

**Investigating the influence of data quality on ecological niche models for
alien plant invaders**

by

René Wolmarans

Submitted in partial fulfilment of the requirements for the degree

M.Sc. Zoology

in the Faculty of Natural and Agricultural Science

University of Pretoria

Pretoria

December 2009



I, declare that the
thesis/dissertation, which I hereby submit for the degree
..... at the University of Pretoria, is my
own work and has not previously been submitted by me for a degree at this or any other
tertiary institution.

SIGNATURE:

DATE:

Investigating the influence of data quality on ecological niche models for alien plant invaders

Student: René Wolmarans

Supervisor: Dr. M.P. Robertson

Co-Supervisor: Dr. B. Janse van Rensburg

Department: Department Zoology & Entomology and Centre for Invasion Biology, Faculty of Natural and Agricultural Science, University of Pretoria, Pretoria, 0002, South Africa

Degree: M.Sc. Zoology

Abstract:

Ecological niche modelling is a method designed to describe and predict the geographic distribution of an organism. This procedure aims to quantify the species-environment relationship by describing the association between the organism's occurrence records and the environmental characteristics at these points. More simply, these models attempt to capture the ecological niche that a particular organism occupies. A popular application of ecological niche models is to predict the potential distribution of invasive alien species in their introduced range. From a biodiversity conservation perspective, a pro-active approach to the management of invasions would be to predict the potential distribution of the species so that areas susceptible to invasion can be identified. The performance of ecological niche models and the accuracy of the potential range predictions depend on the quality of the data that is used to calibrate and evaluate the models. Three different types

of input data can be used to calibrate models when producing potential distribution predictions in the introduced range of an invasive alien species. Models can be calibrated with native range occurrence records, introduced range occurrence records or a combination of records from both ranges. However, native range occurrence records might suffer from geographical bias as a result of biased sampling or incomplete sampling. When occurrence records are geographically biased, the underlying environmental gradients in which a species can persist are unlikely to be fully sampled, which could result in an underestimation of the potential distribution of the species in the introduced range. I investigated the impact of geographical bias in native range occurrence records on the performance of ecological niche models for 19 invasive plant species by simulating two geographical bias scenarios (six different treatments) in the native range occurrence records of the species. The geographical bias simulated in this study was sufficient to result in significant environmental bias across treatments, but despite this I did not find a significant effect on model performance. However, this finding was perhaps influenced by the quality of the testing dataset and therefore one should be wary of the possible effects of geographical bias when calibrating models with native range occurrence records or combinations thereof. Secondly, models can be calibrated with records obtained from the introduced range of a species. However, when calibrating models with records from the introduced range, uncertainties in terms of the equilibrium status and introduction history could influence data quality and thus model performance. A species that has recently been introduced to a new region is unlikely to be in equilibrium with the environment as insufficient time will have elapsed to allow it to disperse to suitable areas, therefore the occurrence records available would be unlikely

to capture its full environmental niche and therefore underestimate the species' potential distribution. I compared model performance for seven invasive alien plant species with different simulated introduction histories when calibrated with native range records, introduced range records or a combination of records from both ranges. A single introduction, multiple introduction and well established scenario was simulated from the introduced range records available for a species. Model performance was not significantly different when compared between models that were calibrated with datasets representing these three types of input data under a simulated single introduction or multiple introduction scenario, indicating that these datasets probably described enough of the species environmental niche to be able to make accurate predictions. However, model performance was significantly different for models calibrated with introduced range records and a combination of records from both ranges under the well established scenario. Further research is recommended to fully understand the effects of introduction history on the niche of the species.

Keywords: Ecological niche models, invasive alien plant species, potential distribution, data quality, geographical bias, introduction history, environmental niche, MAXENT, model performance.

Acknowledgements

First of all, a sincere thanks to my supervisors, Mark Robertson and Berndt Janse van Rensburg. Thank you for your patience, wisdom and the many hours you have spent working with me. Most of all, for the encouragement and good humour when it was most needed. You made this a worthwhile learning experience and I will be forever grateful.

To my parents and my brother; for your unwavering love and support through out my studies. Thank you for always providing me with a home where I felt welcome. For never discouraging any of my new ideas (even the wackiest of them all!) and for teaching me the basic principles of hard work and perseverance.

To all my friends and family. Special thanks to Elenor and Machteld, for providing me with endless hours of entertainment and always helping me to see things in perspective. Chantal and Sonja at my Pretoria base, for listening patiently and running with me when sitting behind the computer became too much. As the senior member of lab 2-9, to my lab-maatjies; MacD and Andrew, for not only academic and statistical advice, but also for all those coffee's and muffins. And to Jesse, for introducing me to the world of R. Your friendship has all contributed in a unique manner to keeping me motivated!

Finally, I would like to acknowledge the DST-NRF Centre of Excellence for Invasion Biology and the University of Pretoria for funding and academic support.

Disclaimer

Chapters 2 and 3 in this thesis have been prepared for submission to different scientific journals. As a result styles and formats may vary between all chapters in the thesis and overlap in content may occur throughout the thesis to secure publishable entities. For ease of reading, tables and figures have been placed on separate pages.



Table of contents

Contents	Page
Declaration	2
Abstract	3
Acknowledgements	6
Disclaimer	7
Table of contents	8
 Chapter 1:	
Introduction	9
 Chapter 2: Geographical bias	
Influence of geographical bias in native range occurrence records on ecological niche model performance	26
 Chapter 3: Introduction history	
Comparing model performance for species with different introduction histories following three different approaches	87
 Chapter 4:	
Summary and conclusions	125



Chapter 1

Introduction

Introduction

Ecological niche modelling

Ecological niche modelling is a method designed to describe and predict the geographic distribution of an organism (Peterson 2001; McPherson *et al.* 2004; Guisan & Thuiller 2005; McPherson & Jetz 2007). The extent of a species geographic distribution is determined by a complex interaction of several factors (Soberon & Peterson 2005). Firstly, climate conditions might impose physiological limits on a species (abiotic factors) and secondly, interactions with other species could either have a positive or negative effect on their distribution (biotic factors). Furthermore, due to the configuration of the environment, not all areas are always accessible to a species and therefore inflict dispersal limitations. The environmental niche occupied by a species describes all the environmental conditions that meet the physiological requirements that enable the species to grow and persist (Pearman *et al.* 2008). In addition, the fundamental niche describes all the requirements of a species that have to be met in order to maintain a positive population growth rate, disregarding biotic interactions (Pearman *et al.* 2008). These biotic and abiotic interactions result in species only occupying some parts of their fundamental niche, known as the realized niche (Pearson & Dawson 2003; Soberon & Peterson 2005; Pearman *et al.* 2008). Ecological niche models attempt to describe the environmental niche that an organism occupies in a particular area by quantifying the species-environment relationship (McPherson & Jetz 2007). This relationship is described by the association between the organism's occurrence records and the environmental characteristics at these points (Guisan & Zimmerman 2000;

Guisan & Thuiller 2005). However, the occurrence records used to make predictions already include the effect of biotic factors and dispersal limitation and as a result most likely only provide information on a species' realized niche (Pearman *et al.* 2008). A wide variety of correlative modelling algorithms are currently available (Guisan & Zimmermann 2000; Elith *et al.* 2006) and the extent to which the realized niche can be described by the correlative model and extent to which the realized and fundamental niches overlap may influence the success of the model (Soberon & Peterson 2005; Phillips *et al.* 2006).

Various applications of ecological niche models exist in the ecology and conservation biology disciplines (Guisan & Thuiller 2005). These models can be used to assess the impact of land-use, climate and other environmental changes on species distributions (Guisan & Thuiller 2005). Additionally, they can serve as a tool in conservation management and reserve selection (Araujo *et al.* 2004; Guisan & Thuiller 2005), and they can assist in the identification of sites where rare and endangered species have a high potential of occurring (Engler *et al.* 2004; Guisan & Thuiller 2005). A popular application of ecological niche models is to predict the potential distributions of invasive alien organisms in their introduced range so that areas that are at risk of invasion can be identified (Roura-Pascual *et al.* 2004; Mau-Crimmins *et al.* 2006; Schussman *et al.* 2006; Broennimann *et al.* 2007; De Meyer *et al.* 2007; Steiner *et al.* 2008).

Invasive species

Biological invasions are either the accidental or deliberate introduction of a species into regions outside of their natural range (native range), where these organisms overcome a

series of barriers (Richardson & van Wilgen 2004), that allow them to become established and eventually spread throughout this new region (introduced or adventive range; Vitousek *et al.* 1997; Mack *et al.* 2000). Escalating global trade and human movements across continents in conjunction with habitat transformation, increase the opportunity for organisms to disperse to novel environments and to occupy vacant niches (Vitousek *et al.* 1997; Mack *et al.* 2000; Pimental *et al.* 2001; Richardson & van Wilgen 2004; Yurkonis *et al.* 2005). Biological invasions impact upon ecosystem services, human health and natural economies at a global, regional and local scale (Vitousek *et al.* 1997; Pimentel *et al.* 2001; Richardson & van Wilgen 2004; Yurkonis *et al.* 2005; van Wilgen *et al.* 2008). Worldwide, biological invasions are known for their significant negative impacts on biodiversity (Vitousek *et al.* 1997; Pimentel *et al.* 2001; Richardson & van Wilgen 2004; Yurkonis *et al.* 2005; van Wilgen *et al.* 2008). Since it is difficult, if not impossible to eradicate invasive species once they have become naturalized and started to spread (Peterson 2003; Thuiller *et al.* 2005), a pro-active approach to the management of invasions is to predict the potential distribution of the species so that areas susceptible to invasion can be identified (Peterson 2003; Nel *et al.* 2004; Roura-Pascual *et al.* 2004; Welk 2004; De Meyer *et al.* 2007; Mgidi *et al.* 2007). First, a model describing the environmental niche of an invasive species can be produced using occurrence records and environmental predictor variables from the species' known distribution, followed by the projection of the niche model onto other regions where the species distribution is unknown (Peterson 2003; Nel *et al.* 2004; Roura-Pascual *et al.* 2004; De Meyer *et al.* 2007; Mgidi *et al.* 2007).

Data quality

The performance of ecological niche models and the accuracy of the potential range predictions depend on the quality of the data that is used to calibrate and evaluate the models (Fielding & Bell 1997; Welk 2004; Hernandez *et al.* 2006; Phillips *et al.* 2006). Several studies have shown that the number of occurrence records used to calibrate distribution models are an important determinant of model performance (Stockwell & Peterson 2002; McPherson *et al.* 2004; Hernandez *et al.* 2006; Phillips *et al.* 2006). Collections of species' distribution records are usually maintained in herbaria or museums and have also recently been made readily available as electronic databases (Graham *et al.* 2004; Hortal *et al.* 2007; Raes & ter Steege 2007; Hortal *et al.* 2008). Occurrence records available from these collections data might be incomplete due to incomplete sampling (Hernandez *et al.* 2006; Phillips *et al.* 2006; Hortal *et al.* 2007; Loiselle *et al.* 2008) or might suffer from geographical and environmental bias as a result of biased sampling. Biased sampling can be the result of the collectors mainly focussing their attention on specific areas e.g. areas that are easily accessible (Funk & Richardson 2002; Reddy & Davalos 2003; Kadmon *et al.* 2004; Loiselle *et al.* 2008), or areas with high species richness and diversity (Dennis & Thomas 2000; Loiselle *et al.* 2008). When occurrence records are geographically biased the underlying environmental gradients in which a species can persist will most likely also not be fully sampled, which could lead to environmental bias (Raes & ter Steege 2007; Hortal *et al.* 2008). This would occur if the areas that were not sampled contain portions of the environmental gradients in which the species can persist that were not represented with the records obtained from the sampled areas (Welk 2004; Guisan & Thuiller 2005; Hortal *et al.* 2008; Beaumont *et al.* 2009).

The suite and quality of environmental predictor variables also play an important role in determining the success of distribution models (Anderson *et al.* 2003; Pearson & Dawson 2003; Rouget *et al.* 2004; Thuiller *et al.* 2005; Araujo & Guisan 2006; Phillips *et al.* 2006). The spatial scale at which the predictions will be applied should determine the type of predictor variables incorporated. That is, at very broad scales climate conditions are often important driving forces of species' distributions (Pearson *et al.* 2002; Pearson & Dawson 2003; Guisan & Thuiller 2005; Soberon & Peterson 2005; Phillips *et al.* 2006) but at finer scales topography and biotic interactions becomes increasingly important (Welk *et al.* 2002; Pearson & Dawson 2003; Guisan & Thuiller 2005; Soberon & Peterson 2005). Model accuracy will be improved if the occurrence records and predictor variables incorporated into the model are of similar resolution (Guisan & Thuiller 2005; Phillips *et al.* 2006).

A limitation when modelling invasive alien species is that the majority of occurrence records from a species known distribution consist of presence-only data (Welk 2004; Soberon & Peterson 2005; Elith *et al.* 2006). Even when absence data are available they are usually unreliable as the species may have been recorded as absent merely because insufficient time had elapsed to allow for invasion and not necessarily due to an unsuitable habitat (Welk 2004; Wilson *et al.* 2007; Peterson *et al.* 2008). Models for invasive organisms can be calibrated with either native range records (Roura-Pascual *et al.* 2004; De Meyer *et al.* 2007; De Meyer *et al.* 2009), introduced range records (Robertson *et al.* 2001, 2003; Nel *et al.* 2004; Rouget *et al.* 2004; Schussman *et al.* 2006; Roura-Pascual *et al.* 2008) or a combination of native and introduced range records (Broennimann & Guisan 2008; Steiner *et al.* 2008). Obtaining occurrence

records that represent an invasive species' entire known distribution and the associated environmental requirements in its native range can prove to be a difficult and time-consuming task (Mau-Crimmins *et al.* 2006). Occurrence records available from collections data might be incomplete, for instance if a species has a native range that spans several countries and sampling effort was better in some parts of the range than others (Raes & ter Steege 2007), or if some records maintained in the museums or herbaria have not yet been added to the electronic databases. If parts of the native range of a species cannot be sampled or are poorly sampled then the potential distribution of the species in the introduced range is likely to be underestimated (Welk 2004; Guisan & Thuiller 2005; Hortal *et al.* 2008). The issue of genetic variation in the source populations relative to the introduced populations could result in the converse being true. If only a fraction of the genetic variability of an invasive species' native population (i.e. specific ecotype) has been introduced into a new region, the potential distribution of the species will be over-predicted if all the native range occurrence records are used to calibrate the niche models (Mau-Crimmins *et al.* 2006).

When calibrating models with occurrence records obtained from the introduced range, uncertainties in terms of the equilibrium status and introduction history could influence data quality and hence model performance (Kadmon *et al.* 2003, 2004; Welk 2004; Mau-Crimmins *et al.* 2006; Loo *et al.* 2007; Wilson *et al.* 2007; Roura-Pascual *et al.* 2008). A species that has recently been introduced to a new region is unlikely to be at equilibrium with the environment as insufficient time will have elapsed to allow it to disperse to suitable areas (Welk 2004; Wilson *et al.* 2007; Peterson *et al.* 2008). The occurrence records of this species would be unlikely to capture its full environmental

requirements and therefore underestimate the species' potential distribution. Similarly in terms of introduction history, the set of occurrence records for a species that has recently been introduced to a single locality would likely underestimate the potential distribution of that species to a greater extent than if the species had been introduced to several different localities simultaneously.

Model evaluation

The final step in the modelling process is to evaluate the models and determine how successful the model was in predicting the invasive species' potential distribution in the introduced range (Fielding & Bell 1997). This procedure relies on introduced range records being available for a species, since it is assumed that these records can give an indication of the species' realized distribution in the introduced range (Mau-Crimmins *et al.* 2006; Loo *et al.* 2007). A limitation of this procedure is that models predict the potential distributions of the species while the evaluation dataset evaluates the realized distribution of these species (Kadmon *et al.* 2003). This geographical distribution may be influenced by biological interactions (e.g. competitions and natural enemies) and dispersal limitation (Pearson & Dawson 2003; Kadmon *et al.* 2003; Soberon & Peterson 2005). Records from the introduced range of a species are not always a true reflection of a species' spreading potential in that region. These records might be an under-representation of a species' distribution if a species is not at equilibrium because insufficient time has elapsed since its introduction into a new region to spread to all the suitable areas (Mau-Crimmins *et al.* 2006; Loo *et al.* 2007; Wilson *et al.* 2007; Broennimann & Guisan 2008).

Ongoing advances are being made in terms of the methodology and applications of ecological niche modelling, and although studies have also recently started to investigate the influence of data quality on model performance (Araujo & Guisan 2006; Maurer-Crimmins *et al.* 2006; Segurado *et al.* 2006; Peterson & Nakazawa 2008) very few have focused on the issues that influence modelling of invasive species distributions.

Thesis outline

This thesis comprises of four chapters, the first chapter being a general introduction to ecological niche models and specifically the influence of data quality on performance of ecological niche models for invasive species. In chapter two I investigate the impact of geographical bias in native range occurrence records on the performance of ecological niche models for invasive alien plant species. The aim of chapter three was to compare the performance of ecological niche models when calibrated with (1) native range records (2) introduced range records and (3) a combination of both, for invasive alien plant species with different simulated introduction histories. In chapter four I discuss the general conclusions of this study, the problems and limitations encountered and recommendations for future research.

References:

ANDERSON, R.P., LEW, D. & PETERSON, A.T. 2003. Evaluating predictive models of species' distributions: criteria for selecting optimal models. *Ecological Modelling* **162**: 211-232.

ARAUJO, M.B., CABEZA, M., THUILLER, W., HANNAH, L. & WILLIAMS, P.H. 2004. Would climate change drive species out of reserves? An assessment of existing reserve-selection methods. *Global Change Biology* **10**: 1618-1626.

ARAUJO, M.B. & GUISAN, A. 2006. Five (or so) challenges for species distribution modeling. *Journal of Biogeography* **33**: 1677-1688.

BEAUMONT, L.J., GALLAGHER, R.V., THUILLER, W., DOWNEY, P.O., LEISHMAN, M.R. & HUGHES, L. 2009. Different climatic envelopes among invasive populations may lead to underestimations of current and future biological invasions. *Diversity and Distributions* **15**: 409-420.

BROENNIMANN, O., TREIER, U.A., MULLER-SCHARER, H., THUILLER, W., PETERSON, A.T. & GUISAN, A. 2007. Evidence of climatic niche shift during biological invasion. *Ecology Letters* **10**: 701-709.

BROENNIMANN, O. & GUISAN, A. 2008. Predicting current and future biological invasions: both native and invaded ranges matter. *Biological Letters* **4**: 585-589.

DE MEYER, M., ROBERTSON, M.P., PETERSON, A.T. & MANSELL, M.W. 2007. Ecological modelling and potential geographic distributions of Mediterranean fruit fly (*Ceratitidis capitata*) and Natal fruit fly (*Ceratitidis rosa*). *Journal of Biogeography* **35**: 270-281.

DE MEYER, M., ROBERTSON, M.P., MANSELL, M.W., EKESI, S., TSURUTA, K., MWAIK, W., VAYSIERES, J-F. & PETERSON, A.T. 2009. Ecological niche and potential geographic distribution of the invasive fruit fly *Bactrocera invadens* (Diptera, Tephritidae). *Bulletin of Entomological Research*. In press.

