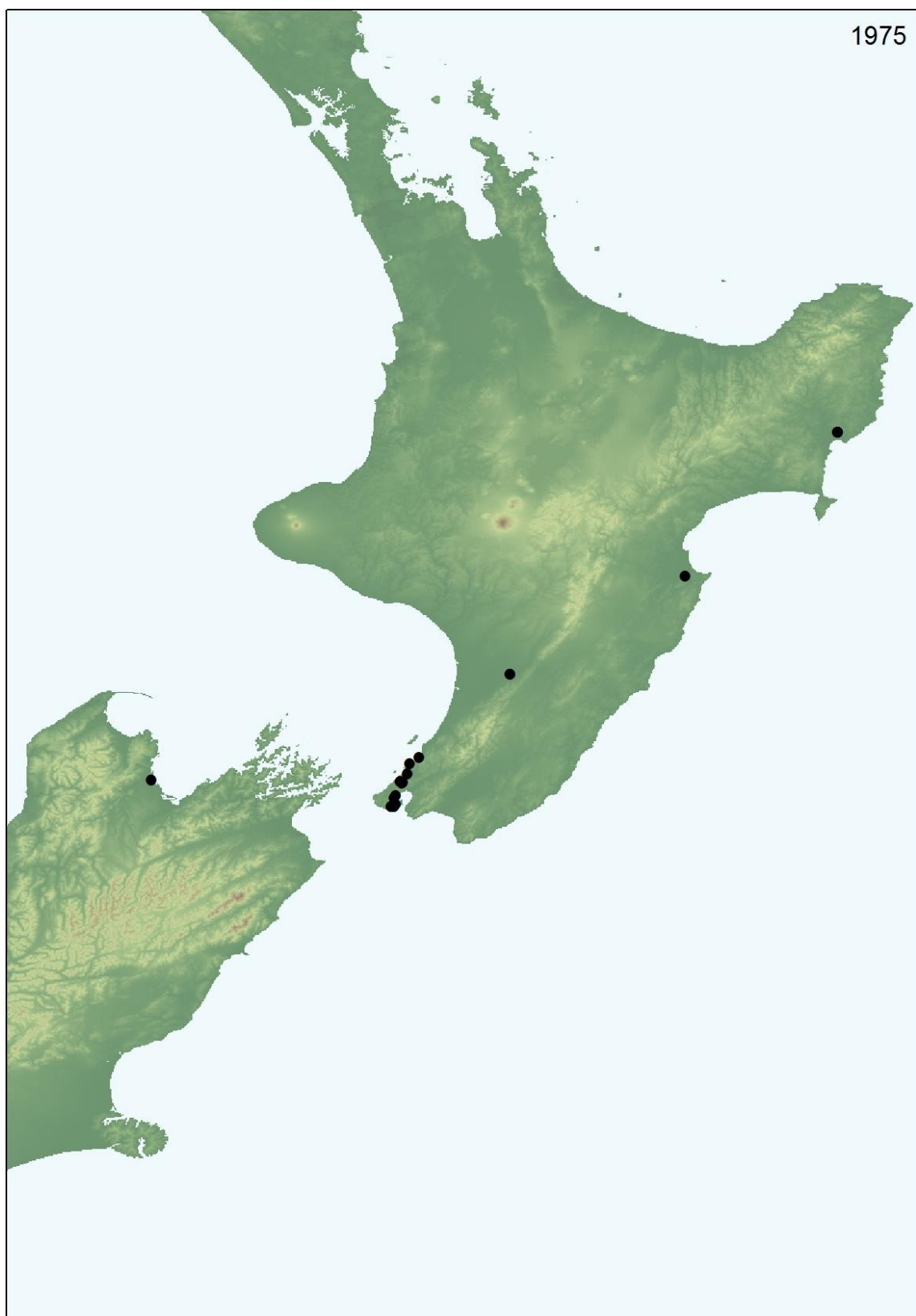
4.5 References

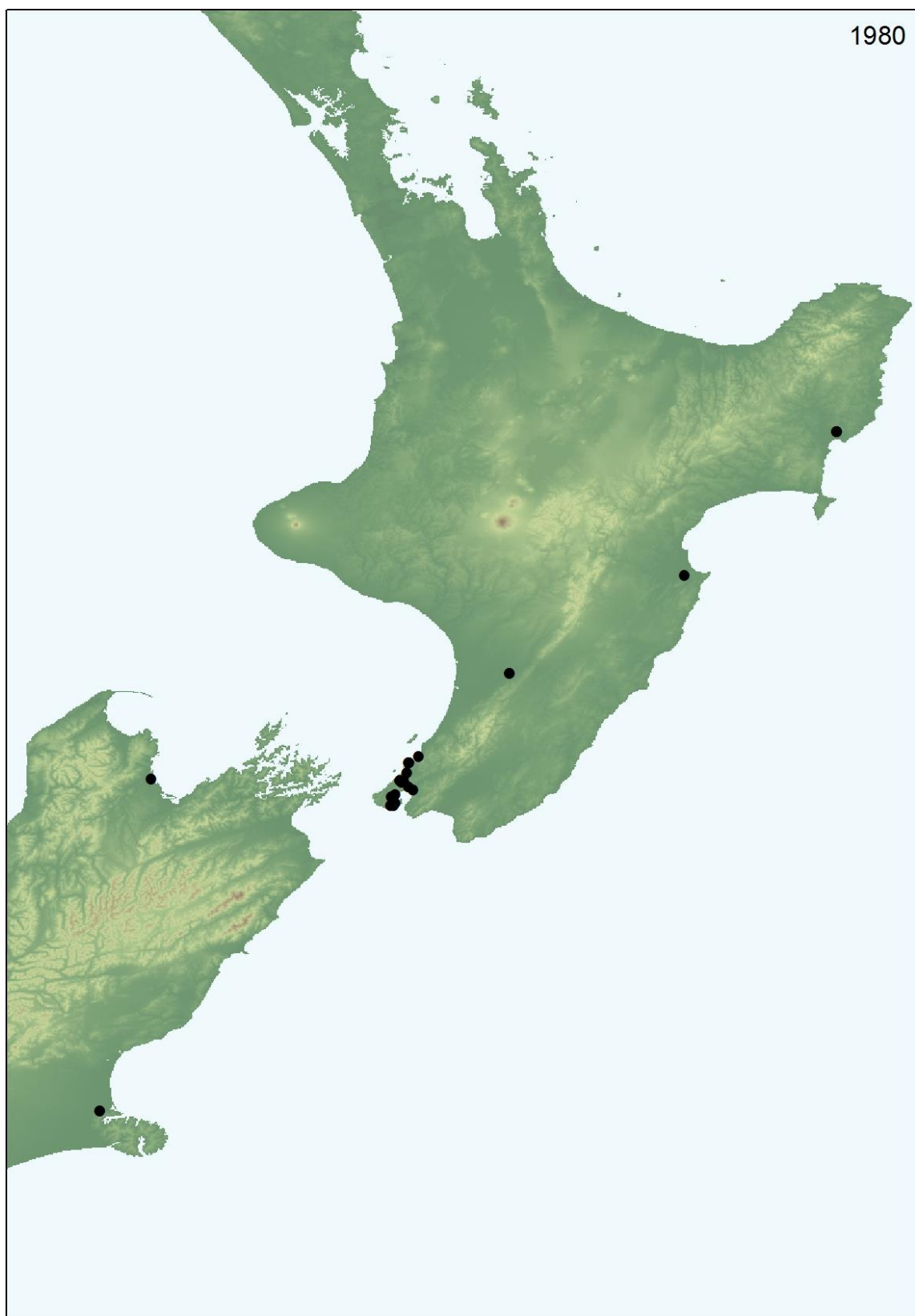
- Elith, J., S. J. Phillips, et al. (2011). "A statistical explanation of MaxEnt for ecologists." Diversity and Distributions **17**(1): 43-57.
- Lobo, J. (2008). "More complex distribution models or more representative data?" Biodiversity Informatics **5**: 14-19.

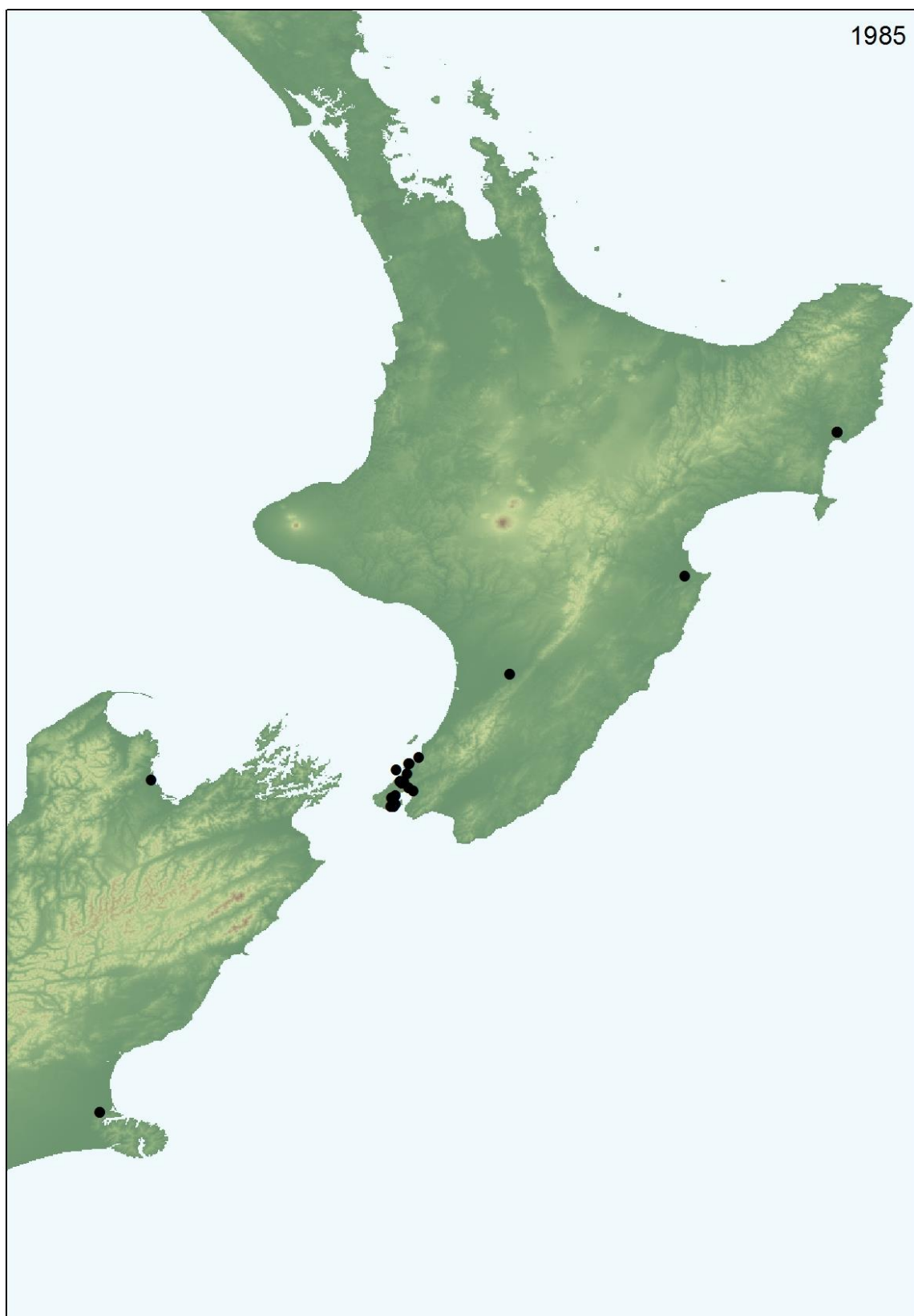
Appendix 2.1 Recorded historical spread of *S. glastifolius* in New Zealand.

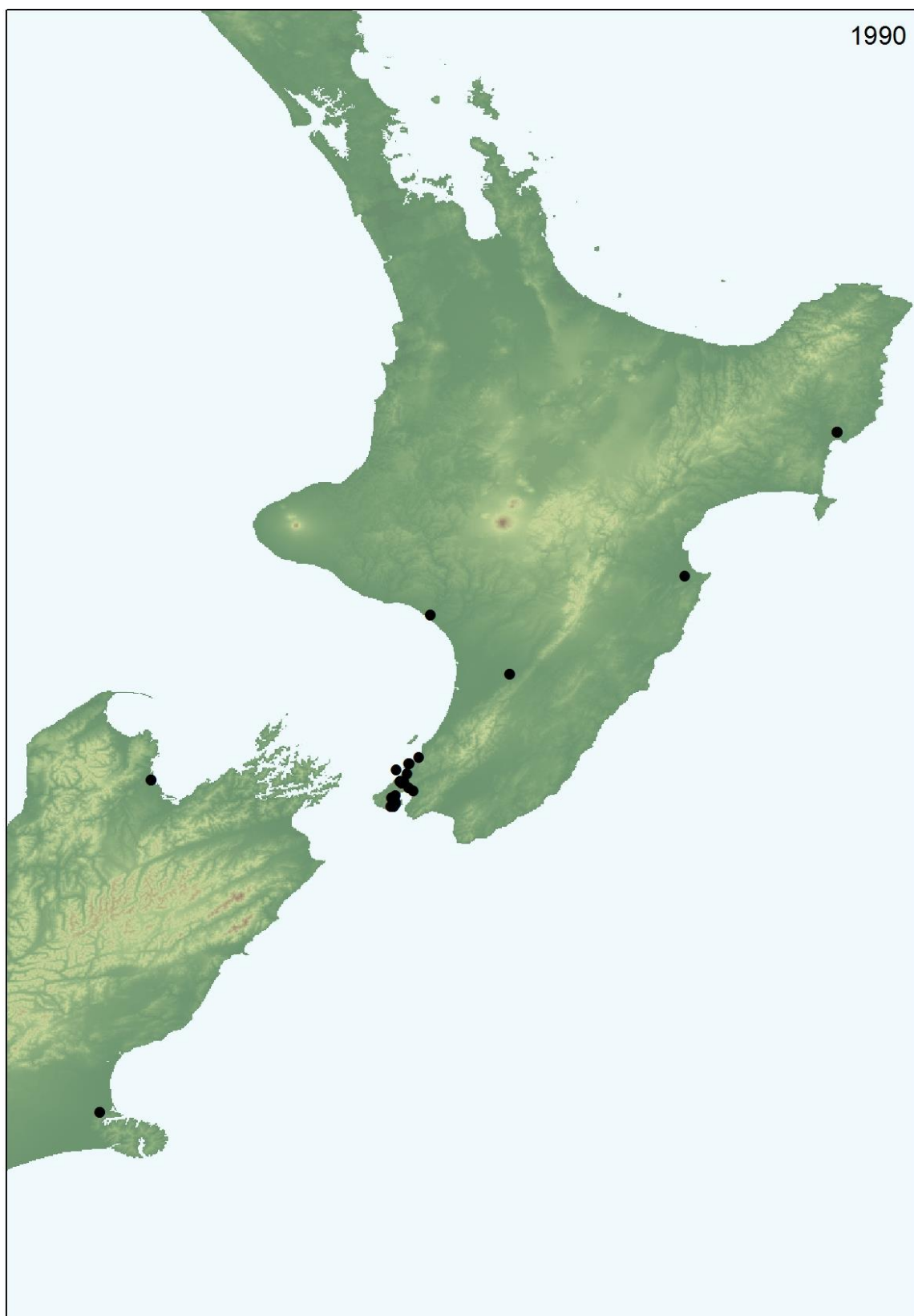


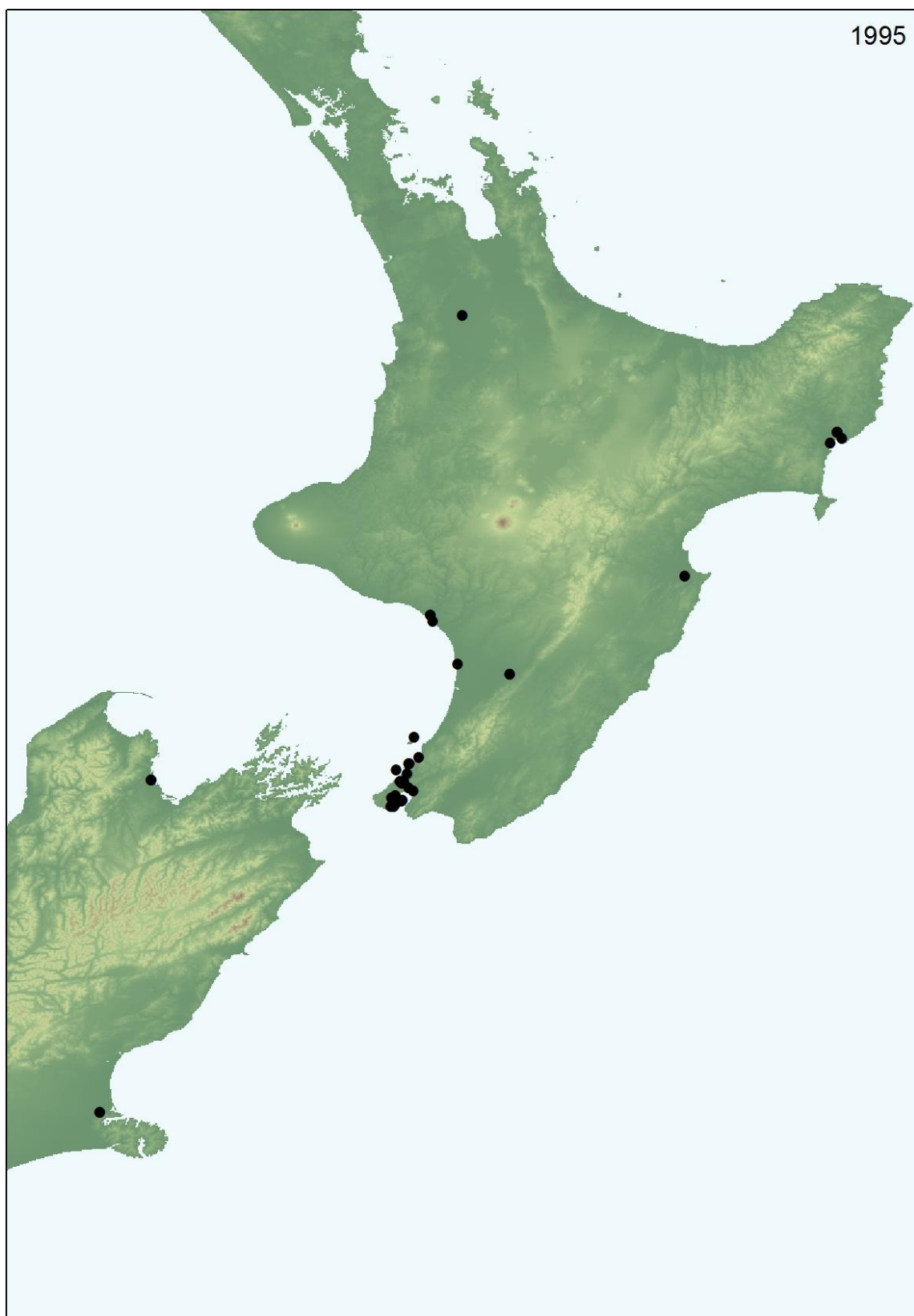


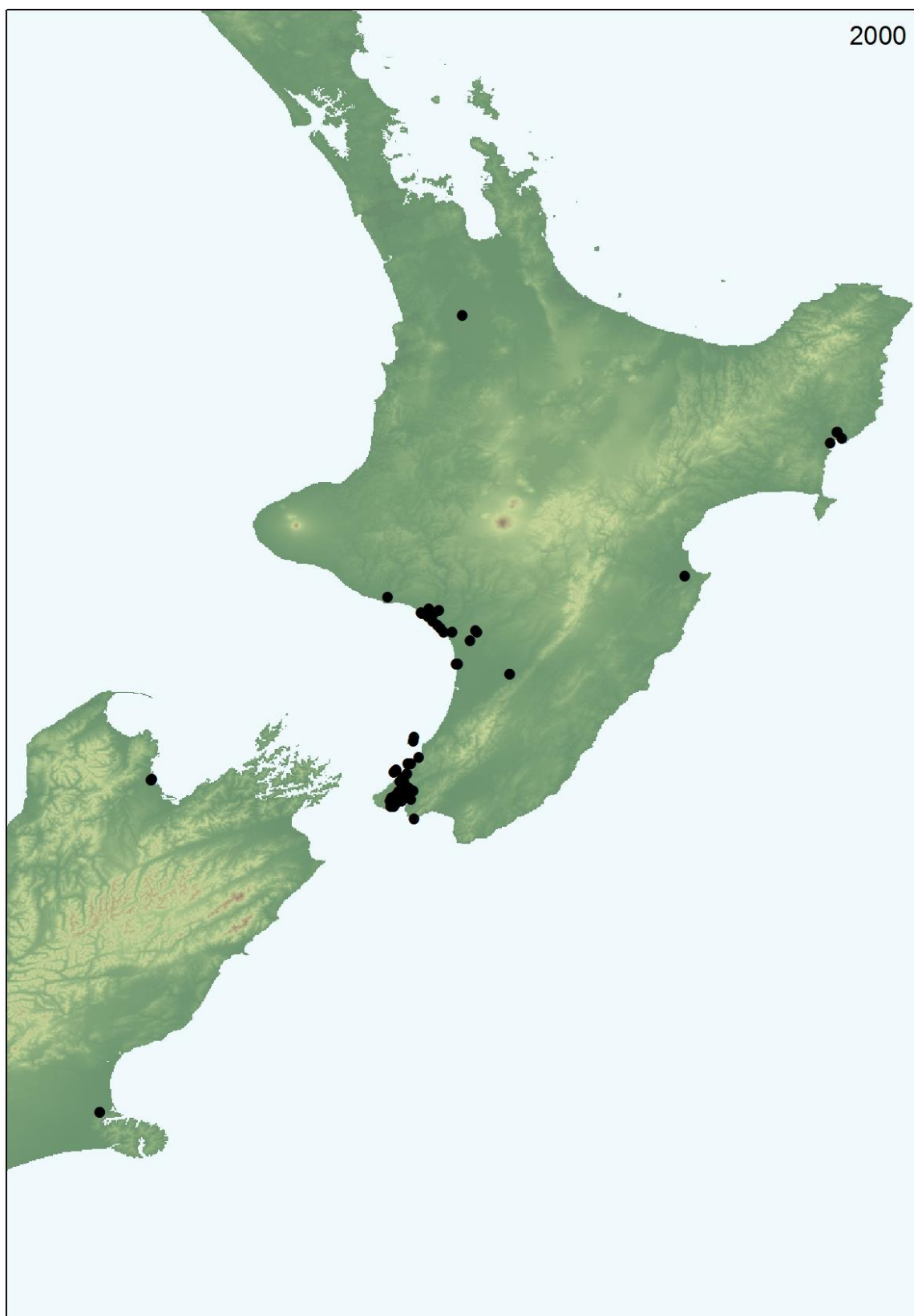


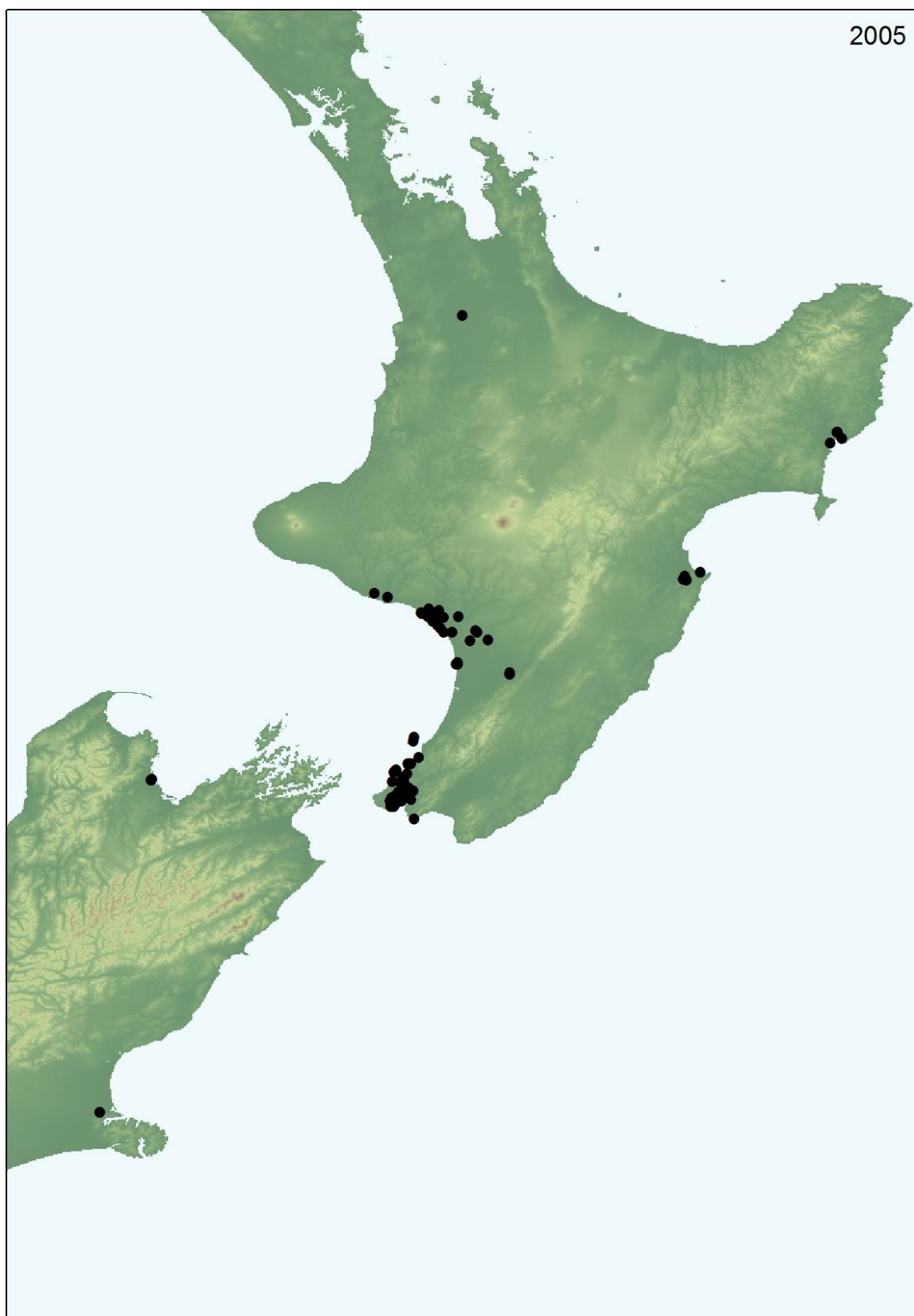


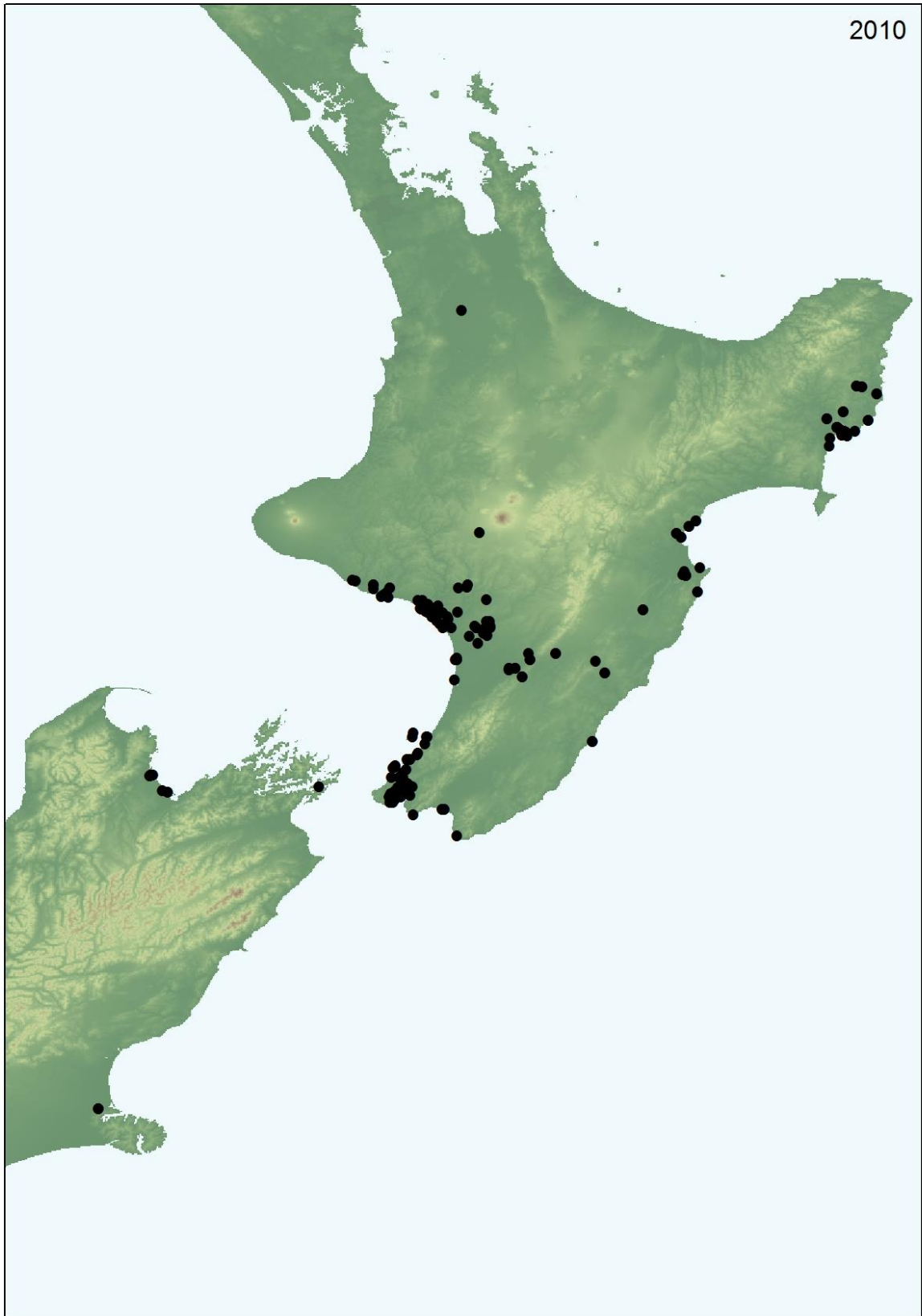


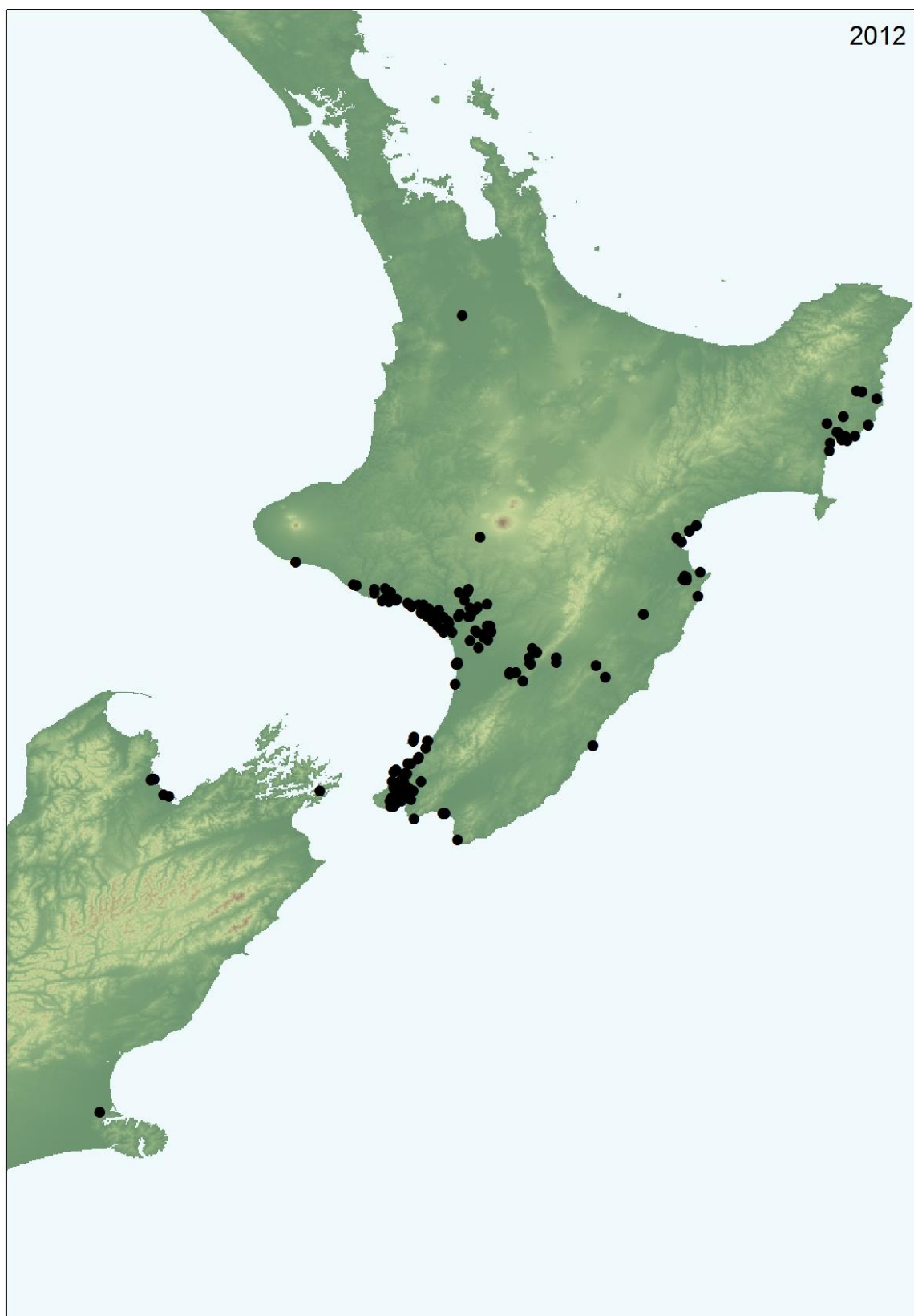












Appendix 2.2 Herbaria cited

AK: Auckland War Memorial Museum, Auckland, New Zealand.

AVH: Australia's Virtual Herbarium.

CHR: Allan Herbarium, Landcare Research, Lincoln, New Zealand.

MPN: Dame Ella Campbell Herbarium, Massey University, Palmerston North, New Zealand.

NSW: Royal Botanic Gardens, National Herbarium of New South Wales, Sydney, Australia.

NZFRI: National Forestry Herbarium, Scion, Rotorua, New Zealand.

PERTH: Western Australian Herbarium, Perth, Australia.

PRE: South African National Botanical Institute, Pretoria, South Africa.