DENNIS, R.L.H. & THOMAS, C.D. 2000. Bias in butterfly distribution maps: the influence of hot spots and recorder's home range. *Journal of Insect Conservation* **4**: 73-77.

ELITH, J., GRAHAM, C.H., ANDERSON, R.P., DUDI, M., FERRIER, S., GUIBAN, A., HIJMAN, R.J., HUETTMAN, F., LEATHWICK, J.R., LEHMAN, A., LI, J., LOHMANN, L.G., LOISELLE, B.A., MANION, G., MORITZ, C., NAKAMURA, M., NAKAZAWA, Y., OVERTON, J.M., TOWNSEND-PETERSON, A., PHILLIPS, S., RICHARDSON, K., SCACHETTI-PEREIRA, R., SCHAPIRE, R.E., SOBERON, J., WILLIAMS, S., WISZ, M.S. & ZIMMERMANN, N.E. 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* **29**: 129-151.

ENGLER, R., GUIBAN, A. & RECHSTEINER, L. 2004. An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. *Journal of Applied Ecology* **41**: 263-274.

FIELDING, A.H. & BELL, J.F. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation* **24**: 38-49.

FUNK, V.A. & RICHARDSON, K.S. 2002. Systematic data in biodiversity studies: Use it or lose it. *Systematic Biology* **51**: 303-316.

GRAHAM, C.H., FERRIER, S., HUETTMAN, F. MORITZ, C. & PETERSON, A.T. 2004. New developments in museum based informatics and applications in biodiversity analysis. *Trends in Ecology and Evolution* **19**: 497-503.

GUIBAN, A. & ZIMMERMAN, N.E. 2000. Predictive habitat distribution models in ecology. *Ecological Modelling* **135**: 147-186.

GUIBAN, A. & THUILLER, W. 2005. Predicting species distribution: offering more than simple habitat models. *Ecology Letters* **8**: 993-1009.

HERNANDEZ, P.A., GRAHAM, C.H., MASTER, L.L. & ALBERT, D.L. 2006. The effect of sample size and species characteristics on performance of different species

distribution modeling methods. *Ecography* **29**: 773-785.

HORTAL, J., LOBO, J.M. & JIMINEZ-VALVERDE, A. 2007. Limitations of biodiversity databases: Case study on seed-plant diversity in Tenerife, Canary Islands. *Conservation Biology* **21**: 853-863.

HORTAL, J., JIMINEZ-VALVERDE, A., GOMEZ, J.F., LOBO, J.M. & BASELGA, A. 2008. Historical bias in biodiversity inventories affects the observed environmental niche of the species. *Oikos* **117**: 847-858.

KADMON, R., FARBER, O. & DANIN, A. 2003. A systematic analysis of factors affecting the performance of climatic envelope models. *Ecological Applications* **13**: 853-867.

KADMON, R., FARBER, O. & DANIN, A. 2004. Effect of roadside bias on the accuracy of predictive maps produced by bioclimatic models. *Ecological Applications* **14**: 401-413.

LOISELLE, B.A., JORGENSEN, P.M., CONSIGLIO, T., JIMENEZ, I., BLAKE, J.G., LOHMANN, L.G. & MONTIEL, O.M. 2008. Predicting species distributions from herbarium collections: does climate bias in collection sampling influence model outcomes? *Journal of Biogeography* **35**: 105-116.

LOO, A.E., MACNALLY, R. & LAKE, P.S. 2007. Forecasting New Zealand mudsnail invasion range: model comparison using native and invaded ranges. *Ecological Applications* **17**: 181-19.

MACK, R.N., SIMBERLOFF, D., LONSDALE, W.M., EVANS, H., CLOUT, M. & BAZZAZ, F. 2000. Biotic invasions: Causes, Epidemiology, Global Consequences and Control. *Issues in Ecology* **5**: 1-20.

MAU-CRIMMINS, T., SCHUSSMAN, H.R. & GEIGER, H. 2006. Can the invaded range of a species be predicted sufficiently using only native-range data? Lehmann lovegrass (*Eragrostis lehmanniana*) in the southwestern United States. *Ecological Modelling* **19**: 736-746.

McPHERSON, J.M., JETZ, W. & ROGERS, D.J. 2004. The effects of species' range sizes on the accuracy of distribution models: ecological phenomenon or statistical artifact? *Journal of Applied Ecology* **41**: 811-823.

MCPHERSON, J.M. & JETZ, W. 2007. Effects of species' ecology on the accuracy of distribution models. *Ecography* **30**: 135-151.

MGIDI, T., LE MAITRE, D.C., SCHONEGEVEL, L., NEL, J.L., ROUGET, M. & RICHARDSON, D.M. 2007. Alien plant invasions – incorporating emerging invaders in regional prioritization: A pragmatic approach for Southern Africa. *Journal of Environmental Management* **84**: 173-187.

NEL, J.L., RICHARDSON, D.M., ROUGET, M., MIGIDI, T.N., MDZEKE, N., LE MAITRE, D.C., VAN WILGEN, B.W., SCHONEGEVEL, L., HENDERSON, L. & NESER, S. 2004. A proposed classification of invasive alien plant species in South Africa: towards prioritizing species and areas for management action. *South African Journal of Science* **100**: 53-64.

PEARMAN, P.B., GUISAN, A., BROENNIMANN, O. & RANDIN, C.F. 2008. Niche dynamics in space and time. *Trends in Ecology and Evolution* **23**: 149-158.

PEARSON, R.G., DAWSON, T.P., BERRY, P.M. & HARRISON, P.A. 2002. SPECIES: A spatial evaluation of climate impact on the envelope of species. *Ecological Modelling* **154**: 289-300.

PEARSON, R.G. & DAWSON, T.P. 2003. Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Global Ecology & Biogeography* **12**: 361-371.

PETERSON, A.T. 2001. Predicting species' geographic distributions based on ecological niche modelling. *The Condor* **103**: 599-605.

PETERSON, A.T. 2003. Predicting the geography of species' invasions via ecological niche modeling. *The Quarterly Review of Biology* **78**: 419-433.

PETERSON, A.T. & NAKAZAWA, Y. 2008. Environmental data sets matter in ecological niche modelling: an example with *Solenopsis invicta* and *Solenopsis richteri*. *Global Ecology and Biogeography* **17**: 135-144.

PETERSON, A.T., PAPES, M. & SOBERON, J. 2008. Rethinking receiver operating characteristic analysis applications in ecological niche modeling. *Ecological Modelling* **213**: 63-72.

PHILLIPS, S.J., ANDERSON, R.P. & SCHAPIRE, R.E. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* **190**: 231-259.

PIMENTEL, D., MCNAIR, S., JANECKA, J., WIGHTMAN, J., SIMMONDS, C., O'CONNELL, C., WONG, E., RUSSEL, L., ZERN, J., AQUINO & TSOMONDO. 2001. Economic and environmental threats of alien plant, animal and microbe invasions. *Agriculture, Ecosystems and Environment* **84**: 1-20.

RAES, N. & TER STEEGE, H. 2007. A null-model for significance testing of presence only species distribution models. *Ecography* **30**: 727-736.

REDDY, S. & DAVALOS, L.M. 2003. Geographical sampling bias and its implications for conservation priorities in Africa. *Journal of Biogeography* **30**: 1719-1727.

RICHARDSON, D.M. & VAN WILGEN, B. 2004. Invasive alien plants in South Africa: how well do we understand the ecological impacts? *South African Journal of Science* **100**: 45-52.

ROBERTSON, M.P., CAITHNESS, N. & VILLET, M.H. 2001. A PCA-based modelling technique for predicting environmental suitability for organisms from presence records. *Diversity and Distributions* **7**: 15-27.

ROBERTSON, M.P., PETER, C.I., VILLET, M.H. & RIPLEY, B.S. 2003. Comparing models for predicting species potential distribution: a case study using correlative and mechanistic predictive modelling techniques. *Ecological Modelling* **164**: 153-167.

ROUGET, M., RICHARDSON, D.M., NEL, J.L., LE MAITRE, D.C., EGOH, B. & MGIDI, T. 2004. Mapping the potential ranges of major plant invaders in South Africa, Lesotho and Swaziland using climatic suitability. *Diversity and Distributions* **10**: 475-484.

ROURA-PASCUAL, N., SUAREZ, A.V., GOMEZ, C., PONS, P., TOUYAMA, Y., WILD, A.L. & PETERSON, A.T. 2004. Geographical potential of Argentine ants (*Linepithema humile* Mayr) in the face of global climate change. *Proceedings of the Royal Society of London B Biological Sciences* **271**: 2517-2534

ROURA-PASCUAL, N., BROTONS, L., PETERSON, A.T. & THUILLER, W. 2008. Consensual predictions of potential distributional areas for invasive species: a case study of Argentine ants in the Iberian Peninsula. *Biological Invasions* **11**: 1017-1031.

SCHUSSMAN, H., GEIGER, E., MAU-CRIMMINS, T. & WARD, J. 2006. Spread of current and potential distribution of an alien grass, *Eragrostis lehmanniana* Nees, in the southwestern USA: comparing historical data and ecological niche models. *Diversity and Distributions* **12**: 582-592.

- SEGURADO, P., ARAUJO, M.B. & KUNIN, W.E. 2006. Consequences of spatial autocorrelation for niche-based models. *Journal of Applied Ecology* **43**: 433-444.
- SOBERON, J. & PETERSON, A.T. 2005. Interpretation of models of fundamental ecological niches and species' distributional areas. *Biodiversity Informatics* **2**: 1-10.
- STEINER, F.M., SCHLICK-STEINER, B.C., VANDERWAL, J., REUTHER, K.D., CHRISTIAN, E., STAUFFER, C., SUAREZ, A.V., WILLIAMS, S.E. & CROZIER, R.H. 2008. Combined modeling of distribution and niche in invasion biology: a case study of two invasive *Tetramorium* ant species. *Diversity and Distributions* **14**: 538-545.
- STOCKWELL, D.R.B. & PETERSON, A.T. 2002. Effects of sample size on accuracy of species distribution models. *Ecological Modelling* **148**: 1-13.
- THULLER, W., RICHARDSON, D.M., PYSEK, P., MIDGLEY, G.F., HUGHES, G.O. & ROUGET, M. 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.
- VAN WILGEN, B.W., REYERS, B., LE MAITRE, D.C., RICHARDSON, D.M. & SCHONEGEVEL, L. 2008. A biome-scale assessment of the impact of invasive alien plants on ecosystem services in South Africa. *Journal of Environmental Management* **89**: 336-349.
- VITOUSEK, P.M., D'ANTONIO, C.M., LOOPE, L.L., REJMANEK, M. & WESTBROOKS, R. 1997. Introduced species: a significant component of human-caused global change. *New Zealand Journal of Ecology* **21**: 1-16.
- WELK, E., SCHUBERT, K. & HOFFMANN, M.H. 2002. Present and potential distribution of invasive garlic mustard (*Alliaria petiolata*) in North America. *Diversity and Distributions* **8**: 219-233.

WELK, E. 2004. Constraints in range predictions of invasive plant species due to non equilibrium distribution pattern: Purple loosestrife (*Lythrum salicaria*) in North America. *Ecological Modelling* **179**: 551-567.

WILSON, J.R.U., RICHARDSON, D.M., ROUGET, M., PROCHE, S., AMIS, M.A., HENDERSON, L. & THUILLER, W. 2007. Residence time and potential range: crucial considerations in modelling plant invasions. *Diversity and Distributions* **13**: 11-22.

YURKONIS, K.A., MEINERS, S.J. & WACHHOLDER, B.E. 2005. Invasion impacts diversity through altered community dynamics. *Journal of Ecology* **93**: 1053-1061.



Chapter 2

Running title: Geographical bias

Predicting invasive alien plant distributions: how geographical bias in occurrence records influences model performance

René Wolmarans, Mark P. Robertson and Berndt J. van Rensburg

Department of Zoology and Entomology, Centre for Invasion Biology, University of Pretoria, Pretoria 0002, South Africa.

ABSTRACT

Aim

To investigate the impact of geographical bias on performance of ecological niche models for invasive plant species.

Location

South Africa and Australia.

Methods

We selected 10 Australian plants invasive in South Africa and nine South African plants invasive in Australia. Geographical bias was simulated in occurrence records obtained from the native range of a species to represent two scenarios. For the first scenario (A, worst-case) a proportion of records were excluded from a specific region of a species' range and for the second scenario (B, less extreme) only some records were excluded from that specific region of the range. Introduced range predictions were produced with the Maxent modelling algorithm where models were calibrated with these biased occurrence records datasets and 19 bioclimatic variables. Models were evaluated with

independent test data obtained from the introduced range of the species. Geographical bias was quantified as the proportional difference in geographical position between the occurrence records from a control and biased dataset and environmental bias was expressed as either the difference in marginality or tolerance between this treatment. Model performance was compared between niches predicted to be occupied with occurrence records from a biased dataset and a control dataset.

Results

We found considerable variation in the relationship between geographical and environmental bias and no significant correlation. Environmental bias expressed as the difference in marginality differed significantly across treatments. Model performance did not differ significantly among treatments. Regions predicted as suitable for most of the species were very similar when compared between a biased and control dataset, with only a few exceptions.

Main conclusions

No significant relationship was found between geographical bias and environmental bias. The geographical bias simulated in this study was sufficient to result in significant environmental bias across treatments, but despite this we did not find a significant effect on model performance.

Keywords

Geographical bias, environmental bias, ecological niche modelling, invasive alien plants, Maxent, model performance, marginality, tolerance.

INTRODUCTION

Natural ecosystems, overall biodiversity, human health and natural economies are impacted upon at global, regional and local scales as a result of an ever-increasing number of destructive alien species (Vitousek *et al.*, 1997; Pimentel *et al.*, 2001; Richardson & van Wilgen, 2004; Yurkonis *et al.*, 2005). Due to the fact that it is difficult, if not impossible to eradicate alien plant species once they have become naturalized and started to spread (Peterson, 2003; Thuiller *et al.*, 2005), conservation biologists and land managers have to make informed decisions on how to manage them (Rouget *et al.*, 2004). If not managed properly, it can be assumed that the associated problems will be amplified in the future (Welk *et al.*, 2002; Peterson, 2003). Consequently a pro-active management strategy for invasive plant species would be the implementation of early detection and prevention systems (Thuiller *et al.*, 2005; Mau-Crimmins *et al.*, 2006; Richardson & Thuiller, 2007). These systems can be implemented with the aid of statistical modeling techniques.

Correlative models that quantify the relationship between occurrence records and environmental data have been applied to a number of problems in biology (see Guisan & Zimmermann, 2000; Pearson & Dawson, 2003; Welk, 2004; Guisan & Thuiller, 2005; Zimmermann *et al.*, 2007). Not only have models been particularly useful for predicting species distributions and ecological niches of invasive organisms (Robertson *et al.*, 2001; 2003; Rouget *et al.*, 2004; Mau-Crimmins *et al.*, 2006; De Meyer *et al.*, 2007; Steiner *et al.*, 2008), they have also provided valuable insight into the potential spread of invasive organisms (Peterson, 2003; Welk, 2004; Rouget *et al.*, 2004; Richardson & Thuiller, 2007) and have predicted environmental suitability of regions that have not yet been

invaded (Mgidi *et al.*, 2007; De Meyer *et al.*, 2007). Predictions such as these can be used to identify areas where management and monitoring efforts should be focused.

Most models have attempted to predict the distribution of invasive species using the relationship between occurrence records from the introduced range (adventive range) of the species and environmental data (Robertson *et al.*, 2001; 2003; Rouget *et al.*, 2004). This relationship assumes that the best indicator of species environmental requirements is its current distribution (Robertson *et al.*, 2003; Thuiller *et al.*, 2005; Pearson *et al.*, 2007). The implicit assumption of this approach is that these species are in equilibrium with their environment. Hence, enough time has passed since their introduction to allow for all the environmentally suitable sites to be occupied. This may not be the case for all species (Wilson *et al.*, 2007); for example sleeper weeds that have a long lag-phase between introduction and naturalization (Richardson, 2004; Guisan & Thuiller, 2005). In cases where a species has only recently started to invade a new region or for which invasion risk needs to be assessed (De Meyer *et al.*, 2007; Mgidi *et al.*, 2007), it may be preferable or necessary to make predictions using native range records (in an attempt to quantify the niche of an organism). Recent studies have indicated that niches can differ between native and introduced ranges of invasive species (Mau-Crimmins *et al.*, 2006; Broennimann *et al.*, 2007; Fitzpatrick *et al.*, 2007; Beaumont *et al.*, 2009), prompting the use of both native and introduced range occurrence records in models (Mau-Crimmins *et al.*, 2006; Steiner *et al.*, 2008; Beaumont *et al.*, 2009).

Models that are able to quantify the fundamental niche of an organism are likely to be most successful at predicting potential spread of an invasive species. The fundamental niche describes all the environmental conditions in which a species would

potentially be able to grow and persist (Pearman *et al.*, 2008). In reality, however, a species' geographic distribution is influenced by a complex interaction of several main factors (Soberon & Peterson, 2005). Climate conditions might impose physiological limits on a species (abiotic factors) and interactions with other species could either have a positive or negative effect on their distribution (biotic factors). Due to the configuration of the environment, not all areas are always accessible to a species and therefore inflict dispersal limitations. These interactions result in species only occupying some parts of their fundamental niche, known as the realized niche (Pearson & Dawson, 2003; Soberon & Peterson, 2005). As a result, the occurrence records used to make predictions already include the effect of these factors and their interaction and therefore most likely only provide information on a species' realized niche (Pearman *et al.*, 2008). The extent to which the realized niche can be described by the correlative model and extent to which the realized and fundamental niches overlap may influence the success of the model (Soberon & Peterson, 2005).

Geographical bias often exists in record collections due to biased sampling. Biased sampling can be the result of the collectors mainly focussing their attention on specific areas e.g. areas that are easily accessible (near roads or rivers) (Funk & Richardson, 2002; Reddy & Davalos, 2003; Kadmon *et al.*, 2004; Loiselle *et al.*, 2008), or areas with high species richness and diversity (Dennis & Thomas, 2000; Loiselle *et al.*, 2008). When occurrence records are geographically biased, the underlying environmental gradients in which a species can persist will most likely also not be fully sampled, which could result in environmental bias (Raes & ter Steege, 2007; Hortal *et al.*, 2008). The records available from herbaria or electronic databases may only sample a

part of a species geographic range or in some cases sampling effort in certain parts of the range, may be better than others (Raes & ter Steege, 2007). This is likely to be the case when a species has a native range that spans several countries. In addition, acquisition of all the records can be time consuming and expensive, and also difficult if sources are in different languages (Mau-Crimmins *et al.*, 2006). If parts of the native range of a species cannot be sampled or are poorly sampled then the potential distribution of the species in the introduced range is likely to be underestimated. This would occur if the areas that were not sampled contain portions of the environmental gradients in which the species can persist that were not represented with the records obtained from the sampled areas (Welk, 2004; Guisan & Thuiller, 2005; Hortal *et al.*, 2008; Beaumont *et al.*, 2009).

The aim of this study was to examine the impact of geographical bias on the performance of ecological niche models of invasive plant species. We examine models that were calibrated with native range occurrence records and utilized to predict the potential distribution in the introduced range. Different treatments consisting of biased datasets were created with the native range occurrence records to simulate different geographical bias scenarios. Specifically, we wanted to test the influence of the resultant underlying environmental bias on model performance and in so doing we addressed the following questions: 1) What is the relationship between geographical bias and environmental bias in the native range of a species? 2) Does environmental bias differ across treatments that simulate geographical bias? 3) Is there sufficient environmental bias to result in significant differences in model performance across the treatments?

MATERIALS AND METHODS

Species selection and distribution records

We selected South Africa and Australia as the study region since both countries have exchanged several plant species that have become naturalized and in many cases invasive (Randall, 2002; Henderson, 2007). Both countries have good sources of distribution records that are readily available through electronic databases. We compiled a list of ten Australian plants that are invasive alien species in South Africa and a list of nine South African species that are invasive alien species in Australia (Table 1). For the Australian species invading South Africa, native range occurrence records were obtained from the Australian Virtual Herbarium public access database (AVH; <http://www.anbg.gov.au/avh/>). This is an online botanical information resource, developed by the Council of Heads of Australian Herbaria, representing a collection of plant specimen data for all the major Australian herbaria that have actively maintained their collections for many decades. Additional occurrence records were requested from the Queensland herbarium (Paul Robins *pers. comm.*) and the Western Australia Herbarium (FloraBase database). Introduced range occurrence records were obtained from the Southern African Plant Invaders Atlas (SAPIA; <http://www.agis.agric.za/>), which contains the most comprehensive set of distribution records for alien invasive plant species in southern Africa (Henderson, 1998; 2006b; 2007).

For the South African species invading Australia, native range occurrence records were obtained from South Africa's National Herbarium Computerised Information System (PRECIS) database. Introduced range occurrence records were derived from the AVH, Queensland herbarium and Western Australia herbarium.

In both cases, all native range records were assumed to be representative of the entire native range of the species concerned. The precision of the distribution records varied among databases. As a result, the distribution records were sufficiently precise to calibrate and evaluate models using a 15 minute grid. The coordinates of the occurrence records were used to assign each record to a grid cell in the map region. Only one record per grid cell was used when several occurrence records were present in a grid cell (Stockwell & Peterson, 2002a; Hernandez *et al.*, 2006).

Environmental predictors

We selected the 19 bioclimatic variables available from the Worldclim database (<http://www.worldclim.org>; Hijmans *et al.*, 2005; Table 2). Bioclimatic variables derived from monthly temperature and precipitation data are commonly used in biogeographical modeling (De Meyer *et al.*, 2007; Loiselle *et al.*, 2008). These variables represent annual trends (e.g. annual mean temperature), seasonality (e.g. annual range in precipitation), and extreme or limiting environmental factors (e.g. precipitation of the wettest month) (Hijmans *et al.*, 2005). All environmental predictors were resampled to 15 minute grids using ArcGIS 9.2.

Ecological niche modeling

Maxent is a relatively new statistical modeling technique that has been applied to model the potential distribution of species and estimating niche occupation (Phillips *et al.*, 2006; Pearson *et al.*, 2007; Peterson *et al.*, 2008; Phillips & Dudik, 2008). Maxent estimates a species distribution by finding the probability distribution of maximum entropy (i.e.

closest to uniform) subject to the constraint that the mean environmental conditions predicted by the model should be close to the empirical average of the conditions at the known presence localities (Phillips *et al.*, 2006; Phillips & Dudik, 2008). An advantage of this approach is that it uses presence records to contrast the distribution of the occurrence sites with the environmental conditions of the entire area of interest, or a random background sample of sites, also referred to as pseudo-absences (Gibson *et al.*, 2007, Pearson *et al.*, 2007). For most invasive species, occurrence data have been recorded without a planned sampling design. Therefore the majority of occurrence records consist of presence-only data obtained from herbarium collections (Soberon & Peterson, 2005; Elith *et al.*, 2006). Even when absence data are available they are usually unreliable as the species may have been recorded as absent merely because insufficient time had elapsed to allow for invasion (Wilson *et al.*, 2007; Peterson *et al.*, 2008). Maxent probability responses can either be raw, logistic or cumulative, and for this study we selected the logistic probability response following advice given by Phillips & Dudik (2008). This is a continuous variable ranging from 0-1, where high values indicate higher suitability for a species in a particular grid cell (Phillips *et al.*, 2006; Gibson *et al.*, 2007).

Simulating biased sampling

We start by defining the full range of a species as all of the native range distribution records available for that species. To simulate geographical bias we defined two scenarios that could occur. In the first scenario (A) a proportion of the distribution records were excluded from a specific geographic region of the full range of a species (Fig. 2a). This represents an extreme case where no records are available for the species

from a specific part of the range. This could occur for species with ranges that include a politically unstable country in which no records have been collected for that species. The second scenario (B) is less extreme than the first because only some records were excluded from that region (Fig. 2b), instead of excluding all records from a specific region. This simulates a case where sampling effort is poor in a particular part of the range. For each species we generated three datasets for each scenario. For scenario A we excluded 10%, 20% and 40% of the records from the North or West of the range (whichever showed the greatest change in annual precipitation (variable BIO 12)) to produce three datasets A10, A20 and A40. For scenario B, an initial sample of 50% of the records from either the East or South of the range was used and then a further 40% of the remaining records (from the West or North respectively) was randomly sampled and added to the initial sample (B10). This was repeated by adding 30% and 10% respectively to the initial sample of 50% to generate the remaining two datasets (B20 and B40). We also generated datasets by randomly excluding 10%, 20% and 40% from the full range (Fig. 2c) to serve as controls (R10, R20, and R40). Datasets e.g. A10, B10 and R10 contained equal numbers of occurrence records.

Models

All models were calibrated with samples of records taken from the native range (the calibration set) (Fig. 1). Environmental variable values corresponding to each grid cell of the calibration set were extracted for each species and used as input data for Maxent (v. 3.1.0). In Maxent, predictor variables are transformed into feature types (linear, quadratic, product, threshold or discrete) and may be the mean of variables, their square,

and product with other variables, thresholds or binarizations of categorical variables. The auto features option was selected allowing the set of feature types used to depend on the number of presence records for that particular species. Algorithm parameters were set to a maximum number of 500 iterations, a regularization multiplier of one, and a convergence threshold of 0.00001.

We wanted to ensure that native range predictions were accurate and acceptable before making projections to the introduced range. Five native range predictions were made, in each case all the native range occurrence records obtained were partitioned into a calibration set (training set) and a testing set (validation set) using a k-fold method (Pearson *et al.*, 2006; Phillips *et al.*, 2006). A random selection of 70% of the occurrence records comprised the calibration set and the remaining 30% comprised the testing set. For the testing set, pseudo-absences were generated by randomly selecting grid cells from the map region that did not contain presence records. For each species the number of pseudo-absence records was equal to the number of presence records in the testing set (Jiminez-Valverde & Lobo, 2007; McPherson & Jetz, 2007; Roura-Pascual *et al.*, 2008). Models projected on to the introduced range were evaluated with a test set that comprised of occurrence records available for each species from the introduced range and pseudo-absence records generated in a similar fashion as described above.

Data quality evaluation

The quality of all model predictions depend on the data used to calibrate and evaluate the model (Fielding & Bell, 1997). In particular, models for alien invasive organisms are influenced by several factors related to the invasion of the species, e.g. whether a species

is at equilibrium with its environment or if niche shift has occurred between the native and introduced ranges (Mau-Crimmins *et al.*, 2006; Broennimann *et al.*, 2007; Fitzpatrick *et al.*, 2007; Beaumont *et al.*, 2009). To determine if niche shift occurred between ranges quantified with different occurrence records datasets, we used Ecological-Niche Factor Analysis (ENFA: Hirzel *et al.*, 2002) implemented in the *adehabitat* package in R (R Development Core Team, 2004) to compare the ecological niche occupied by a species in a specific range (Steiner *et al.*, 2008). ENFA compares the prevailing climatic conditions in a specific range to background data that represents the global environmental conditions (Steiner *et al.*, 2008). We used the same 19 bioclimatic variables derived from the Worldclim database to serve as global environmental background data, to allow for direct comparisons among species (Steiner *et al.* 2008). ENFA characterises two components of a species ecological niche namely marginality and tolerance. Marginality measures the niche position by comparing the mean of a specific range to the background mean, and tolerance measures the niche width by comparing the variance in a specific range to the background environmental variation. A difference in the marginality values indicate a shift in the mean environmental niche occupied by a species and a difference in the tolerance values correspond to a shift in the niche width (Steiner *et al.*, 2008). To determine whether niche shift occurred between the native and introduced range of a species, we compared marginality values between the native and introduced range with Student's t-tests, and contrasted tolerance values with Bartlett's homogeneity of variance tests (Steiner *et al.*, 2008).

In addition, we wanted to visually illustrate whether each species occupied the same environmental space in the native and introduced range, since models were

calibrated with native occurrence records and evaluated with records obtained from the introduced range. We performed a principal components analysis (PCA) on the environmental data extracted from native and introduced datasets for each species (Robertson *et al.*, 2001; Mau-Crimmins *et al.*, 2006), and plotted the native records used for calibration and the introduced records used for evaluation in the resulting environmental space. This allowed us to establish whether the records used for evaluation and for calibration overlapped in environmental space.

Model performance was examined for models that predicted the native range of a species and models that were projected to the introduced range. This allowed us to determine how well the models could predict a species known distribution in the native range and could therefore verify that models were calibrated with a reliable set of predictor variables.

Model evaluation

Model performance is frequently assessed using the Area Under Curve (AUC) from receiver operating characteristic (ROC) plots (Fielding & Bell, 1997; Lobo *et al.*, 2008; Peterson *et al.*, 2008). The ROC curve is a threshold independent measure that plots sensitivity values (true positive fraction) as a function of 1- specificity values (false positive fraction) across a series of thresholds (Fielding & Bell, 1997). At every possible threshold the continuous probability score is converted to a binary value. If the predicted occurrence is above the particular threshold then it is considered to be a presence, if below the threshold then it is considered an absence. These binary transformed probabilities are then compared to the actual occurrence records from the test set,

allowing for the production of a confusion matrix (Fielding & Bell, 1997). A confusion matrix calculates the number of true presences (sensitivity), false positives (commission error), false negatives (omission error) and true absences (specificity) at each particular threshold (Fielding & Bell, 1997). All AUC values range between 0-1, where values of 0.5 indicate that predictions are no better than random, poor when values range between 0.5-0.7, and useful if values fall between 0.7-0.9. Predictions that obtain values greater than 0.9, are considered to be good or even excellent (Fielding & Bell, 1997; Broennimann *et al.*, 2007). Recently the use of AUC statistics for model evaluation has been criticized (Lobo *et al.*, 2008). A modified AUC approach was proposed by Peterson *et al.* (2008) to overcome some of these criticisms. This method plots the true positive fraction as a function of the proportion of the overall area predicted present. This evaluation method eliminates the reliance on commission error where areas might be classified as unsuitable based on uncertain absences, i.e. pseudo-absences (Peterson *et al.*, 2008). These ROC results are expressed as ratios of the area under the observed curve to the area under the random line. We used both the conventional and modified version of ROC plots. Additionally, we used the True Skill Statistic (TSS) instead of the popular Kappa statistic, as Kappa is influenced by prevalence (Allouche *et al.*, 2006). TSS values range from -1 to +1, where +1 indicates a perfect fit and values of zero or less indicate a performance no better than random (Allouche *et al.*, 2006). Since the TSS value was calculated for all possible thresholds ranging from 0-1 with set increments of 0.01, only the maximum TSS (mTSS) value for each species prediction was reported. All three model performance measures were calculated as described in the literature (Fielding &

Bell, 1997; Allouche *et al.*, 2006; Peterson *et al.*, 2008) as functions written in R (R Development Core Team, 2004).

Analysis

In order to describe the relationship between geographical and environmental bias we had to quantify these biases. We quantified environmental bias as the difference in either the marginality (niche position) or tolerance (niche width) values between a control (e.g. R10) and bias dataset (e.g. A10 or B10). The marginality and tolerance value for a dataset comprising biased records is likely to be smaller than a dataset comprising randomly selected records as a biased dataset will tend to sample less environmental variation. We quantified geographical bias as the proportional difference between a control and bias dataset using the x- or y-coordinates of a particular dataset, depending on the direction from which records were excluded in order to simulate bias. We show how the bias created in the occurrence records relates to the geographic and environmental distribution of the species (see Fig. S1 in Supplementary Material). To assess whether the amount of environmental bias differed on average across treatments we compared the difference in marginality or difference in tolerance values for all the species between a control and bias dataset (e.g. R10 and A10). Non-parametric comparisons were carried out with Kruskal-Wallis tests, where after the significant differences were identified with multiple comparisons carried out with the `npmc` function in R (R Development Core Team, 2004). We describe the relationship between environmental bias and model performance by determining the correlation between the difference in marginality or tolerance and difference in model performance between a control (e.g. R10) and bias

dataset (e.g. A10) across all treatments. We show the potential distribution predicted for two species in the introduced range with the different biased and corresponding control datasets.

RESULTS

Data quality evaluation

Comparisons of marginality values between the native and introduced range indicated that there was no significant difference in the position of the niche occupied between these two ranges for any of the species (see Table S1 in Supplementary Material). However, a comparison of the tolerance values between these two ranges indicated that the environmental variation between the native and introduced range did differ significantly for 12 of the 19 species (Table S1).

An assessment of the PCA analyses revealed that for 9 out of the 19 species (*Acacia cyclops*, *A. mearnsii*, *A. saligna*, *Arctotheca calendula*, *Moraea flaccida*, *Oxalis pes-caprae*, *Romulea rosea*, *Sparaxis bulbifera* and *Watsonia meriana*), the environmental space occupied by the species in the introduced range expanded in relation to the environmental space occupied in the native range (see Fig. S1b in Supplementary Material and Table S1). In addition, 5 out of the 19 species (*A. dealbata*, *A. longifolia*, *A. melanoxydon*, *Leptospermum laevigatum* and *Zantedeschia aethiopica*) showed a smaller environmental space occupied in the introduced range relative to the environmental space occupied in the native range (Fig. S1b; Table S1). For all species, native range models obtained AUC values (with conventional ROC analysis) that ranged between 0.94 and 1.00, indicating good/excellent model performance (Fig. S2). Models projected onto the

introduced range showed more variation in performance (Fig. S2). Models for nearly all the species achieved AUC values greater than 0.8, indicating that model performance in the introduced range was useful. However, *A. decurrens*, *A. dealbata* and *A. mearnsii* obtained values close to and below 0.7 (Fig. S2), indicating useful but poor performance in the introduced range.

The relationship between geographical bias and environmental bias

The environmental space occupied with all the native range records compared to the environmental space occupied when 40% of the records were excluded from the dataset (A40 treatment, biased records) was similar for some species (e.g. *A. decurrens* in Fig. 3b, *Hakea sericea*, *Paraserianthes lophantha*, *Ehrharta calycina* and *Polygala myrtifolia*), but for the majority of species the environmental space occupied with these two datasets were different (Fig. S1). There was considerable variation in the relationship between geographical bias and environmental bias across treatments (Fig. 4). For some species an increase in geographical bias, expressed as the proportional geographic difference between the bias dataset and the control, lead to an increase in the amount of environmental bias, expressed as the difference in marginality, but no clear trends were evident. The correlation between the proportional geographic difference and marginality difference was only significant for the B10 treatment (Table 3), the rest of the treatments showed no significant correlation (Table 3). None of the treatments showed a significant correlation when environmental bias was expressed as the difference in tolerance.

Differences in environmental bias across treatments

Environmental bias, expressed as the difference in marginality between treatment and control datasets for all species (Fig. 5a), differed significantly across treatments ($H = 26.13$, $p < 0.05$). Non-parametric multiple comparisons (multiple Steel-tests) identified that these differences were especially large in the comparisons of A10 and A40 ($p = 0.004$) as well as B10 and B40 ($p = 0.057$). When measured as the difference in tolerance (Fig. 5b), environmental bias showed no significant differences across treatments ($H = 10.18$, $p = 0.069$).

Environmental bias and model performance across treatments

Model performance did not differ significantly among treatments with different levels of bias (see Table S2 in Supplementary Material). Model performance was assessed using conventional ROC analysis, maximum TSS and the modified version of the ROC analysis. To establish exactly how model performance was influenced, we evaluated the relationship between environmental bias and the difference in model performance (AUC) between models based on biased and control datasets, across all treatments (Fig. 6). Some species (e.g. *A. decurrens*, *A. longifolia* and *Hakea sericea*), had a large difference in model performance measured between the control and bias dataset even though there was only a small amount of environmental bias present in the bias dataset. The converse was also true. Nearly all treatments showed a positive correlation between the difference in marginality and the difference in AUC, although the relationships were not significant (see Table S3 in Supplementary Material). Similarly all treatments showed a positive correlation between the difference in tolerance and difference in AUC, a significant

relationship was evident for the A40 and B40 treatment (Table S3). The region predicted as suitable for *A. saligna* in the introduced range was very similar when compared between a biased (e.g. A10 or B10) and control dataset (e.g. R10) across all treatments (see Fig. S3 in Supplementary Material). This is one of the species that show very little difference in performance across treatments. Range predictions for *A. dealbata* however, were rather different when the ranges predicted with control datasets were compared to ranges predicted with the biased datasets (Fig. 7). The differences were especially large when the prediction obtained with the R40 dataset was compared to the ranges obtained with the A40 and B40 dataset. This is one of the species that showed a large difference in performance across treatments.

DISCUSSION

Several studies have reported niche shifts between the native and introduced ranges of alien invasive organisms (Mau-Crimmins *et al.*, 2006; Broenniman *et al.*, 2007; Fitzpatrick *et al.*, 2007; Beaumont *et al.*, 2009). This ecological occurrence has been recognized when an invasive species occupies habitats or climatic zones different from those encountered in the species' native range (Mau-Crimmins *et al.*, 2006; Broenniman *et al.*, 2007; Fitzpatrick *et al.*, 2007; Beaumont *et al.*, 2009). A study done by Broennimann *et al.* (2007) provides evidence of an invasive species occupying a climatically different niche compared between its native and introduced range. Similar results were found by Fitzpatrick *et al.* (2007), who showed that the invasive fire ant species initially invaded areas in their introduced range that showed conditions similar to that of their native range, and then started to spread to regions that were climatically

dissimilar to their native range. These shifts in niche occupation were argued to be the result of either genetic change that occurred after introduction, e.g. hybridization (shifts in the fundamental niche) or release from biotic constraints, e.g. competitors (shifts in the realized niche) or a combination of these two factors (Mau-Crimmins *et al.*, 2006; Broenniman *et al.*, 2007; Fitzpatrick *et al.*, 2007; Beaumont *et al.*, 2009). We calibrated models for 19 species using native range occurrence records and evaluated them with occurrence records obtained from their introduced range. To ensure that these evaluations were fair we had to assume that the niches occupied by these species were conserved between the native and introduced ranges (Wiens & Graham, 2005). Following Steiner *et al.* (2008), we tested for niche shift using Ecological-Niche Factor Analyses (ENFA; Hirzel *et al.*, 2002). Marginality values were not significantly different between the native and introduced ranges. Tolerance values showed significant differences for 12 out of the 19 species (Table S1). This indicated that for all the species considered here, no significant shift in the niche position had occurred, i.e. that the centers of the niches occupied in the native and introduced ranges were in the same place of the environmental space but that the amount of environmental variation differed significantly between the native and introduced ranges.