WELT: Museum of New Zealand Te Papa Tongarewa Herbarium, Wellington, New Zealand.

Appendix 3.1 Comparison of MaxEnt models using raw and transformed data for MinTemp

Methods

To illustrate the effect of transforming the MinTemp variable (average daily minimum air temperature of the coldest month) from the Celsius temperature scale to the Kelvin temperature scale, equivalent MaxEnt models were trained: one using transformed data (Kelvin), the other using raw (Celsius). All MaxEnt settings were held equal, and the same presence and background points used for each. Predictor variables used were MinTemp, Aridity Index and Annual Temperature Range.

Results

Models varied in several key ways: AUC score, predicted suitabilities, number of features created, and contribution of each variable to the model (other differences not reported). The maximum absolute difference between predicted suitability values when projected to New Zealand was a substantial 0.36 (see Figure A3.1.3). Differences between predictions were spatially patterned, with the greatest differences occurring in the northern third of the North Island.

Table A3.1.1. Summary of some key differences between models.

	AUC (training)	Minimum Training Presence	Number of Features
Celsius (Raw)	0.891	0.014	18
Kelvin (Transformed)	0.890	0.025	17

Table A3.1.2. Contribution of predictor variables to MaxEnt model.

Variable	Percentage Contribution (C)	Percentage Contribution (K)
Annual Temperature Range	46.1	45
MinTemp	36.6	38.6
Aridity Index	17.4	16.5

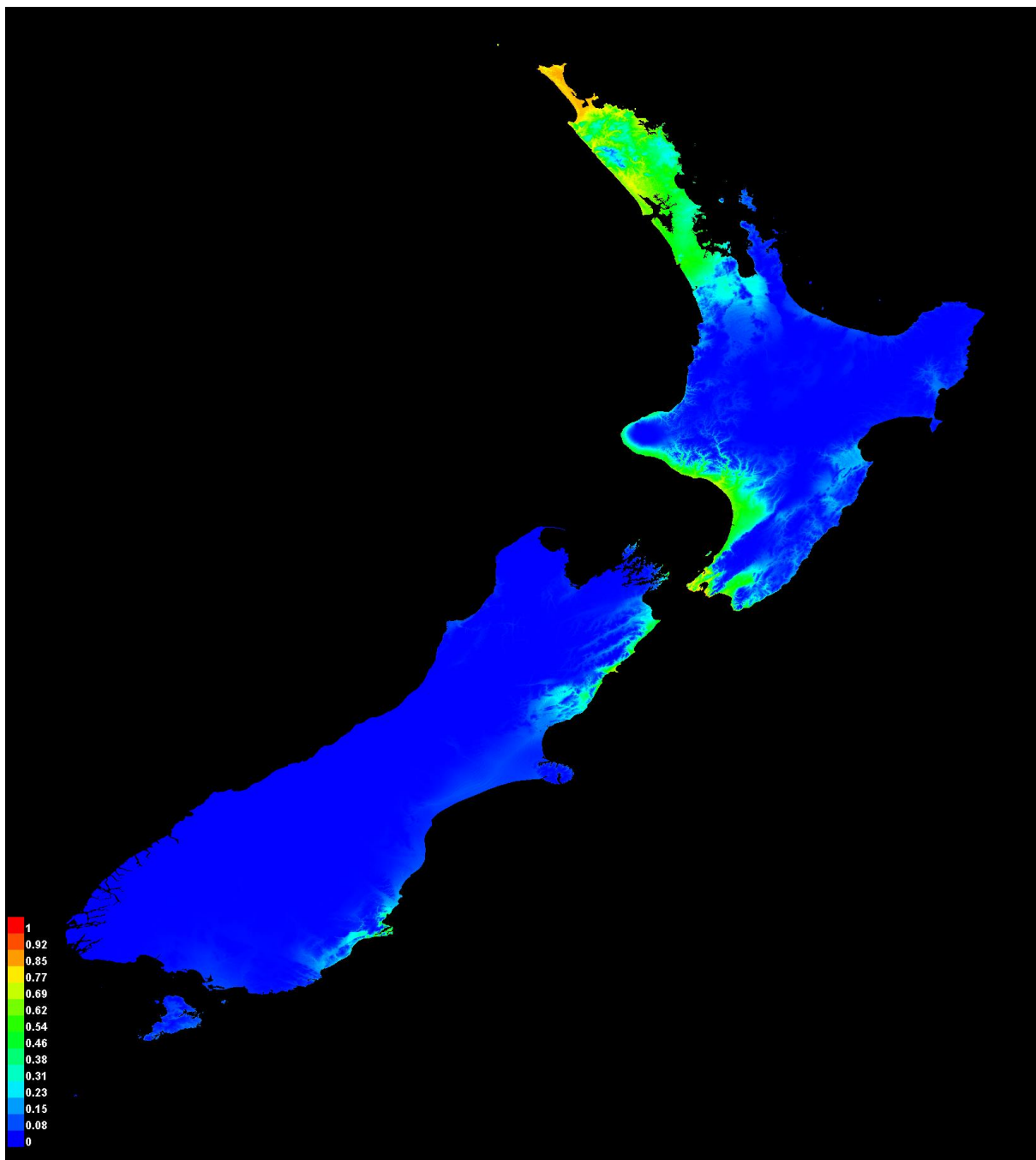


Figure A3.1.1. Mapped predictions for New Zealand using transformed (Kelvin scale) data for MinTemp.