Introduced range models were only useful for all the species except for *A. dealbata*, *A. decurrens* and *A. mearnsii*, which achieved poor model performance. When models are calibrated with native range records and evaluated with introduced range records, model performance may be influenced by how much overlap exists in the environmental space between the environmental variation of a species' native and introduced range. To visualise the environmental space occupied by a species in its

native and introduced range we plotted component scores of components 1 and 2 from a PCA performed on the environmental data associated with the occurrence records in the native and introduced range. Some species did show good overlap in the resulting environmental space, but the majority of species did show an expansion or contraction of the niche occupied in the introduced range (Fig. S1b). This could account for the poor model performance achieved by some species.

We proceeded to model this suite of species, since we argue that at least the overall niche position was conserved between the native and introduced ranges of a species despite a difference in the amount of environmental variation and relatively poor overlap between the two ranges, which most likely suggest early signs of niche shift. The presence of niche shift between the native and introduced range of the species, could potentially have complicated our analyses since it could be argued that the geographical bias that was simulated in the native range occurrence records could be part of the environmental space that was actually not occupied by the species in the introduced range, hence that niche shift could mask the influence of the resulting environmental bias on model performance. However, for most species the records (native range) that were excluded to create a biased dataset, did not reside in the region of the environmental space occupied with the introduced range records that were used for model evaluation (Fig. S1b).

Relationship between geographical bias and environmental bias

We found considerable variation in the relationship between geographical bias and environmental bias when examined across species. This result is consistent with that of

Kadmon *et al.* (2004) and suggests that the effect of geographical bias on the underlying environmental bias most likely depends on where the geographical range of the species is located in relation to the underlying environmental gradients and the steepness of these gradients. For example, a species with a geographical range that covers a steep environmental gradient can be expected to show considerable environmental bias in the presence of a small amount of geographical bias in the dataset. We focussed on the annual precipitation gradient (BIO 12) in both native ranges when describing the main environmental gradient encountered in a range as this variable was identified to have the largest amount of variation. In Australia this gradient gradually changes from very dry interior areas to wetter areas all along the coast. Species native to Australia showed geographical distributions all along the coastal areas that varied from the south western coast to the south eastern coast excluding Tasmania (*A. cyclops*), south western to eastern coast including Tasmania (*A. dealbata*, *A. longifolia*, *A. mearnsii* and *Paraserianthes lophantha*), from the south western coast all along to the east coast excluding Tasmania (*A. decurrens*), from the north eastern to the south western coast including Tasmania (*A. melanoxydon*), west to east coast excluding Tasmania (*A. saligna*) south western coast to east coast including Tasmania (*Leptospermum laevigatum*), to an east/south eastern coast only distribution (*Hakea sericea*). In South Africa however, low rainfall areas are encountered in the north western side of the country and gradually increases towards the eastern and north eastern parts of the country. Species native to South Africa showed geographical distributions that stretched from the west coast to the interior part of the country to the south eastern coast (*Arctotheca calendula*), all along the coast to the northern interior (*Ehrharta calycina*, *Polygala myrtifolia* and *Zantedeschia aethiopica*),

along the south western coast only (*Moraea flaccida* and *Sparaxis bulbifera*), west/south western coast (*Oxalis pes-caprae*) and from the south to the south western coast (*Romulea rosea*). We show that the geographical bias that was simulated in the native range occurrence records translated into different quantities of environmental bias across species (Fig. S1). For some species the environmental space occupied with all the native range records and the environmental space occupied with the records that were excluded from the dataset (biased records) were similar, i.e. small amount of environmental bias, but for the majority of the species however, different parts of the environmental space were occupied with records from these two datasets indicating larger amounts of environmental bias. We did not find any significant relationship between geographical and environmental bias for any of the treatments analysed. The sample size in this study ($n = 19$) may have been too small to detect a significant relationship given the amount of variation.

Environmental bias across treatments

There was considerable environmental variation between treatments, and as expected bias was greater in the A40 and B40 treatments compared to the A10 and B10 treatments. This variation suggests that even though we did not find a significant relationship between geographical and environmental bias, the geographical bias simulated in the occurrence records resulted in some environmental bias.

Steiner *et al.* (2008) compared the introduced niche position of two invasive ant species in terms of marginality values when modeled with different datasets. For one species the niche position was considered to be significantly different when the difference

in the marginality value was 0.17. For the other species, when the difference was 0.3. The average difference in marginality values for the treatments analysed here was very small in comparison (largest difference approximately 0.007). Therefore although the amount of environmental bias was significantly different between treatments, the amount of bias present can be considered to be small.

Effect of bias on model performance

On average for all species geographical bias and underlying environmental bias did not influence model performance significantly. This is in contrast to the results found by Kadmon *et al.* (2003) who found a significant decrease in model performance in the presence of environmental bias. In the presence of niche shift between the native and introduced range of the species, the geographical bias that was simulated in the native range occurrence records could be part of the environmental space that was actually not occupied by the species in the introduced range, hence that niche shift could mask the influence of the resulting environmental bias on model performance. However, for the majority of the species the native range records excluded from the dataset (biased records) did not reside in the relevant part of the environmental space which was occupied by the introduced range records used for model evaluation (Fig. S1b). Therefore, this finding in conjunction with the amount of environmental bias considered to be small, could explain why we did find significant environmental bias across treatments but did not find a significant difference in model performance.

For some species (e.g. of *A. saligna* presented in Fig. S3), the potential distribution predictions obtained when models were calibrated using biased datasets and

projected to the introduced range were very similar to the predictions made using records from the corresponding control datasets. Distinct differences in the region predicted as suitable was however evident for a few species (e.g. of *A. dealbata* presented in Fig. 7) suggesting that for some species the environmental gradients may have been sufficiently sampled with the biased datasets but others not. This however, remains speculative since we were unable to detect such differences with the testing sets used to evaluate the models, given that they comprised of records obtained from the introduced range. Loiselle *et al.* (2008) obtained similar results, and concluded that model performance can be quite good despite existing environmental bias, if models are calibrated with a sufficient number of occurrence records. This might explain our findings, since we used quite a large number of records for calibrating the models (Table 1). Perhaps the effect of bias would have been greater if fewer records were available for model calibration.

The evaluation of model performance is reliant on the quality of the testing set. We evaluated models with test sets that consisted of presence records obtained from the introduced range and pseudo-absence records that we generated. Pseudo-absences were randomly selected grid cells from the introduced map region that were not occupied by a presence record. A weakness of our approach is the implicit assumption that the records available are a true reflection of a species' distribution in the introduced range, therefore that a species is at equilibrium with its environment and that sampling effort has been thorough. However, although we specifically selected species that were considered to be well established and well sampled in their introduced ranges, it is possible that some of the species may have been under sampled.

In conclusion, we found that geographical bias did not have a significant negative effect on ecological niche model performance for invasive plant species. Numerous authors have documented how and why biased sampling might lead to records in data collections not being representative of a species' true distribution and a few studies have investigated the effects of bias on model performance. Our results are novel in the sense that we focussed specifically on invasive plant species. When making introduced range predictions for alien invasive species, it may be necessary to calibrate models with native range records. This would most likely be the case when predicting the potential distribution of emerging invaders, when only a few introduced range records are available (Mgidi *et al.*, 2007), or when the risk of a potential invader needs to be assessed in areas where invasion has not yet occurred and therefore no introduced range records are available (De Meyer *et al.*, 2007). Therefore our results are encouraging as we were able to demonstrate for this suite of species that model performance was not significantly reduced when geographical bias was present in the native range records used to calibrate the models. However, we may not have been able to detect the true effect of geographical bias with the testing sets used to evaluate our models. Furthermore, the effects of sample size, spatial resolution of predictor variables, range size and geographical region may influence environmental bias and require further investigation. Loiselle *et al.* (2008) found that species with larger range sizes tended to be affected more by environmental bias than those with narrower ranges.

We recommend that future studies analyse more species to investigate geographical and environmental bias as we found considerable variation among species. WE suggest that a similar study be conducted, to determine how niche shift influences

studies evaluating the influence of geographical bias on model performance. Specifically, when models are calibrated with native range records and evaluated with introduced range records. Instead however, model performance should be compared between projections made to the native range of the species when models are calibrated with a geographically biased dataset and control dataset. Furthermore, our study was conducted using relatively coarse scale (15 minute) predictors and finer scale studies may yield different results. Finally, a wide variety of modeling techniques are available all of which have unique strengths and weaknesses (Elith *et al.*, 2006) and therefore studies on the sensitivity of different modeling techniques to bias specifically, would be very valuable.

ACKNOWLEDGEMENTS

Lesley Henderson kindly provided the SAPIA database, and assisted with the species selection process. Rod Randall provided information about the South African plants invading Australia. The Council of Heads of Australian Herbaria and the custodial institutions, the Australian National Herbarium (Canberra), the National Herbarium of New South Wales (New South Wales), Northern Territory Herbarium (Darwin and Alice Springs), Queensland Herbarium (Brisbane), State Herbarium of South Australia (Adelaide), Tasmanian Herbarium (Hobart), National Herbarium of Victoria (Melbourne) and the Western Australian Herbarium (Perth) are acknowledged for providing Australian plant specimen data, especially Alex Chapman, Murray Fagg and Greg Whitbread for assistance. Paul Robins, Queensland herbarium, Environmental Protection Agency, in charge of the HERBRECS kindly provided distribution data. Many thanks to Steven

Phillips and Miroslav Dudik for help and advice with Maxent. We would also like to express our gratitude to Jeremy van der Wal who provided statistical advice and shared some of his scripts with us. We thank two anonymous reviewers who provided helpful comments on an earlier draft of this manuscript. This project was funded by the University of Pretoria and the DST-NRF Centre of Excellence for Invasion Biology.

REFERENCES

- Allouche, O., Tsoar, A. & Kadmon, R. (2006) Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, **43**, 1223-1232.
- Beaumont, L.J., Gallagher, R.V., Thuiller, W., Downey, P.O., Leishman, M.R. & Hughes, L (2009) Different climatic envelopes among invasive populations may lead to underestimations of current and future biological invasions. *Diversity and Distributions*, **15**, 409-420.
- Broennimann, O., Treier, U.A., Muller-Scharer, H., Thuiller, W., Peterson, A.T. & Guisan, A. (2007) Evidence of climatic niche shift during biological invasion. *Ecology Letters*, **10**, 701-709.
- De Meyer, M., Robertson, M.P., Peterson, A.T. & Mansell, M.W. (2007) Ecological modelling and potential geographic distributions of Mediterranean fruit fly (*Ceratitidis capitata*) and Natal fruit fly (*Ceratitidis rosa*). *Journal of Biogeography*, **35**, 270-281.
- Dennis, R.L.H. & Thomas, C.D. (2000) Bias in butterfly distribution maps: the influence of hot spots and recorder's home range. *Journal of Insect Conservation*, **4**, 73-77.
- Elith, J., Graham, C.H., Anderson, R.P., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A.,

Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M., Townsend-Peterson, A., Phillips, S., Richardson, K., Scachetti-Pereira, R., Schapire, R.E., Soberon, J., Williams, S., Wisz, M.S. & Zimmermann, N.E. (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129-151.

Fielding, A.H. & Bell, J.F. (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, **24**, 38-49.

Fitzpatrick, M.C., Weltzin, J.F., Sanders, N.J. & Dunn, R.R. (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**, 24-33.

Funk, V.A. & Richardson, K.S. (2002) Systematic data in biodiversity studies: Use it or lose it. *Systematic Biology*, **51**, 303-316.

Gibson, L., Barrett, B. & Burbidge, A. (2007) Dealing with uncertain absences in habitat modeling: a case study of a rare ground-dwelling parrot. *Diversity and Distributions*, **13**, 704-713.

Guisan, A. & Zimmermann, N.E. (2000) Predictive habitat distribution models in ecology. *Ecological Modelling*, **135**, 147-186.

Guisan, A. & Thuiller, W. (2005) Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, **8**, 993-1009.

Henderson, L. (1998) Southern Africa plant invaders atlas (SAPIA). *Applied Plant Sciences*, **12**, 31-32.

Henderson, L. (2006b) Comparisons of invasive plants in southern Africa originating from southern temperate, northern temperate and tropical regions. *Bothalia*, **36**, 210-222.

Henderson, L. (2007) Invasive, naturalized and casual alien plants in southern Africa: a summary based on the Southern African Plant Invaders Atlas (SAPIA). *Bothalia*, **37**, 215-248.

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

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965-1978.

Hirzel, A.H., Hausser, J., Chessel, D. & Perrin, N. (2002) Ecological-Niche Factor Analysis: How to compute habitat suitability maps without absence data? *Ecology*, **83**, 2027-2036.

Hortal, J., Jiminez-Valverde, A., Gomez, J.F., Lobo, J.M. & Baselga, A. (2008) Historical bias in biodiversity inventories affects the observed environmental niche of the species. *Oikos*, **117**, 847-858.

Jiminez-Valverde, A. & Lobo, J.M. (2007) Threshold criteria for conversion of probability of species presence to either-or presence-absence. *Acta Oecologia*, **31**, 361-369.

Kadmon, R., Farber, O. & Danin, A. (2003) A systematic analysis of factors affecting the performance of climatic envelope models. *Ecological Applications*, **13**, 853-867.

Kadmon, R., Farber, O. & Danin, A. (2004) Effect of roadside bias on the accuracy of predictive maps produced by bioclimatic models. *Ecological Applications*, **14**, 401-413.

Lobo, J.M., Jimenez-Valverde, A. & Real, R. (2008) AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, **17**, 145-151.

Loiselle, B.A., Jorgenson, P.M., Consiglio, T., Jimenez, I., Blake, J.G., Lohmann, L.G. & Montiel, O.M. (2008) Predicting species distributions from herbarium collections: does climate bias in collection sampling influence model outcomes? *Journal of Biogeography*, **35**, 105-116.

Mau-Crimmins, T., Schussman, H.R. & Geiger, H. (2006) Can the invaded range of a species be predicted sufficiently using only native-range data? Lehmann lovegrass (*Eragrostis lehmanniana*) in the southwestern United States. *Ecological Modelling*, **19**, 736-746.

McPherson, J.M. & Jetz, W. (2007) Effects of species' ecology on the accuracy of distribution models. *Ecography*, **30**, 135-151.

Mgidi, T., Le Maitre, D.C., Schonegevel, L., Nel, J.L., Rouget, M. & Richardson, D.M. (2007) Alien plant invasions – incorporating emerging invaders in regional prioritization: A pragmatic approach for Southern Africa. *Journal of Environmental Management*, **84**, 173-187.

Pearman, P.B., Guisan, A., Broennimann, O. & Randin, C.F. (2008) Niche dynamics in space and time. *Trends in Ecology and Evolution*, **23**, 149-158.

Pearson, R.G. & Dawson, T.P. (2003) Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Global Ecology & Biogeography*, **12**, 361-371.

Pearson, R.G., Thuiller, W., Araujo, M.B., Martinez-Meyer, E., Brotons, L., McClean, C., Miles, L., Segurado, P., Dawson, P. & Lees, D.C. (2006) Model-based uncertainty in species range prediction. *Journal of Biogeography*, **33**, 1704-1711.

Pearson, R.G., Raxworthy, C.J., Nakamura, M. & Peterson, A.T. (2007) Predicting species distribution from small occurrence records: a test case using cryptic geckos in Madagascar. *Journal of Biogeography*, **34**, 102-117.

Peterson, A.T. (2003) Predicting the geography of species' invasions via ecological niche modeling. *The Quarterly Review of Biology*, **78**, 419-433.

Peterson, A.T., Papes, M. & Soberon, J. (2008) Rethinking receiver operating characteristic analysis applications in ecological niche modeling. *Ecological Modelling*, **213**, 63-72.

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

Phillips, S.J. & Dudik, M. (2008) Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography*, **31**, 161-175.

Pimentel, D., McNair, S., Janecka, J., Wightman, J., Simmonds, C., O'Connell, C., Wong, E., Russel, L., Zern, J., Aquino & Tsomondo. (2001) Economic and environmental threats of alien plant, animal and microbe invasions. *Agriculture, Ecosystems and Environment*, **84**, 1-20.

R Development Core Team 2004. 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>.

Raes, N. & Ter Steege, H. (2007) A null-model for significance testing of presence-only species distribution models. *Ecography*, **30**, 727-736.

Randall, R.P. (2002) *A Global Compendium of weeds*, Shannon Books, Melbourne, Australia.

Reddy, S. & Davalos, L.M. (2003) Geographical sampling bias and its implications for conservation priorities in Africa. *Journal of Biogeography*, **30**, 1719-1727.

Richardson, D.M. (2004) Plant invasion ecology – dispatches from the front line. *Diversity and Distributions*, **10**, 315-319.

Richardson, D.M. & van Wilgen, B. (2004) Invasive alien plants in South Africa: how well do we understand the ecological impacts? *South African Journal of Science*, **100**, 45-52.

Richardson, D.M. & Thuiller, W. (2007) Home away from home – objective mapping of high risk source areas for plant introductions. *Diversity and Distributions*, **1**, 1-14.

Robertson, M.P., Caithness, N. & Villet, M.H. (2001) A PCA-based modelling technique for predicting environmental suitability for organisms from presence records. *Diversity and Distributions*, **7**, 15-27.

Robertson, M.P., Peter, C.I., Villet, M.H. & Ripley, B.S. (2003) Comparing models for predicting species potential distribution: a case study using correlative and mechanistic predictive modelling techniques. *Ecological Modelling*, **164**, 153-167

Rouget, M., Richardson, D.M., Nel, J.L., Le Maitre, D.C., Egoh, B. & Mgidi, T. (2004) Mapping the potential ranges of major plant invaders in South Africa, Lesotho and Swaziland using climatic suitability. *Diversity and Distributions*, **10**, 475-484.

Roura-Pascual, N., Brotons, L., Peterson, A.T. & Thuiller, W. (2008) Consensual predictions of potential distributional areas for invasive species: a case study of Argentine ants in the Iberian Peninsula. *Biological Invasions*, **11**, 1017-1031.

Soberon, J. & Peterson, A.T. (2005) Interpretation of models of fundamental ecological niches and species' distributional areas. *Biodiversity Informatics*, **2**, 1-10.

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

Stockwell, D.R.B. & Peterson, A.T. (2002a) *Predicting species occurrences: Issues of scale and accuracy*, (ed. by J.M. Scott, P.J. Heglund, M. Morrison, M. Raphael, J. Haufler, B. Wall) chapter 48. Island Press, Covello, California.

Thuiller, W., Richardson, D.M., Pysek, P., Midgley, G.F., Hughes, G.O. & Rouget, M. (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.