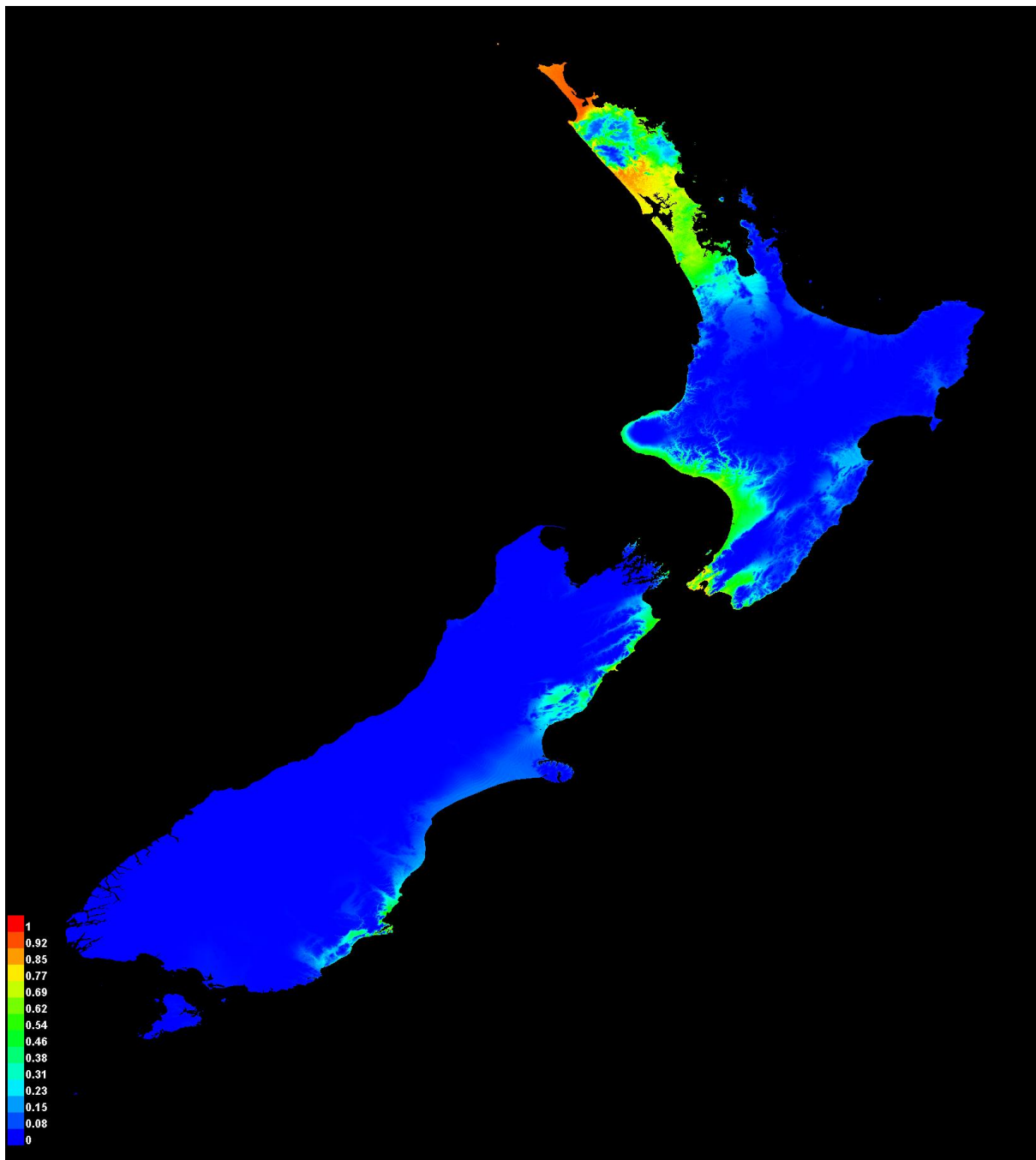


Figure A3.1.2. Mapped predictions for New Zealand using raw data (Celsius scale) for MinTemp.

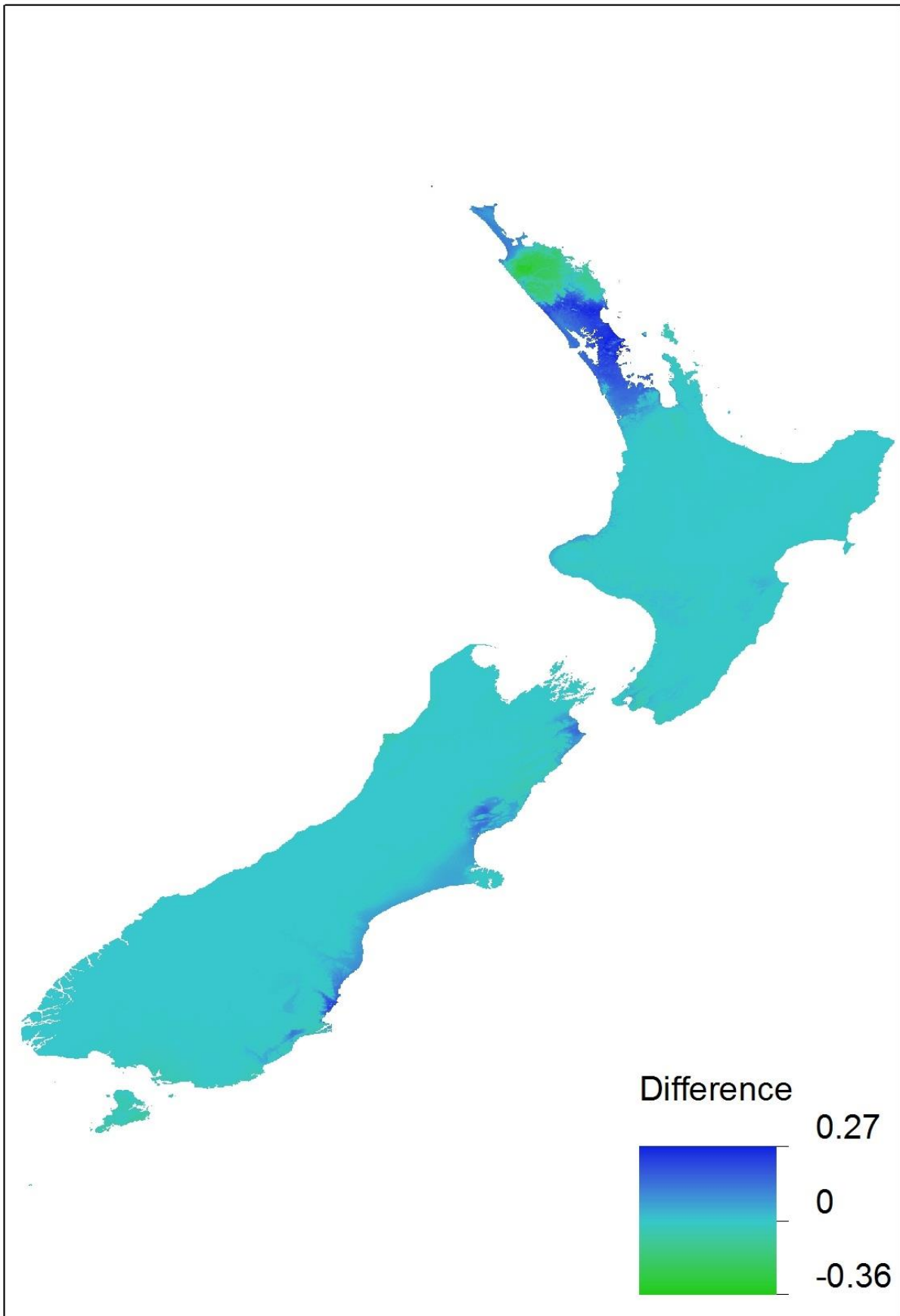


Figure A3.1.3. Difference in predicted values of each model (raw minus transformed), projected onto New Zealand.

Appendix 3.2: Climatic suitability predictions for *Senecio glastifolius* using BIOCLIM.

BIOCLIM models for *S. glastifolius* were created, to explore the hypothesis that a geometric, presence-only technique would provide better prediction to new environments than MaxEnt.

Presence training data and environmental predictor variables used were the same as for the full MaxEnt-based models described in the main text. Models were fit in R version 2.14.2 (R Core Development Team, 2012) using the package ‘dismo’ (Hijmans et al., 2012). Dismo’s implementation of BIOCLIM allows specification of tails (see explanation in main text, section 3.3.6.2). Both tails were used for Aridity Index and Annual Temperate Range; only the lower tail was used for MinTemp, mirroring the post-hoc threshold approach of the MaxEnt-based predictions. Models were evaluated using 4-fold regional cross-validation. Results from the globally fitted model are also reported, resubstituting training data. Statistics reported are the minimum training presence; omission on test data using minimum training presence as the cut-off value, and AUC.