Vitousek, P.M., D'Antonio, C.M., Loope, L.L., Rejmanek, M. & Westbrooks, R. (1997) Introduced species: a significant component of human-caused global change. *New Zealand Journal of Ecology*, **21**, 1-16.

Welk, E., Schubert, K. & Hoffmann, M.H. (2002) Present and potential distribution of invasive garlic mustard (*Alliaria petiolata*) in North America. *Diversity and Distributions*, **8**, 219-233.

Welk, E. (2004) Constraints in range predictions of invasive plant species due to non-equilibrium distribution pattern: Purple loosestrife (*Lythrum salicaria*) in North America. *Ecological Modelling*, **179**, 551-567.

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

Wilson, J.R.U., Richardson, D.M., Rouget, M., Poches, S., Amis, M.A., Henderson, L. & Thuiller, W. (2007) Residence time and potential range: crucial considerations in modelling plant invasions. *Diversity and Distributions*, **13**, 11-22.

Yurkonis, K.A., Meiners, S.J. & Wachholder, B.E. (2005) Invasion impacts diversity through altered community dynamics. *Journal of Ecology*, **93**, 1053-1061.

Zimmermann, N.E., Edwards, T.C., Moisen, G.G., Frescino, T.S. & Blackard, J.A. (2007) Remote sensing-based predictors improve distribution models of rare, early successional and broadleaf tree species in Utah. *Journal of Applied Ecology*, **44**, 1057-1067.

SUPPLEMENTARY MATERIAL

Table S1. Marginality and tolerance values compared between native and introduced ranges.

Table S2. Model performance compared among treatments with Kruskal-Wallis tests.

Table S3. Correlation between environmental bias and model performance.

Figure S1. Geographic and environmental distribution of all the species.

Figure S2. Relationship between native and introduced range model performance.

Figure S3. Introduced range predictions for *A. saligna*.

TABLE AND FIGURE LEGENDS

Table 1. Species selected for analysis. The first ten species indicated the Australian species invading South Africa and the arrow (^) indicates those South African species invading Australia. The number of native range records used for calibration and the number of introduced range records used for evaluation.

Table 2. The 19 bioclimatic variables (code BIO 1- BIO 19) obtained from the Worldclim database (www.worldclim.org/methods.htm) that were used as predictor variables.

Table 3. Relationship between geographical bias (expressed as proportional geographic difference) and environmental bias (expressed as the difference in marginality or tolerance) evaluated across all the different treatments. Regression (R^2 -value) and correlation (r -value) analyses were carried out by comparing values from a control (e.g. R10) and bias dataset (A10 or B10). $F(1,17)$ is the proportion of variation explained between the control and bias group, degrees of freedom ($n-2$), $n=19$. No significant relationship was found ($p > 0.05$), except for the B10 treatment indicated with an asterisk (*).

Figure 1. A flowchart of the different datasets used to make native and introduced range predictions for a species. All datasets comprised of samples of occurrence records obtained from the native range of a species.

Figure 2. A hypothetical example indicating the selection of occurrence records used to simulate geographical bias in datasets that were used to calibrate models. Green symbols indicate occurrence records that were excluded and red symbols indicate those that were selected. a) Illustrates geographical bias scenario (A) where a proportion of records (e.g. 10%; A10) were excluded from a specific geographic region of a species' range. b) Illustrates the geographical bias scenario (B) where only some records were excluded (e.g. 10%; B10) from a specific region of a species' range. c) Indicates random exclusion of occurrence records from the entire region of the species' range for the control datasets (e.g. 10%; R10).

Figure 3. A diagrammatic illustration of the relationship between the a) geographic and b) environmental bias for *A. decurrens*. The geographic distribution of the species shown across the annual precipitation gradient (BIO 12) encountered in the native range, darker colors indicate wetter areas. The extent of the entire native range of the species indicated with red dots and the geographic distribution represented with the biased dataset (A40) indicated with green dots. The environmental distribution of the species shown by the principal components analysis (PCA) of the environmental conditions associated with the native (red dots) and introduced (black dots) range of the species. The green dots represent the records excluded when the A40 dataset was created. The percentage of environmental variation explained by a specific component given in parenthesis.

Figure 4. Relationship between geographical and environmental bias assessed across all treatments for all species. Geographical bias quantified as the proportional difference

between records from a control and bias dataset using the x- or y-coordinates of a particular dataset, depending on the direction from which records were excluded in order to simulate bias. (measured in decimal degrees). Environmental bias expressed as the difference in marginality between a control and bias dataset.

Figure 5. Environmental bias expressed as a) difference in marginality and b) difference in tolerance assessed across all treatments for all species.

Figure 6. The relationship between environmental bias and model performance assessed across all treatments for all species. Environmental bias expressed as the difference in marginality. Model performance expressed as the AUC (obtained from the conventional ROC analysis) difference calculated as the absolute difference in AUC values between a specific control (e.g. R10) and bias dataset (e.g. A10).

Figure 7. Introduced range predictions for *A. dealbata* with all three control datasets (R10, R20 and R40), datasets from treatment A (A10, A20 and A40) as well as datasets from treatment B (B10, B20 and B40).

Table 1.

Botanical name	Family	Common name	Number of native records	Number of introduced records
<i>Acacia cyclops</i>	Fabaceae	Red-eye	100	150
<i>Acacia dealbata</i>	Fabaceae	Silver wattle	200	285
<i>Acacia decurrens</i>	Fabaceae	Green wattle	101	110
<i>Acacia longifolia</i>	Fabaceae	Sydney golden wattle	220	86
<i>Acacia mearnsii</i>	Fabaceae	Black wattle	135	430
<i>Acacia melanoxylon</i>	Fabaceae	Australian blackwood	287	138
<i>Acacia saligna</i>	Fabaceae	Port Jackson willow	268	146
<i>Hakea sericea</i>	Proteaceae	Silky hakea	47	79
<i>Leptospermum laevigatum</i>	Myrtaceae	Australian myrtle	112	36
<i>Paraserianthes lophantha</i>	Fabaceae	Stinkbean	83	45
<i>Arctotheca calendula</i> [^]	Asteraceae	Cape weed	97	431
<i>Ehrharta calycina</i> [^]	Poaceae	Veld grass	240	115
<i>Moraea flaccida</i> [^]	Iridaceae	One leaf cape tulip	16	41
<i>Oxalis pes-caprae</i> [^]	Oxalidaceae	Soursob	39	117
<i>Polygala myrtifolia</i> [^]	Polygalaceae	Sweet pea bush	125	41
<i>Romulea rosea</i> [^]	Iridaceae	Oniongrass	57	109
<i>Sparaxis bulbifera</i> [^]	Iridaceae	Sparaxis	17	58
<i>Watsonia meriana</i> [^]	Iridaceae	Bulbil watsonia	44	36
<i>Zantedeschia aethiopica</i> [^]	Araceae	Arum lily	60	34

Table 2.

Code	Bioclimatic variable
BIO 1	Annual Mean Temperature
BIO 2	Mean Diurnal Range [Mean of monthly (max temperature-min temperature)]
BIO 3	Isothermality (BIO 2/BIO 7) (*100)
BIO 4	Temperature Seasonality (standard deviation *100)
BIO 5	Max Temperature of Warmest Month
BIO 6	Min Temperature of Coldest Month
BIO 7	Temperature Annual Range (BIO 5-BIO 6)
BIO 8	Mean Temperature of Wettest Quarter
BIO 9	Mean Temperature of Driest Quarter
BIO 10	Mean Temperature of Warmest Quarter
BIO 11	Mean Temperature of Coldest Quarter
BIO 12	Annual Precipitation
BIO 13	Precipitation of Wettest Month
BIO 14	Precipitation of Driest Month
BIO 15	Precipitation Seasonality (Coefficient of Variation)
BIO 16	Precipitation of Wettest Quarter
BIO 17	Precipitation of Driest Quarter
BIO 18	Precipitation of Warmest Quarter
BIO 19	Precipitation of Coldest Quarter

Table 3.

Treatment	Difference in marginality			Difference in tolerance		
	R ²	F(1,17)	r	R ²	F(1,17)	r
R10, A10	0.000	0.004	-0.019	0.026	0.450	-0.101
R10, B10	0.069	1.250	-0.510*	0.012	0.199	-0.322
R20, A20	0.033	0.588	-0.060	0.095	1.789	-0.464
R20, B20	0.171	3.510	-0.285	0.000	0.005	-0.150
R40, A40	0.063	1.140	0.168	0.000	0.002	-0.154
R40, B40	0.000	0.007	-0.021	0.000	0.013	-0.140

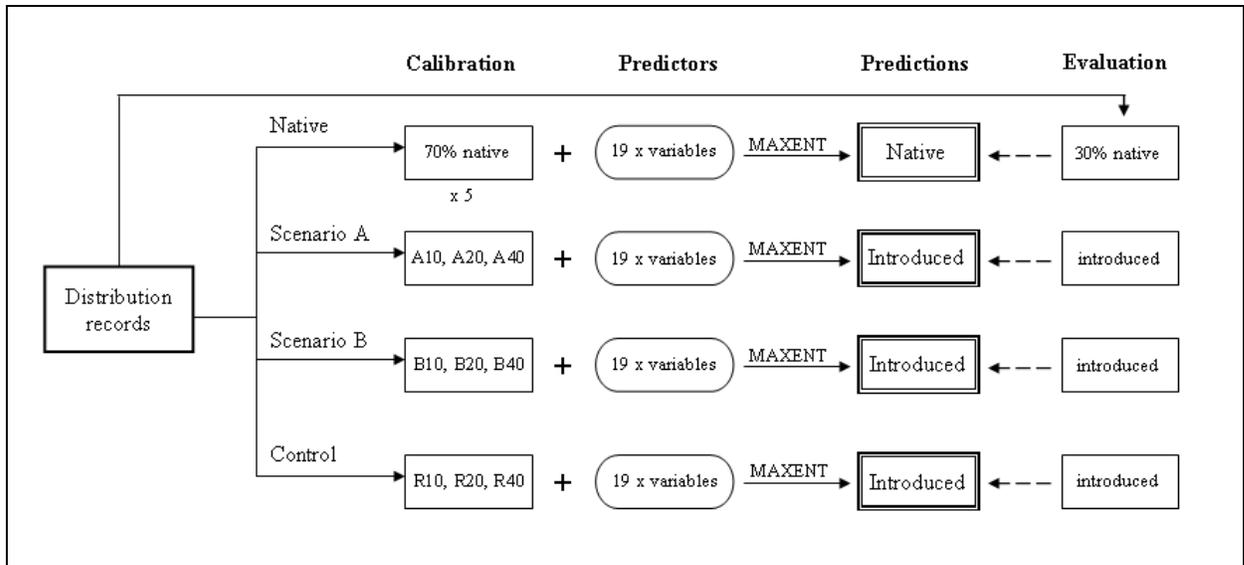


Figure 1.

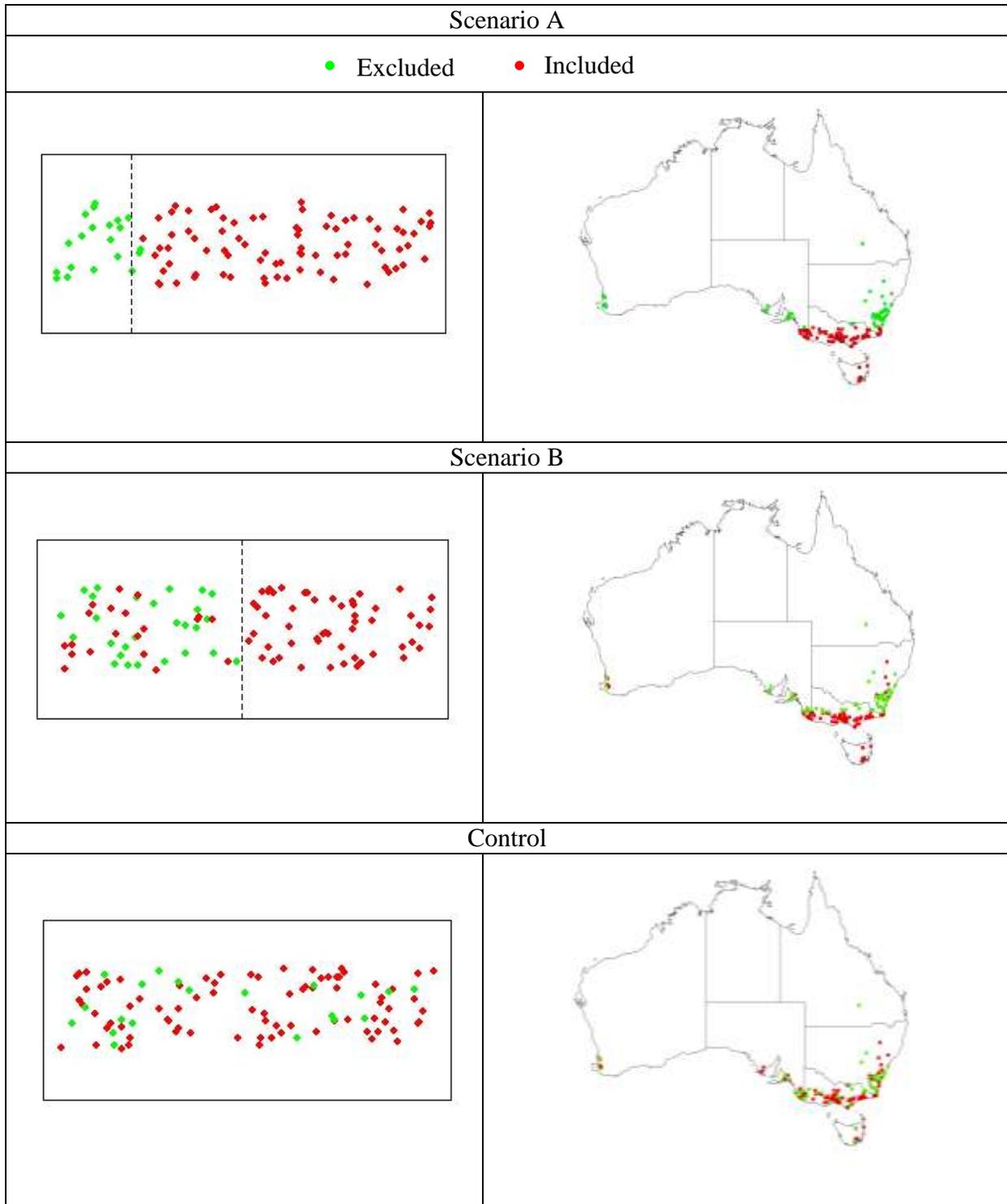


Figure 2.

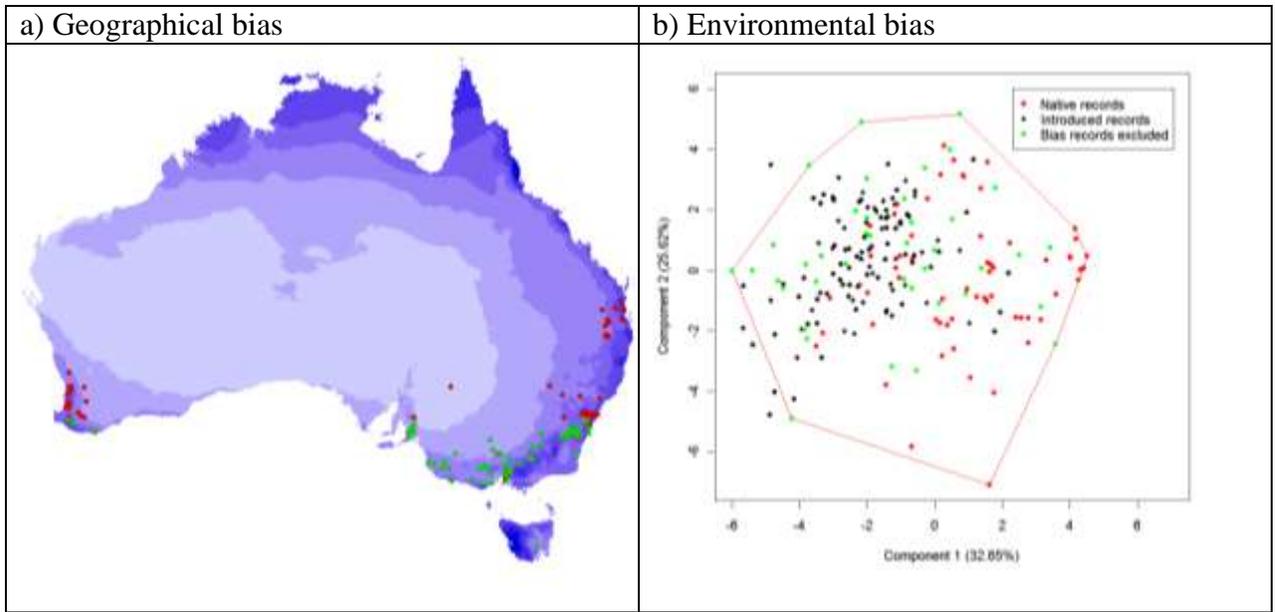


Figure 3.

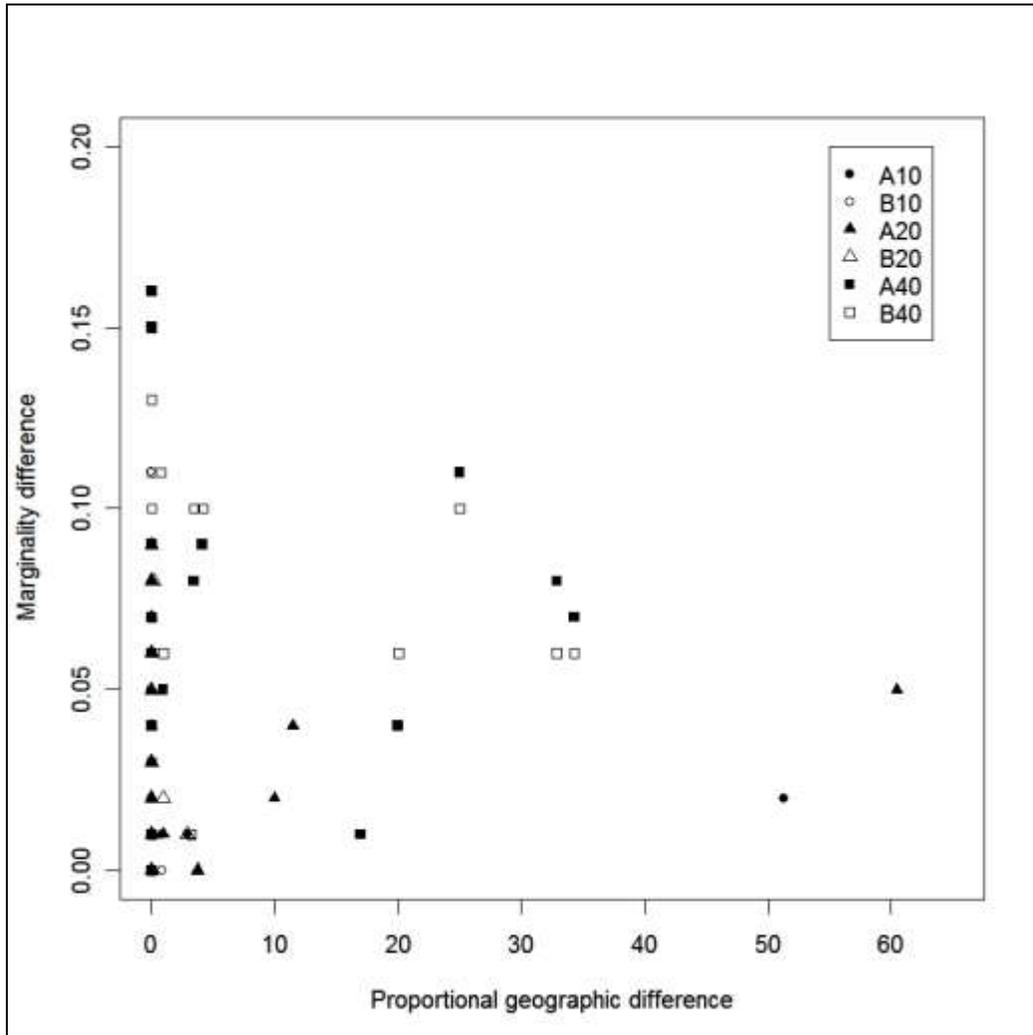


Figure 4.