Table A3.2.1 Results summary for BIOCLIM models

Test Region:	Global	New Zealand	South Africa	Australia	Scilly
Min Training Presence	0.005	0.013	0.006	0.006	0.005
Omission	0	0.02238806	0.27777778	0	1
AUC	0.8179248	0.7977855	0.7907253	0.8471313	0.5

Comparison to MaxEnt

Evaluation scores (omission and AUC) for BIOCLIM models were lower than those for MaxEnt in all cases. Mapped predictions show very high similarity in the spatial pattern of predicted suitability. This is true for all regions. The most noticeable difference is that BIOCLIM is generally less permissive, predicting spatially more restricted areas of suitability. This is because BIOCLIM assigns a value of zero to

any grid cell with one or more environmental variable outside of the range of training data. Conversely, MaxEnt can assign potentially very high suitability to cells with values outside the training range.

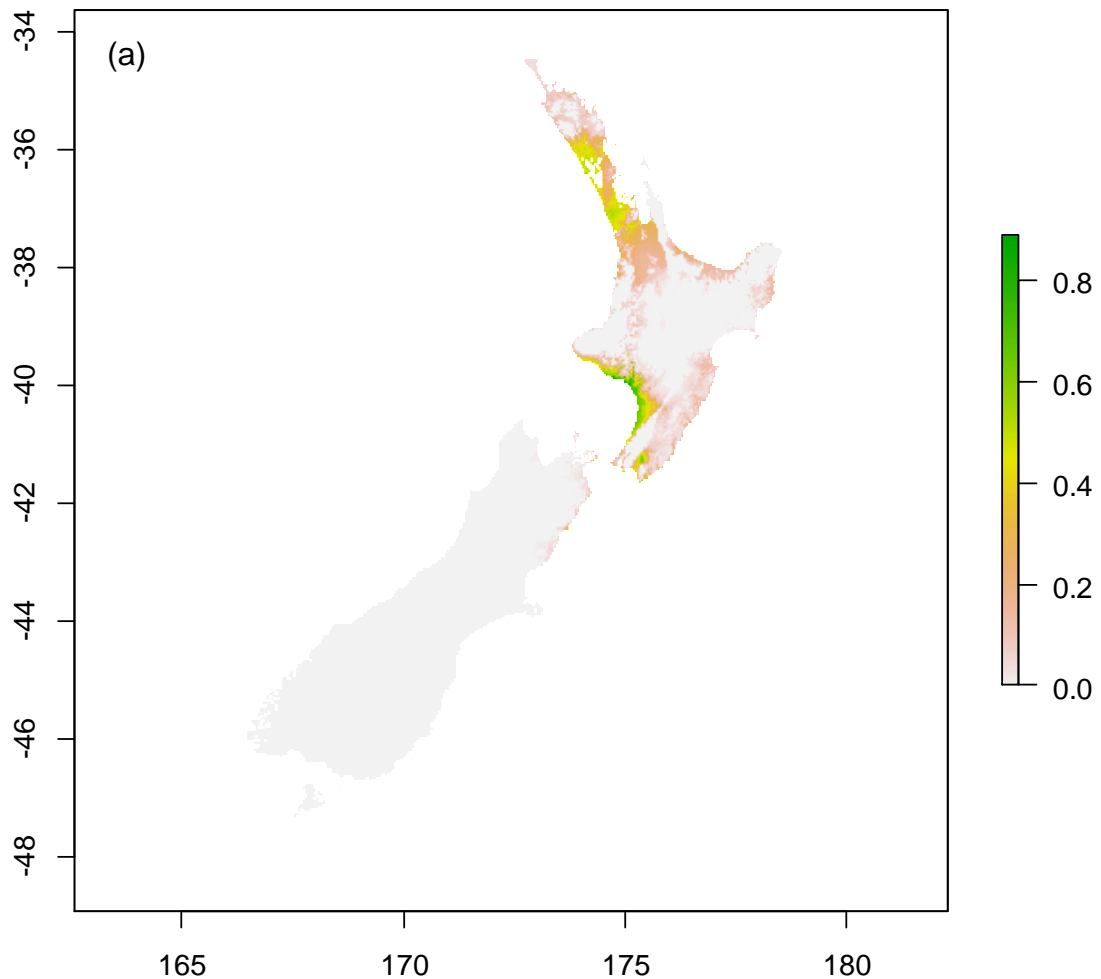
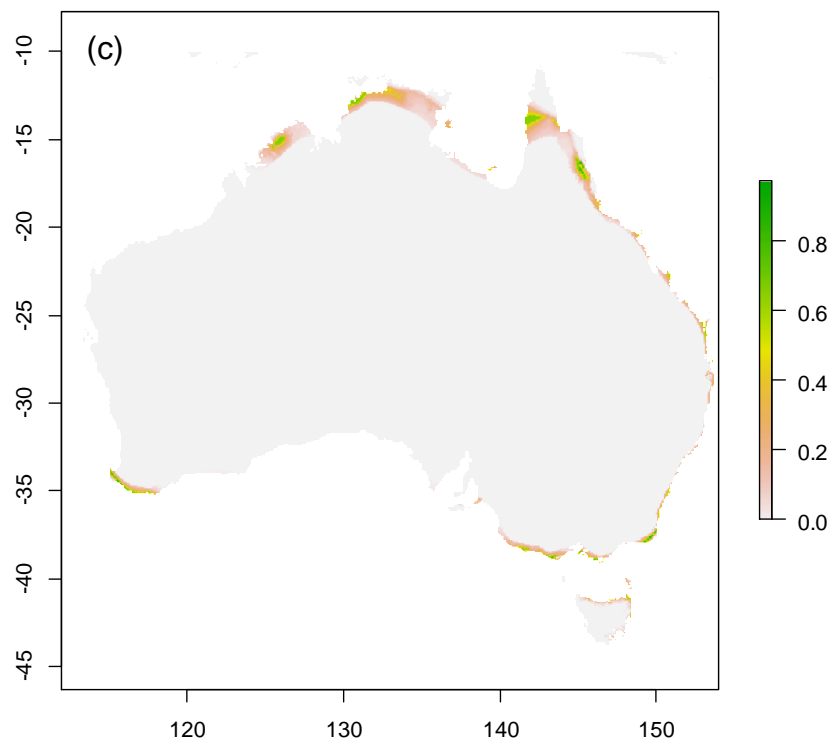
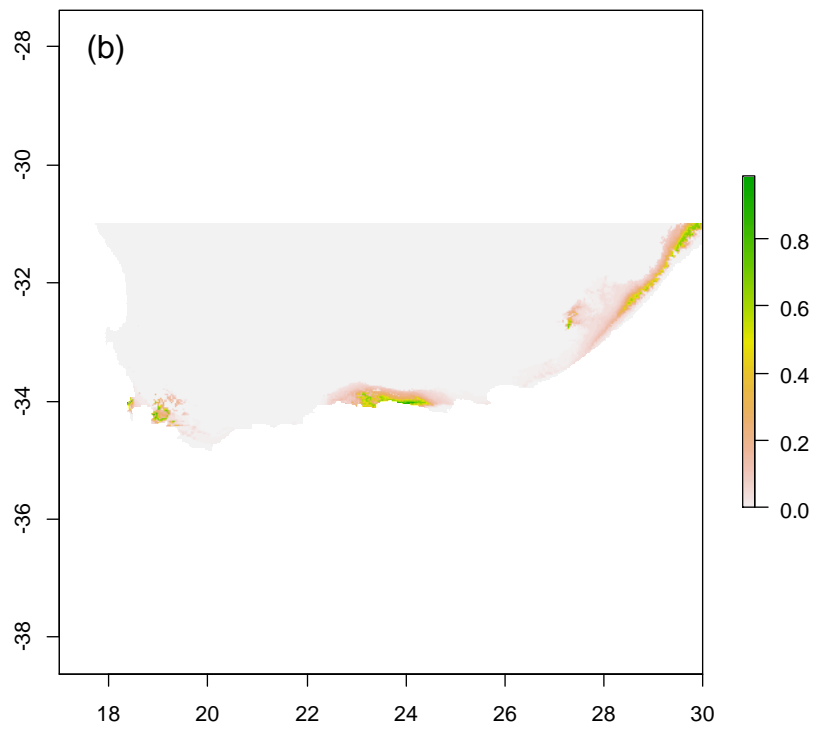
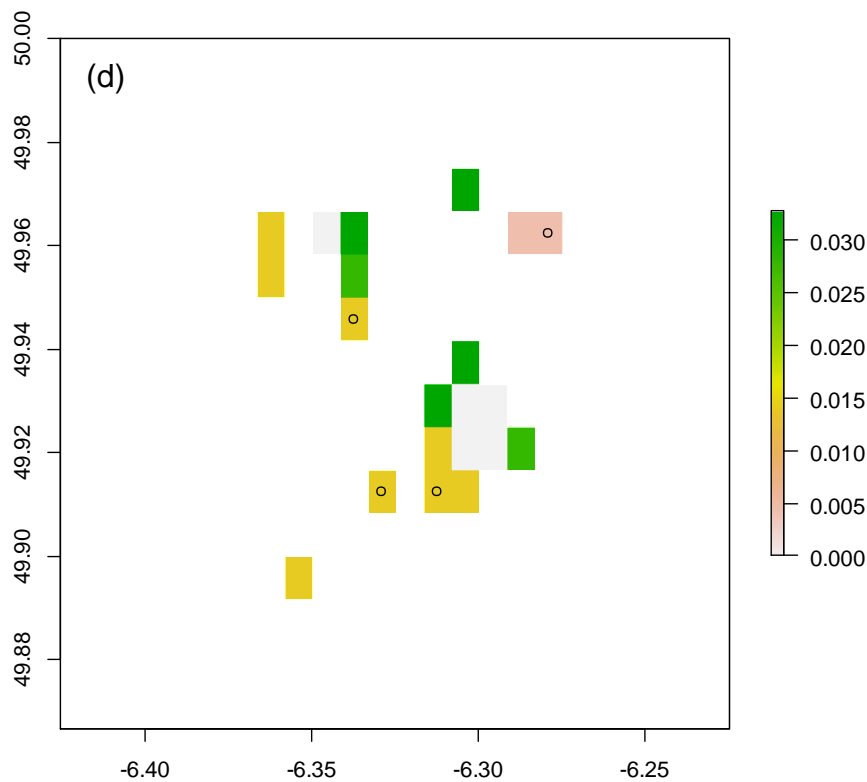


Figure A3.2.1 Globally fitted BIOCLIM model for *S. glastifolius*, projected to New Zealand (a); South Africa (b); Australia (c); and the Isles of Scilly (d). Colour scale represents predicted suitability.





References

Hijmans, R. J., S. Phillips, et al. (2012). "dismo: Species distribution modelling". R package version 0.7-17. <http://CRAN.R-project.org/package=dismo>

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/>.

Appendix 3.3: Correlation structure of predictor variables

Correlation structure of predictor variables requires the careful attention of the modeller. Notably, changing correlation structures between data sets (e.g. study regions) can severely reduce the predictive ability of a model trained on one data set to predict to another (Elith et al., 2010; Harrel 2001).

Correlation between two of the key predictor variables, aridity and annual temperature range, was analysed. Correlation with the third variable, minimum temperature of the coldest month, was not seen as important because this variable was included as contributing only a single limiting threshold value.

Table A3.3.1. Correlation between Aridity and Annual Temperature Range, reporting Pearson correlation coefficients.

Region:	Global	New Zealand	South Africa	Australia	Scilly
r	-0.628	-0.226	-0.812	-0.770	0.118
p-value	<0.001	<0.001	<0.001	<0.001	0.095

Correlation strength ranges from low (New Zealand) to fairly high, with correlation for South Africa approaching a heuristic maximum tolerable value of ± 0.85 (Elith et al., 2006). The direction of relationship (negative) was constant across all regions for which correlation was significant.

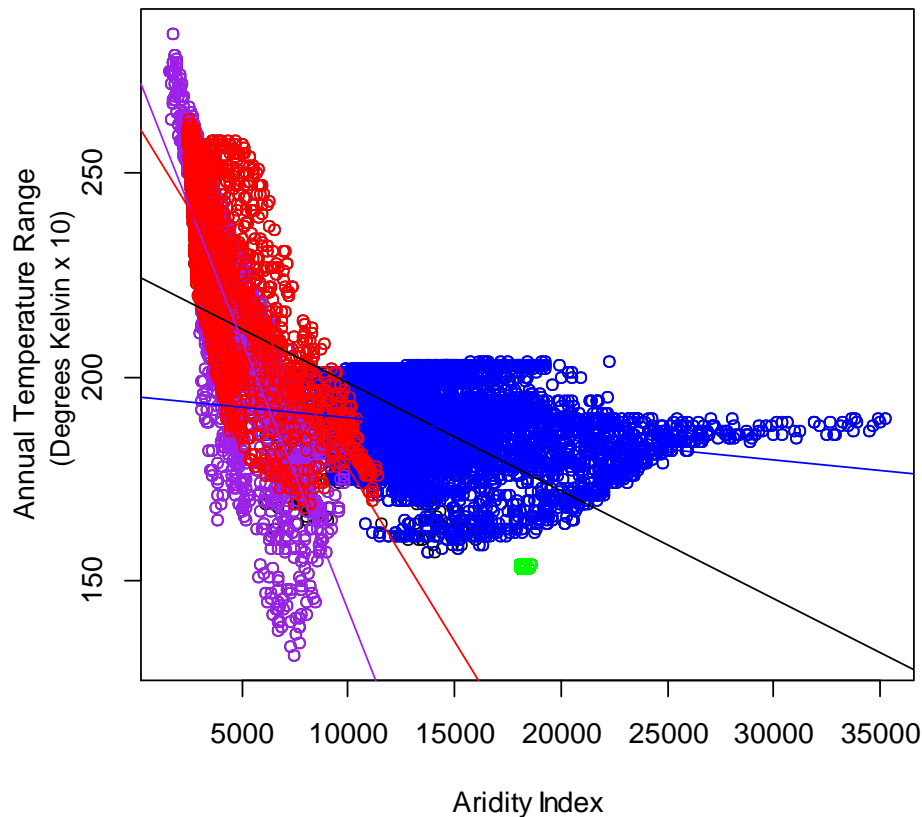


Figure A3.3.1. 'Background' points plotted by study region, with associated lines of best fit. Blue = New Zealand, Purple = South Africa, Red = Australia, green = Scilly. The black line represents the global model. No line is plotted for Scilly because the relationship is not significant.

References

- Elith, J., C. H. Graham, et al. (2006). "Novel methods improve prediction of species' distributions from occurrence data." *Ecography***29**(2): 129-151.
- Elith, J., M. Kearney, et al. (2010). "The art of modelling range-shifting species." *Methods in Ecology & Evolution*.
- Harrell, F.E. (2001) Regression Modeling Strategies with Applications to Linear Models, Logistic Regression and Survival Analysis. Springer Verlag, New York.

Appendix 3.4: Predicted Suitability of Madeira Archipelago for *Senecio glastifolius*

The globally fitted MaxEnt model, described in the main text, section 3.3.6.1, was projected onto Portugal's Madeira Islands. Minimum temperature of the coldest month was not included in the prediction because the lowest interpolated value for Madeira was 3.3°C, higher than the observed limiting threshold of ~2.3°C.

The model predicts most of the land area of Madeira as climatically suitable to *S. glastifolius*. This result reinforces the credibility of the model for use in predicting to new regions.

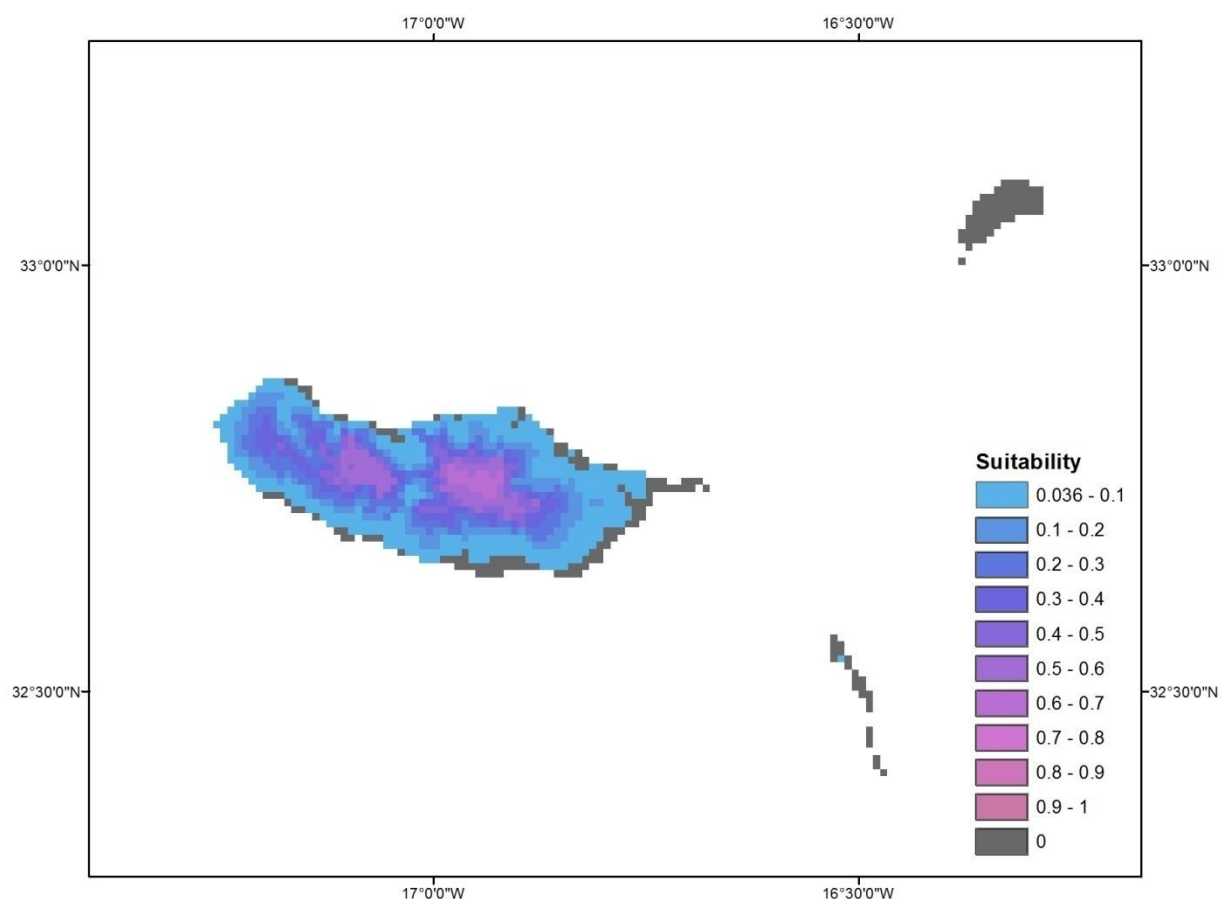


Figure A3.4.1 Predicted climatic suitability for *S. glastifolius* in Madeira.

Areas of low predicted suitability for the archipelago were probably penalised by values for Annual Temperature Range being below the lower bound of training data

(both presence and background). This is not necessarily an error of prediction; rather, no information is available as to the suitability of that climatic space for *S. glastifolius*.