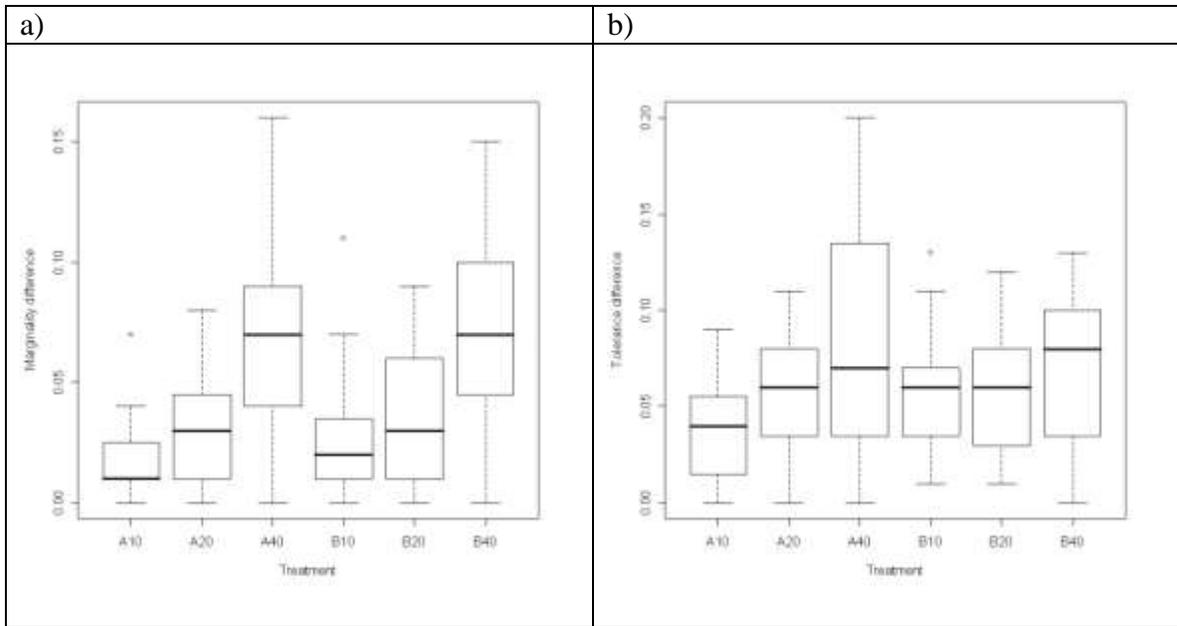


Figure 5.

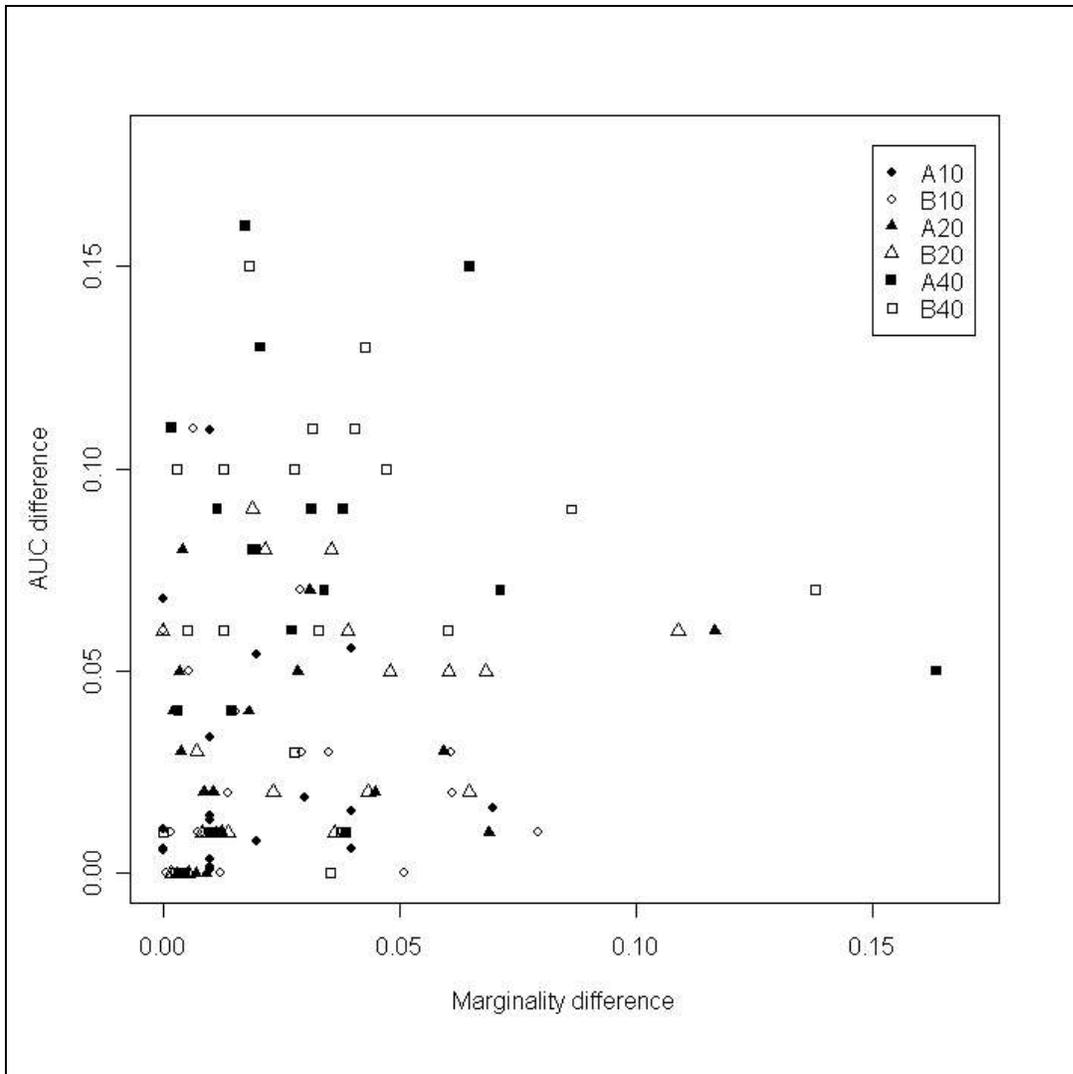


Figure 6.

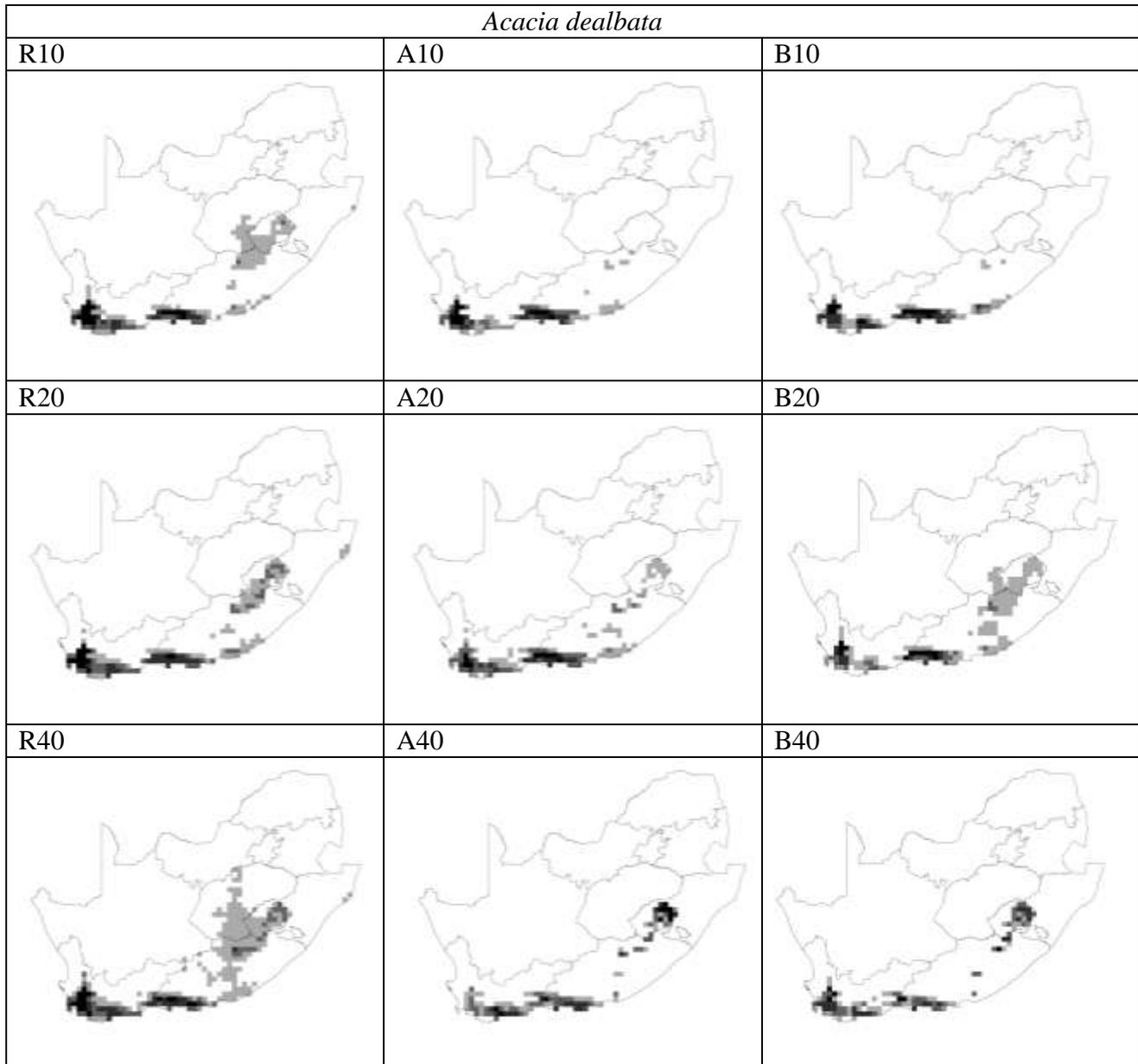


Figure 7.

SUPPLEMENTARY MATERIAL:

Table S1. Marginality and tolerance values compared between the native and introduced range. Marginality values were contrasted with Student's t-tests and tolerance values with Bartlett's test for homogeneity. Values indicated with an asterisk (*) were significantly different ($p < 0.05$) between the different ranges. Species indicated with an arrow (^) are South African species invasive in Australia.

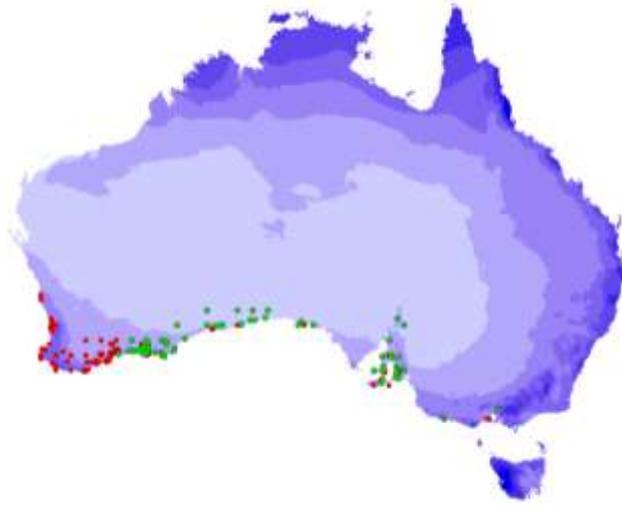
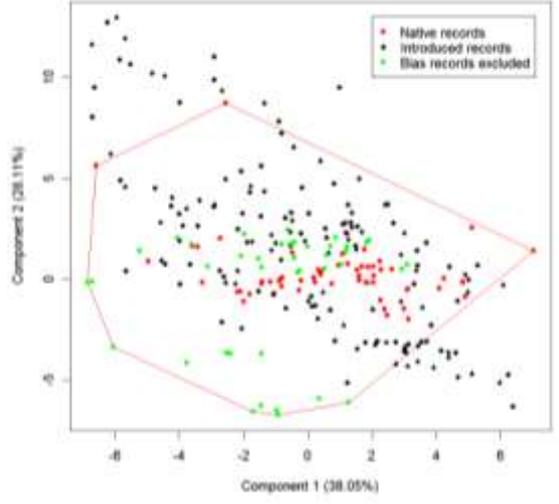
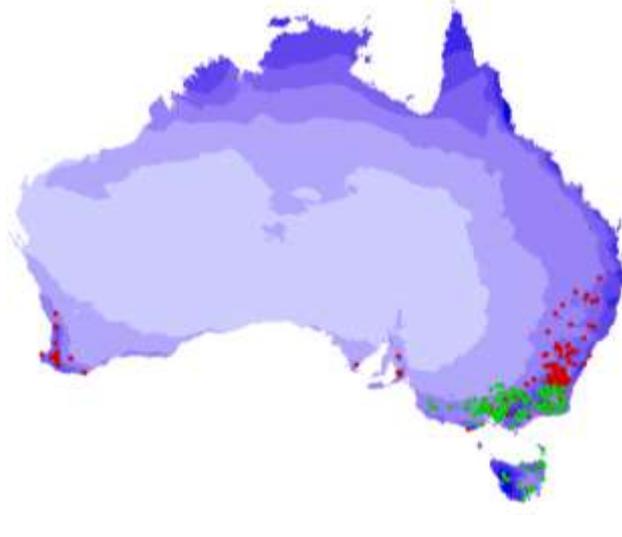
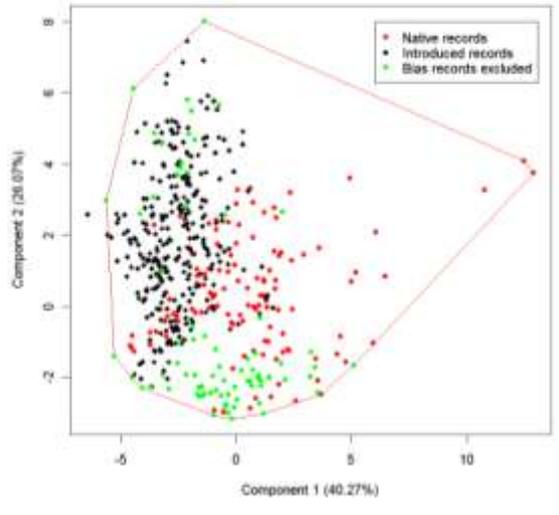
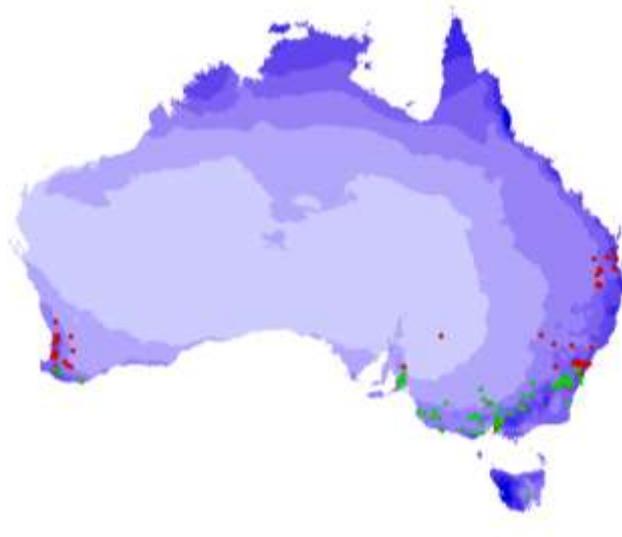
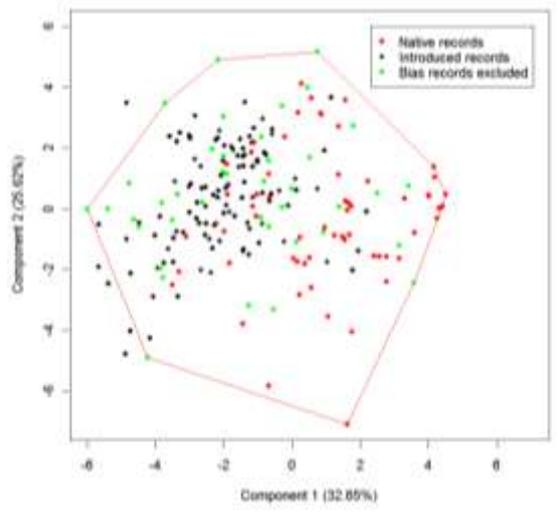
Species	Native Marginality	Introduced Marginality	Student's t-test	Native Tolerance	Introduced Tolerance	Bartlett's test
<i>Acacia cyclops</i>	1.29	1.33	0.82	3.94	2.55	20.37*
<i>Acacia dealbata</i>	1.39	1.22	4.48	2.42	2.31	0.49
<i>Acacia decurrens</i>	1.24	1.26	0.50	2.49	6.23	79.01*
<i>Acacia longifolia</i>	1.37	1.30	1.14	1.79	2.22	5.32*
<i>Acacia mearnsii</i>	1.35	1.19	3.88	2.46	1.96	9.85*
<i>Acacia melanoxylon</i>	1.36	1.19	3.42	1.88	2.09	2.12
<i>Acacia saligna</i>	1.25	1.31	1.37	2.38	2.26	0.54
<i>Hakea sericea</i>	1.38	1.35	0.51	3.50	3.14	0.65
<i>Leptospermum laevigatum</i>	1.46	1.48	0.31	2.09	4.63	24.67*
<i>Paraserianthes lophantha</i>	1.38	1.42	0.77	2.79	4.77	14.23*
<i>Arctotheca calendula</i> [^]	1.20	1.16	0.69	1.82	1.82	0
<i>Ehrharta calycina</i> [^]	1.25	1.24	0.24	1.74	3.24	49.44*
<i>Moraea flaccida</i> [^]	1.34	1.27	3.78	34.64	8.45	25.54*
<i>Oxalis pes-caprae</i> [^]	1.31	1.20	1.66	2.86	2.31	2.41
<i>Polygala myrtifolia</i> [^]	1.26	1.37	2.01	2.03	4.65	30.21*
<i>Romalea rosea</i> [^]	1.27	1.26	0.23	2.90	2.50	1.51
<i>Sparaxis bulbifera</i> [^]	1.35	1.34	0.56	27.82	6.66	28.36*
<i>Watsonia meriana</i> [^]	1.31	1.38	1.43	8.21	3.48	26.66*
<i>Zantedeschia aethiopica</i> [^]	1.23	1.37	2.07	2.07	5.73	32.5*

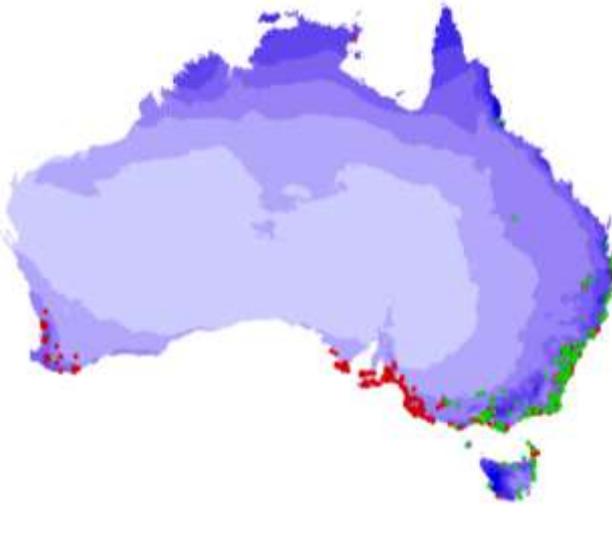
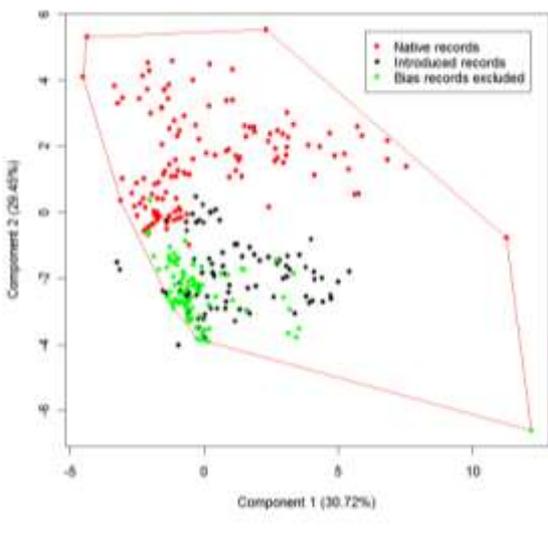
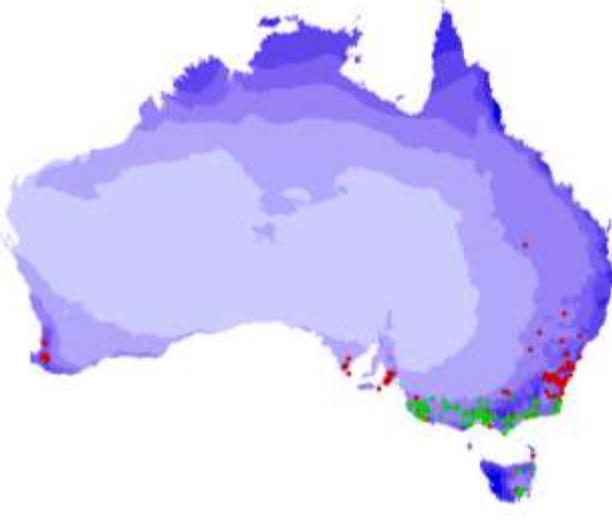
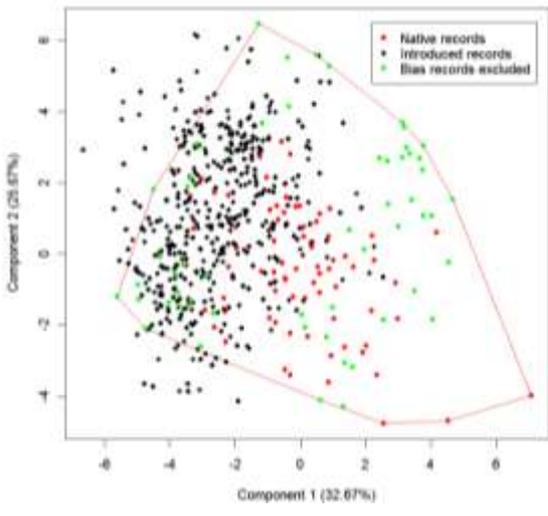
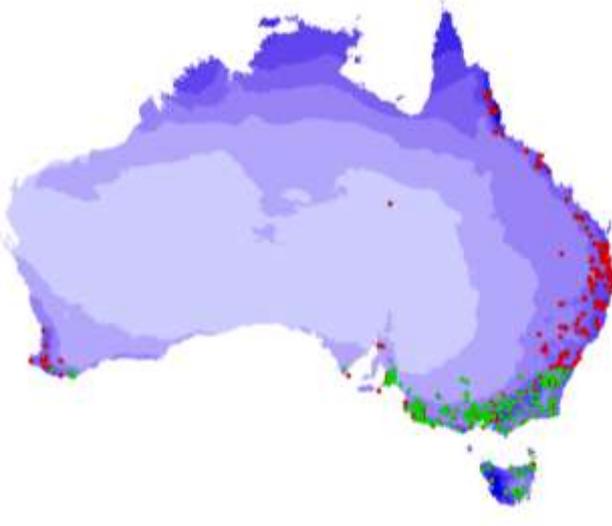
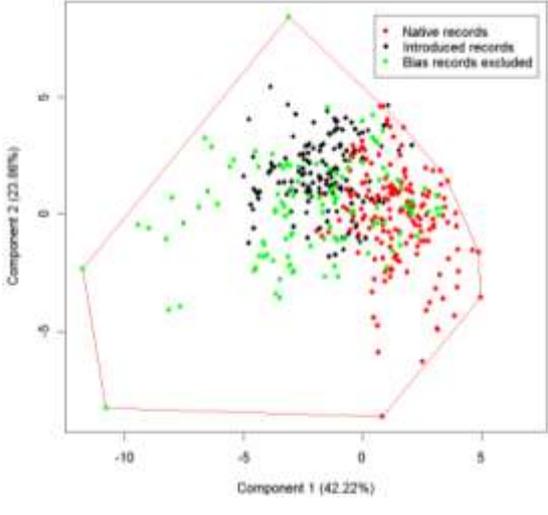
Table S2. Model performance measured with the conventional AUC analysis (AUC), maximum TSS (mTSS) and the modified version of the ROC analysis (Aratio) compared across all treatments with different levels of bias with Kruskal-Wallis tests. Values of the test statistic obtained when each model performance measure was compared between a control and biased datasets ($p > 0.05$).

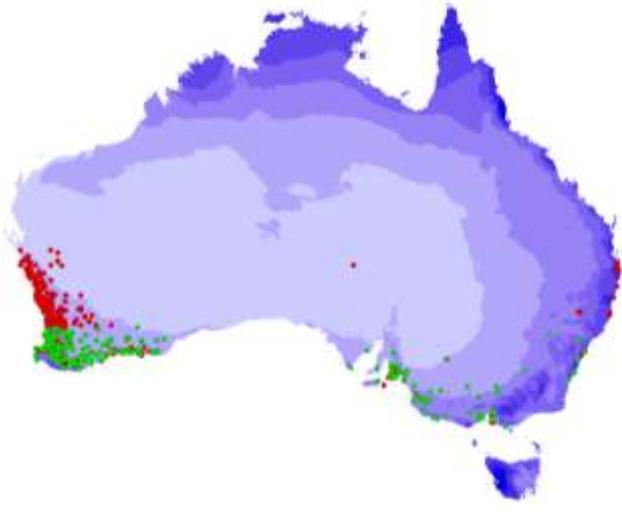
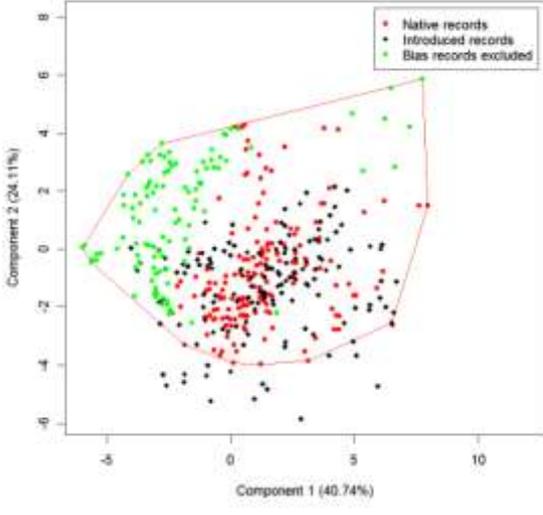
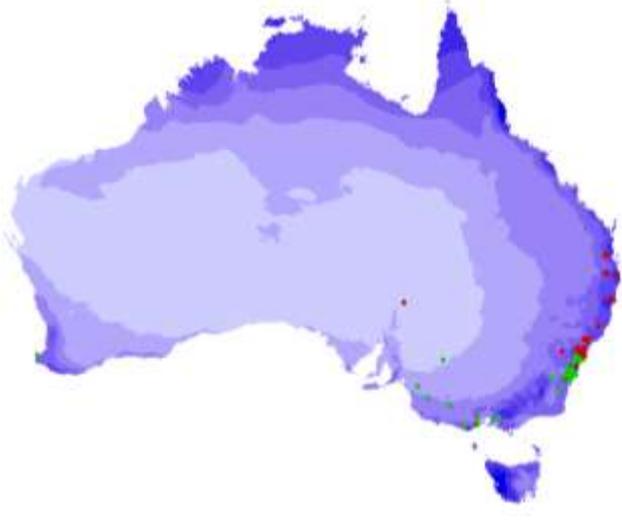
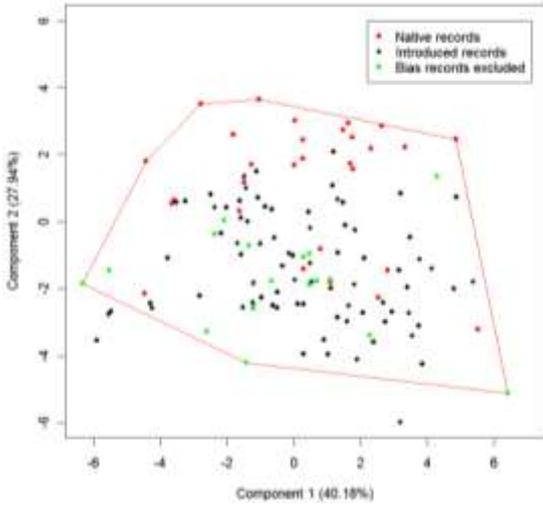
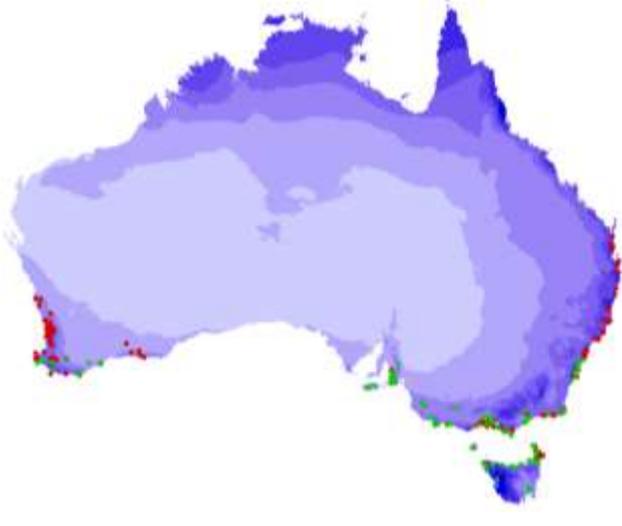
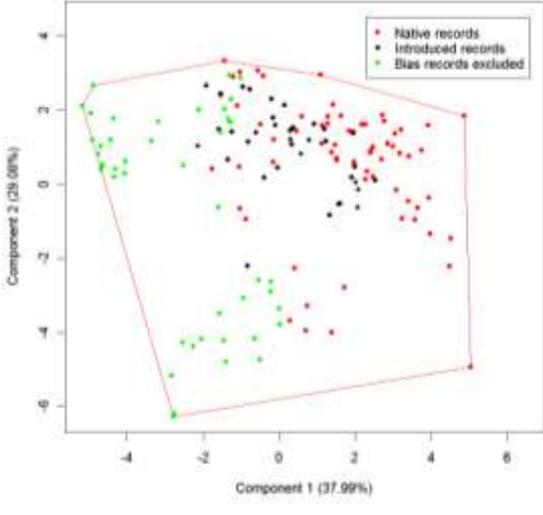
Model performance	Treatments		
	R10, A10, B10	R20, A20, B20	R40, A40, B40
AUC	0.005	0.545	0.024
mTSS	0.062	0.182	0.022
Aratio	0.012	0.292	0.028

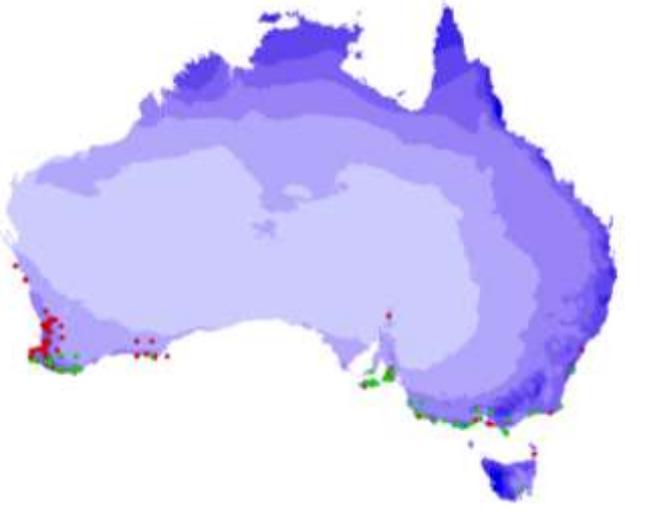
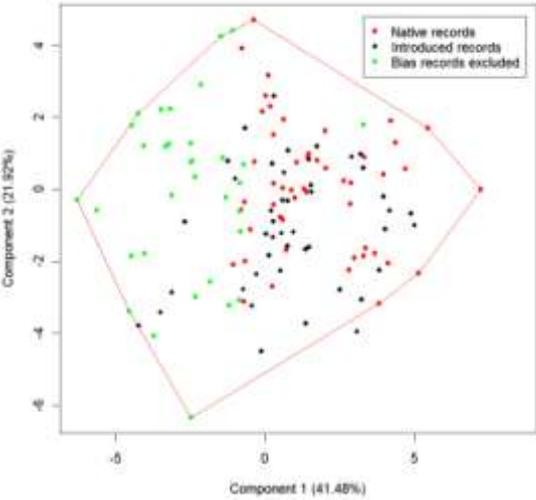
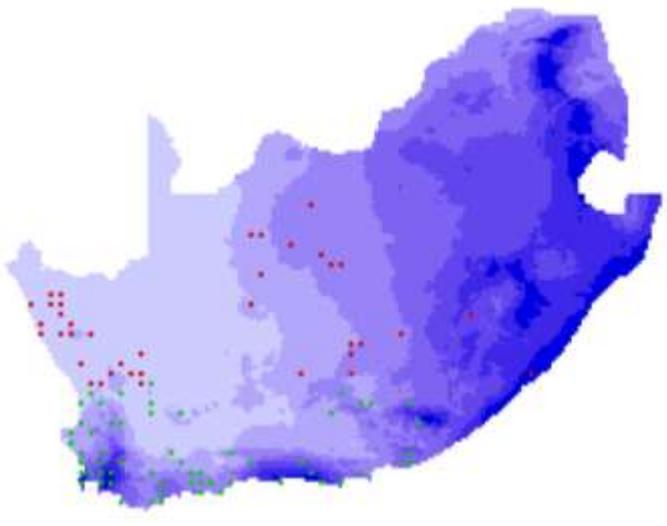
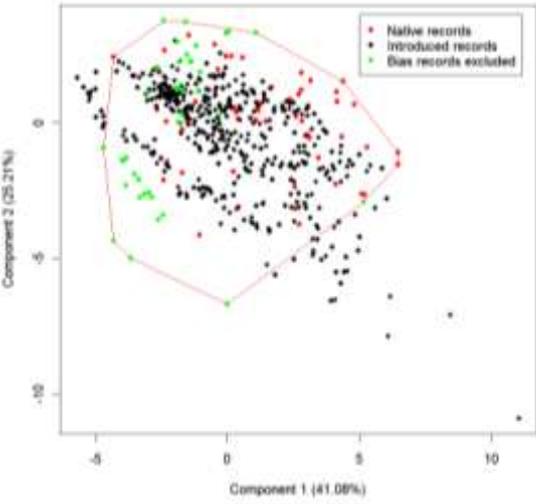
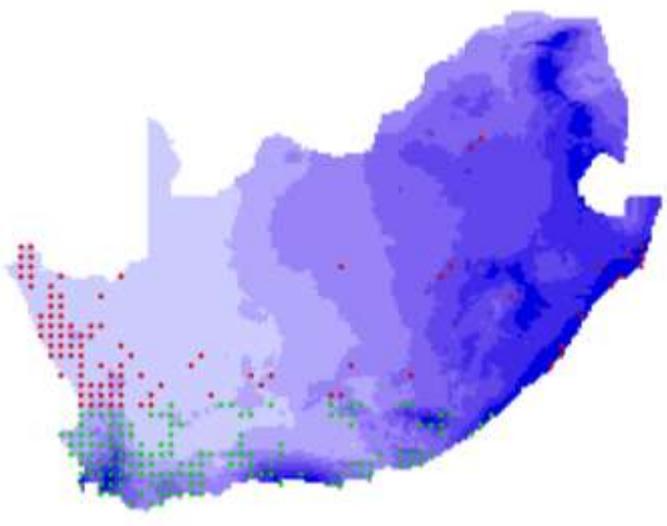
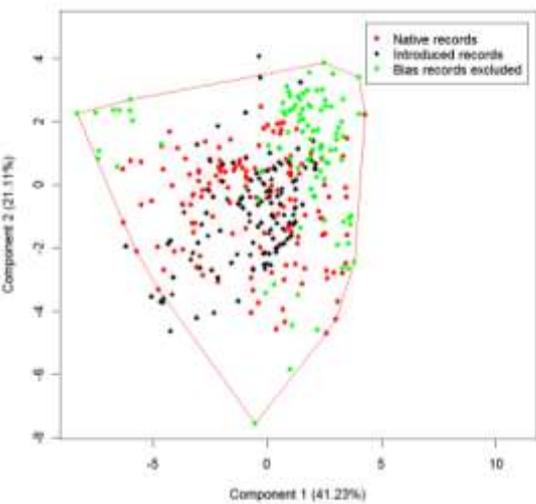
Table S3. Correlation between environmental bias (expressed as the difference in marginality or tolerance) and model performance assessed by comparing a control (e.g. R10) and bias dataset (A10 or B10). Values indicated with an asterisk (*) were significantly different ($p < 0.05$).