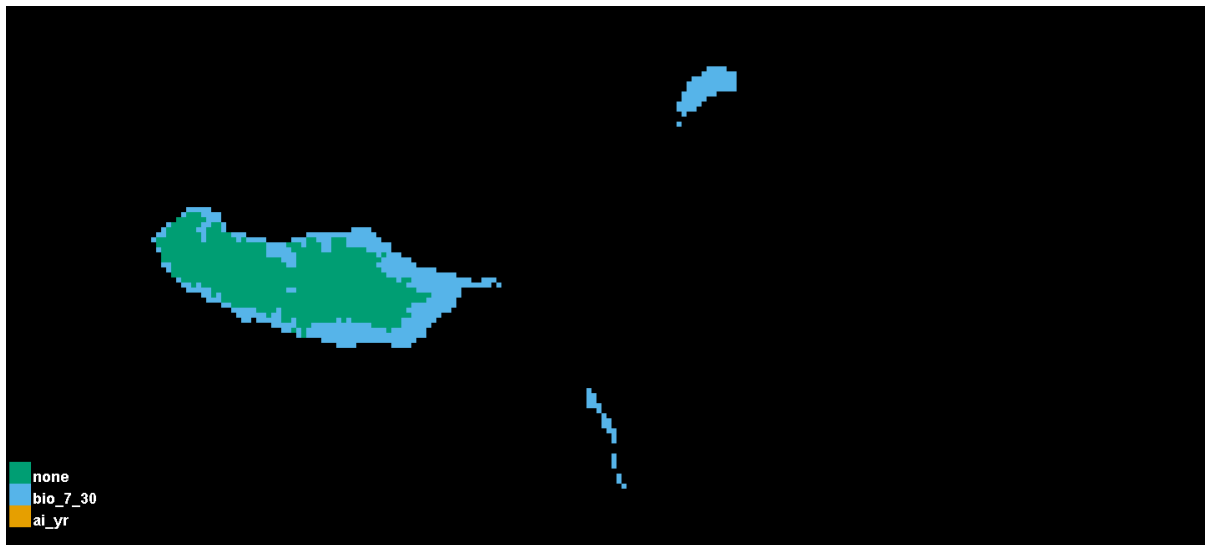


Figure A3.4.2. Predictor variables outside of training data range. “bio_7_30” represents Annual Temperature Range; “ai_yr” represents Aridity Index.

Appendix 3.5: Invasion to New Zealand by *Senecio glastifolius* reveals niche space not available in native range of South Africa

Species introductions to new areas sometimes reveal apparent shifts in ecological niche (section 1.5.3). Section (3.1.1) invited speculation that the fundamental niche of *S. glastifolius* was not fully available in its native range of South Africa. *Senecio glastifolius* shows a strong association with wet conditions in its narrow native range. Within this range, which is more humid than surrounding areas, it shows micro-habitat selection of wet slopes, hollows, stream banks etc. New Zealand is much less arid, and so presumably offers favourable climatic space not available in South Africa. To explore this idea, the realised niches of *S. glastifolius* in these two geographic regions were examined visually in environmental space, along with corresponding biotopes (available conditions).

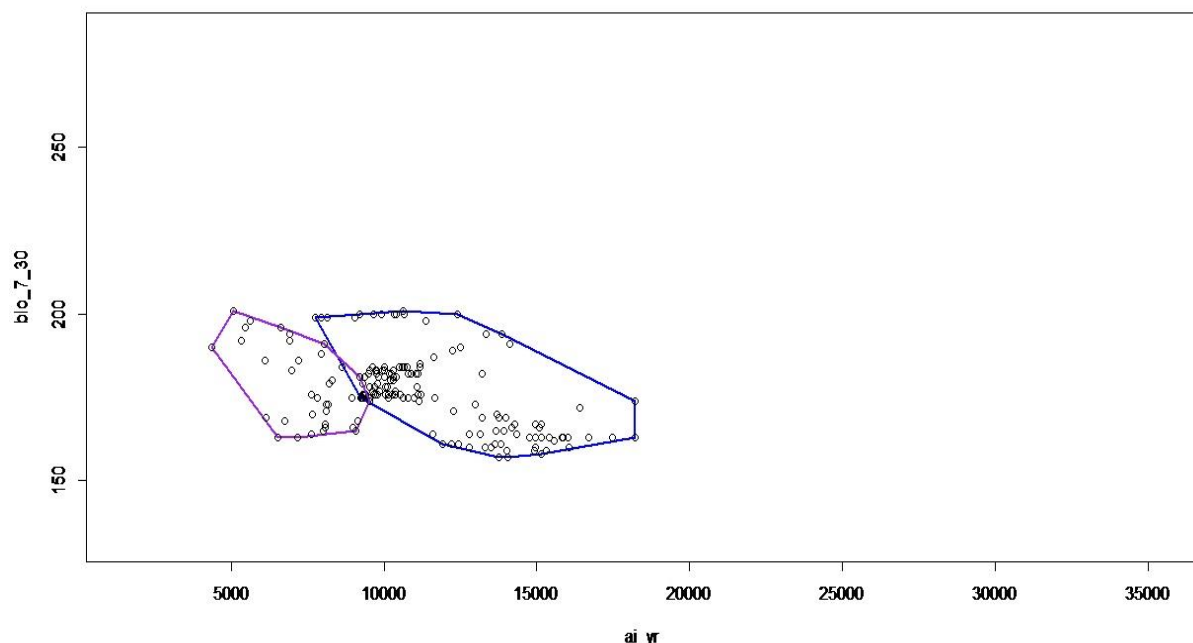


Figure A3.5.1 Apparent niche shift of *S. glastifolius* between South Africa (Purple) and New Zealand (Blue). Recorded presence data are expressed in environmental space, with axes Annual Temperature Range (labelled as “bio_7_30”) and Aridity (labelled as “ai_yr”).

Figure A3.5.1 presents the niche space of *S. glastifolius* in the two climatic dimensions most influencing of its distribution: Annual Temperature Range and

Aridity. Minimum convex polygons are drawn around points of known species occurrence; purple representing occurrence in South Africa, blue representing occurrence in New Zealand. These polygons can be taken to represent the realised niche of *S. glastifolius* in the two respective geographic regions. Observe that the two polygons barely overlap. While the species occupies roughly the same range of values for Annual Temperature Range in both geographic regions, it is associated with much more humid conditions in New Zealand. This pronounced shift in the realised niche could be interpreted in a number of ways, but a likely hypothesis emerges once the available climatic space in each geographic region is considered (Figure A3.5.2). Of the climate space occupied in New Zealand but not South Africa, almost none is available to the species in its native range of South Africa. Note too that due to imperfect sampling of the species' distribution and contingent factors which limit geographic distribution, space unoccupied on this plot might in reality be climatically suitable in terms of the two axes considered.

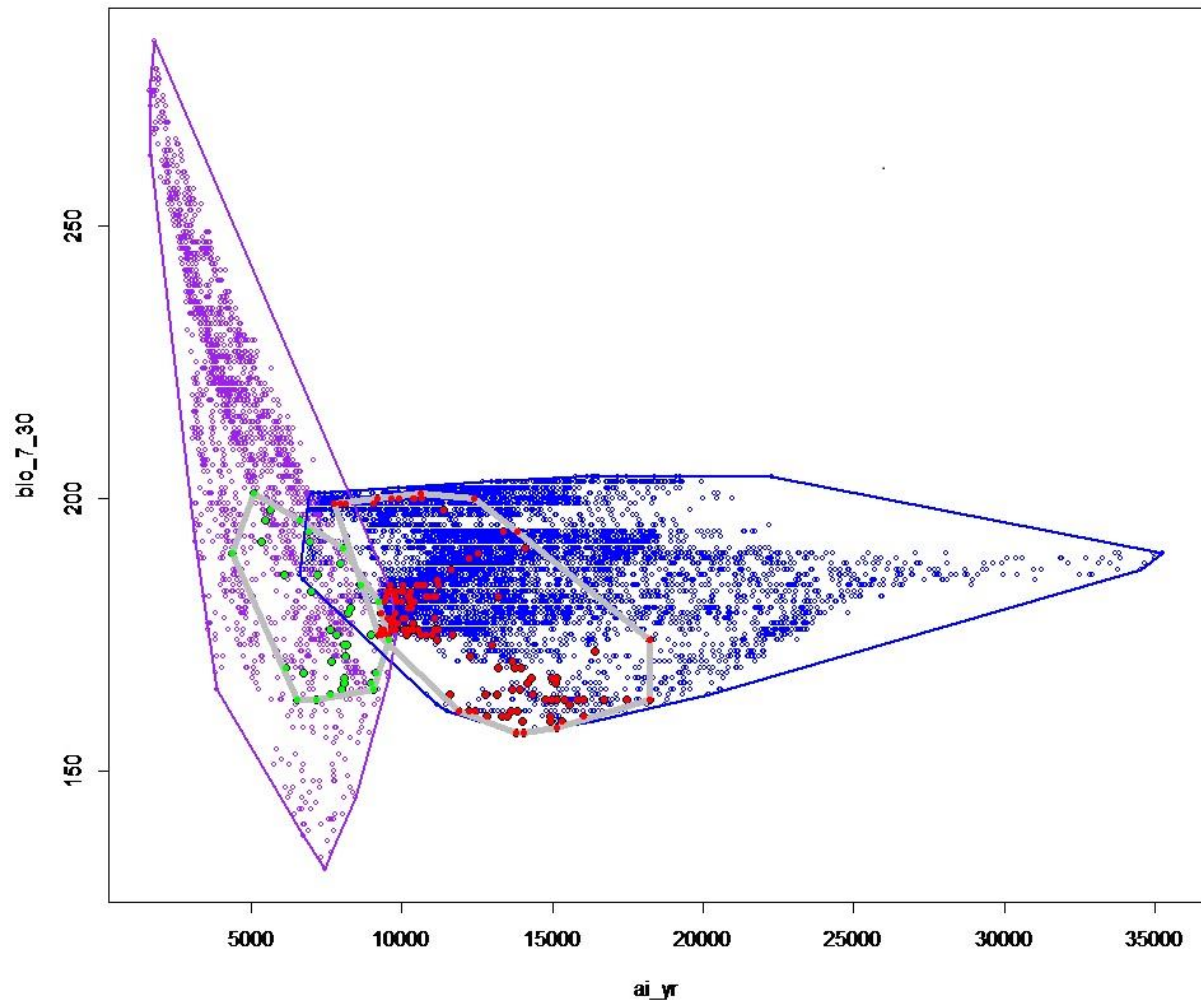


Figure A3.5.2 Change in available environmental space between South Africa and New Zealand. Purple points and associated minimum convex polygon represent the biotope of *S. glastifolius* in South Africa, as measured by a random sample of background points. Blue points and polygon represent the biotope of the New Zealand study region. Presence points for each country are plotted in contrasting colours, and fitted with minimum convex polygons for illustrative purposes.