Treatment	Difference in marginality		Difference in tolerance	
	r	p-value	r	p-value
R10, A10	0.250	0.301	0.015	0.952
R10, B10	-0.082	0.738	0.046	0.852
R20, A20	0.131	0.594	0.204	0.403
R20, B20	0.330	0.168	0.190	0.436
R40, A40	0.199	0.414	0.578	0.009*
R40, B40	0.229	0.346	0.633	0.004*

Geographical bias	Environmental bias
<i>Acacia cyclops</i>	
	
<i>Acacia dealbata</i>	
	
<i>Acacia decurrens</i>	
	

Geographical bias	Environmental bias
<i>Acacia longifolia</i>	
	
<i>Acacia mearnsii</i>	
	
<i>Acacia melanoxylon</i>	
	

Geographical bias	Environmental bias
<i>Acacia saligna</i>	
	
<i>Hakea sericea</i>	
	
<i>Leptospermum laevigatum</i>	
	

Geographical bias	Environmental bias
<i>Paraserianthes lophantha</i>	
	
<i>Arctotheca calendula</i> [^]	
	
<i>Ehrharta calycina</i> [^]	
	



Geographical bias	Environmental bias
<i>Moraea flaccida</i> [^]	
<i>Oxalis pes-caprae</i> [^]	
<i>Polygala myrtifolia</i> [^]	



Geographical bias	Environmental bias
<i>Romulea rosea</i> [^]	
<i>Sparaxis bulbifera</i> [^]	
<i>Watsonia meriana</i> [^]	

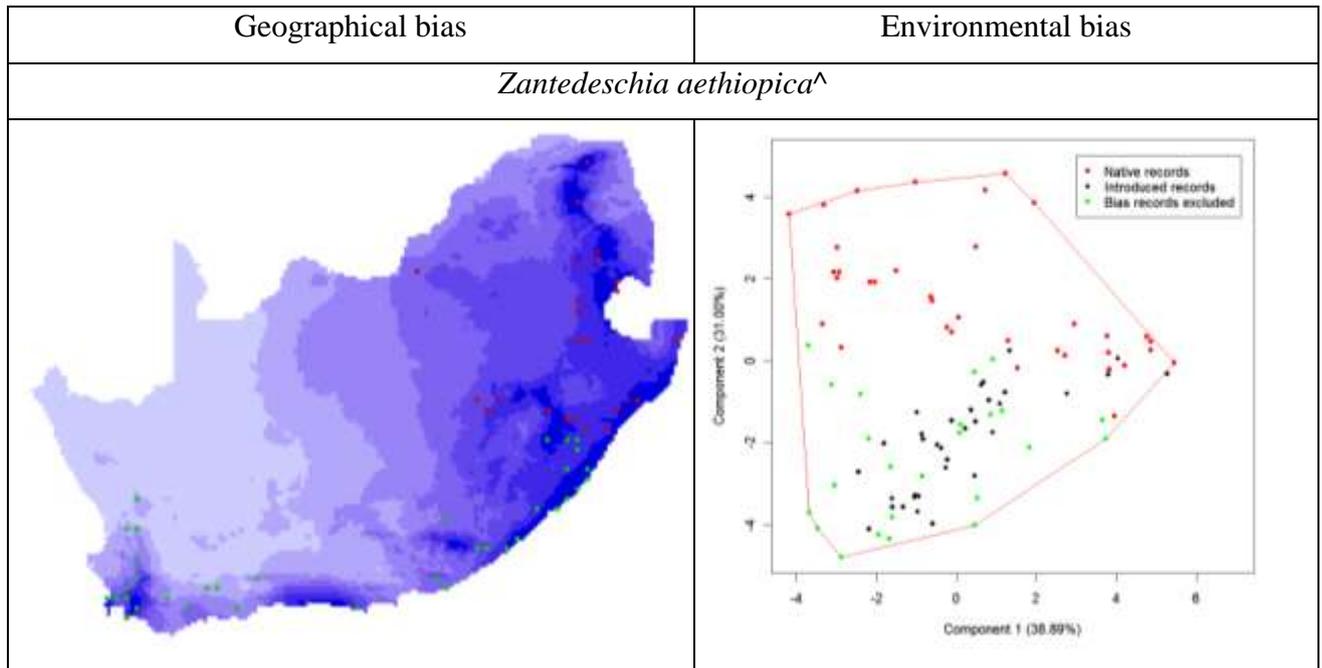


Figure S1. Diagrammatic illustration of the relationship between a) geographic and b) environmental bias for all the species (South African species invading Australia indicated with an arrow (^)). The entire native geographic distribution of the species shown across the annual precipitation gradient (BIO 12) encountered in the native range, darker colors indicate wetter areas. Red dots indicate the subset of the native records used to calibrate the models and the green dots represent the subset of native records excluded from the dataset (biased records) to simulate geographical bias (treatment A40). The environmental distribution of the species shown by the principal components analysis (PCA) of the environmental conditions associated with the native and introduced range of the species. The extent of the entire native range of the species indicated with the red polygon. In this polygon the subset of the native range records used to calibrate the models are indicated with red dots and the subset of the native range records excluded from the dataset to simulate geographical bias (biased records, treatment A40) are indicated with the green dots. The environmental distribution of the introduced range records used to evaluate the models is indicated with the black dots. The percentage of environmental variation explained by a specific component is given in parentheses.

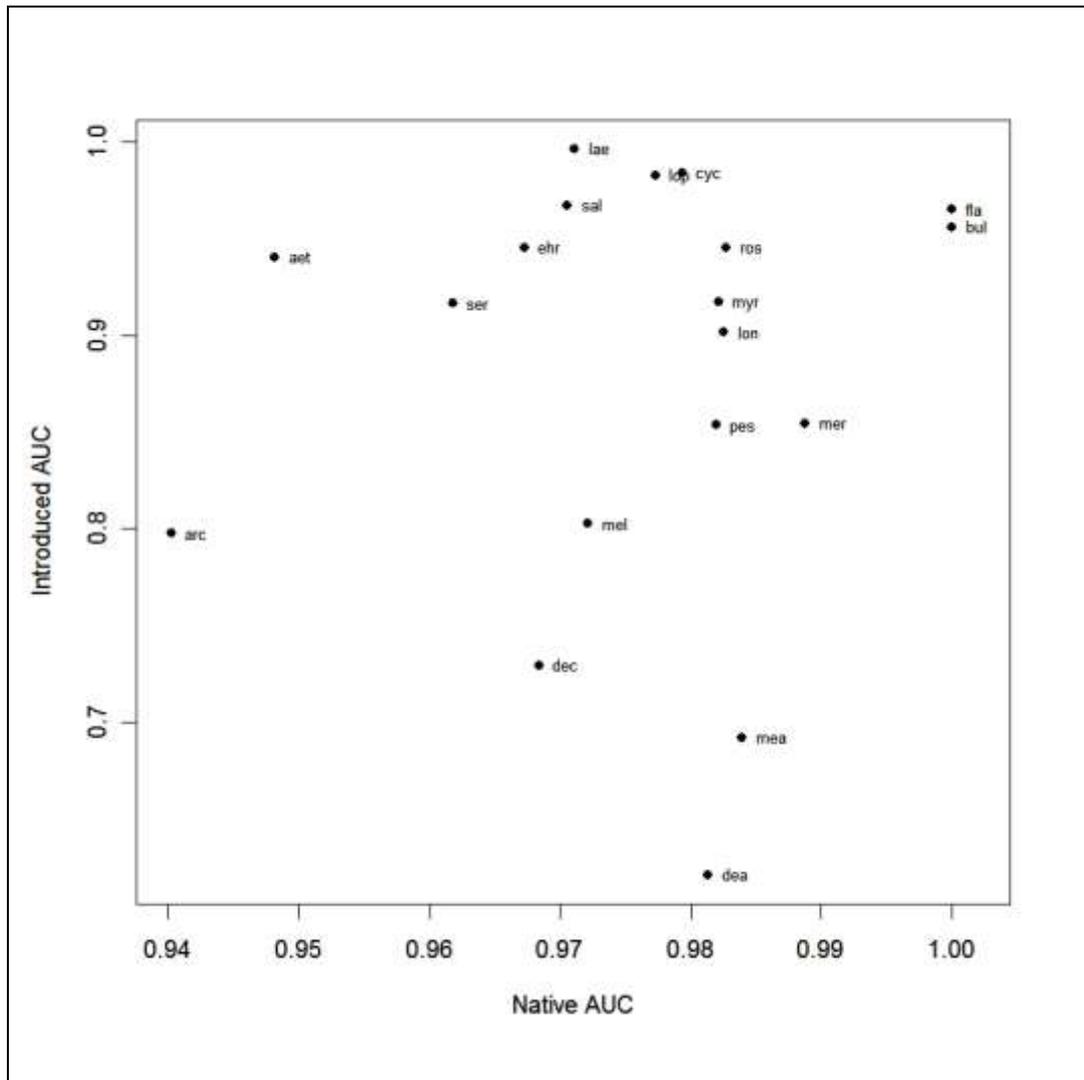


Figure S2. The relationship between native range and introduced range model performance. Introduced range models were calibrated with all the records available from the native range and then projected to the introduced range. Model performance (AUC values obtained from the conventional ROC analysis) was evaluated with a test set that comprised of presence records obtained from the introduced range and pseudo-absence records generated by randomly selecting grid cells from the map region that did not contain presence records. The number of pseudo-absence records was equal to the number of presence records in the testing set. Species codes are defined as the first three letters of the species name (e.g. ser = *Hakea sericea*) or of the genus name when the first three letters of a species name is the same (e.g. arc = *Arctotheca calendula*; ehr = *Ehrharta calycina*).

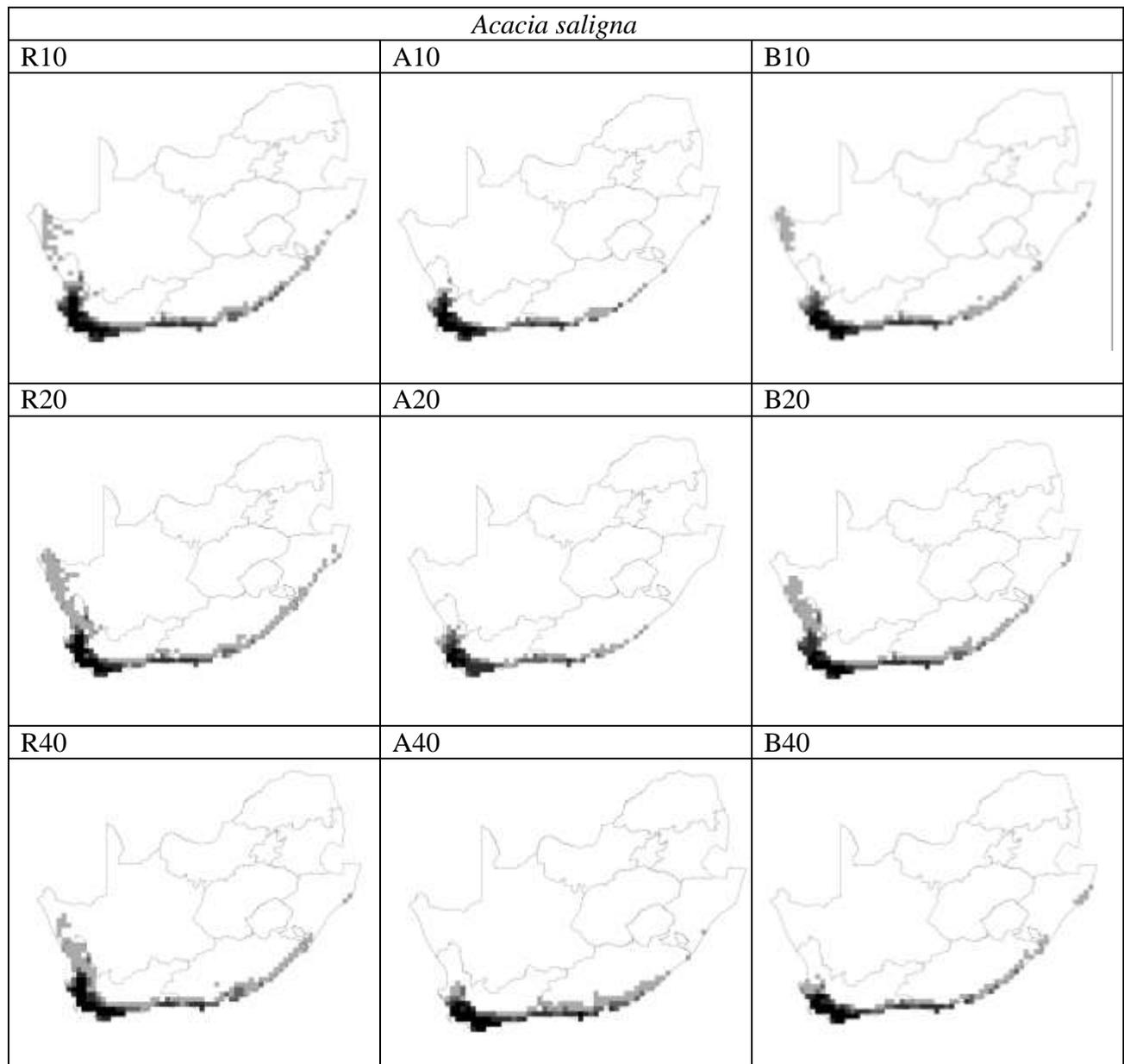


Figure S3. Introduced range predictions for *A. saligna* with all three control datasets (R10, R20 and R40), datasets from treatment A (A10, A20 and A40) as well as datasets from treatment B (B10, B20 and B40).



Chapter 3

Running title: Introduction history

Calibrating ecological niche models with native and introduced range records for invasive alien plant species under different invasion scenarios

René Wolmarans, Mark P. Robertson and Berndt J. van Rensburg

Department of Zoology and Entomology, Centre for Invasion Biology, University of Pretoria, Pretoria 0002, South Africa.

ABSTRACT

The aim of this study was to compare model performance of ecological niche models when calibrated with three different types of input data 1) with native range records 2) introduced range records or 3) a combination of records from both ranges for alien invasive plant species with different simulated introduction histories. Models were calibrated with datasets that represented these three different types of input data and seven environmental predictor variables using the MAXENT modelling algorithm. Three different introduction histories were simulated from the introduced range records available namely, a single introduction (30 records from a single location), a multiple introduction (records from three different localities adding up to 30) and a well established (all introduced records available) scenario. Model performance was not significantly different when compared between models that were calibrated with datasets representing these three types of input data under a simulated single introduction or multiple introduction scenario. However, model performance was significantly different for models calibrated with introduced range records and a combination of records from both ranges under the well established scenario. Our results indicated that models

calibrated with both native and introduced range records over-predicted the potential distribution of the species in the introduced range, yet we still suggest that models should be calibrated following this approach as an over-estimation of an invasive alien species potential distribution is less of a concern than an underestimation.

Keywords: Ecological niche models, invasive plant species, native records, introduced records, both ranges, MAXENT, model performance, introduction history.

INTRODUCTION

The geographic distribution of an organism can be predicted with the aid of distribution models and specifically, ecological niche modelling (Peterson 2001; McPherson *et al.* 2004; Guisan & Thuiller 2005; McPherson & Jetz 2007). This procedure aims to quantify the species-environment relationship by describing the association between the organism's occurrence records and the environmental characteristics at these points (Guisan & Zimmerman 2000; Guisan & Thuiller 2005). More simply, these models attempt to capture the ecological niche that a particular organism occupies.

A popular application of ecological niche models is to predict the potential distribution of invasive alien organisms in their introduced range so that areas that are at risk of invasion can be identified (Roura-Pascual *et al.* 2004; Mau-Crimmins *et al.* 2006; Schussman *et al.* 2006; Broennimann *et al.* 2007; De Meyer *et al.* 2007; Steiner *et al.* 2008). This allows for a pro-active approach to the management of invasions as opposed to a reactive approach. A wide variety of correlative modelling algorithms are currently available for this purpose (Elith *et al.* 2006). First a model describing the ecological niche of the species is produced using occurrence records and environmental predictor variables from a species' known distribution, followed by the projection of the niche model onto other regions so that areas susceptible to invasion can be identified (Peterson 2003; Roura-Pascual *et al.* 2004; De Meyer *et al.* 2007; Mgidi *et al.* 2007).

Models incorporating occurrence records that are able to quantify the geographical extent of the fundamental niche of an organism are likely to be most successful at predicting potential spread of an invasive species. The fundamental niche describes all the environmentally suitable habitats in which a species could potentially

maintain populations without interactions with other species or immigration (Peterson 2001; Peterson 2003; Pearman *et al.* 2008). However, a species' geographic distribution is influenced by a complex interaction of several main factors (Soberon & Peterson 2005). First, climate conditions might impose physiological limits on a species (abiotic factors), and second, interactions with other species could either have a positive or negative effect on their distribution (biotic factors). Finally, geographic barriers may prevent species from dispersing and colonizing all the environmentally suitable areas. These biotic and abiotic interactions result in species only occupying some parts of their fundamental niche, known as the realized niche (Pearson & Dawson 2003; Soberon & Peterson 2005). As a result, the occurrence records incorporated into distribution models already include the effect of these factors and their interaction and therefore most likely only provide information on a species' realized niche (Pearman *et al.* 2008). The extent to which the realized niche can be described by the model and extent to which the realized and fundamental niches overlap will determine the success of the model (Soberon & Peterson 2005; Phillips *et al.* 2006).

Three different types of input data can be used to calibrate models when producing potential distribution predictions in the introduced range of an invasive alien species. Occurrence records obtained from the native range of a species represents the first type of input data (Roura-Pascual *et al.* 2004; De Meyer *et al.* 2007; De Meyer *et al.* 2009). Obtaining records from herbarium collections and other data sources in a species' native range could however be time-consuming and expensive (Mau-Crimmins *et al.* 2006). Furthermore, this approach assumes that an invasive species' environmental requirements would stay unchanged in the introduced range, i.e. niche conservatism

(Wiens & Graham 2005) and could therefore be best described from distribution data obtained from its region of origin (Peterson 2003; Welk 2004). However, recent studies have shown that some invasive species occupy areas that are not environmentally similar to that of their native range, i.e. niche shifts (Mau-Crimmins *et al.* 2006; Broennimann *et al.* 2007; Fitzpatrick *et al.* 2007; Beaumont *et al.* 2009). Possible explanations for this observation could be that invasive species are released from biological enemies in the introduced range, a loss of genetic variation or hybridization in the new region (Fitzpatrick *et al.* 2008). Therefore, calibrating models with native range records only, may not provide adequate information on where a species could potentially spread to in a new region.

The second type of input data consists of records obtained from the introduced range of a species (Robertson *et al.* 2001, 2003; Rouget *et al.* 2004; Schussman *et al.* 2006; Roura-Pascual *et al.* 2008). However, when calibrating models with records from the introduced range only, uncertainties in terms of the equilibrium status and introduction history could influence data quality and thus model performance (Welk 2004; Mau-Crimmins *et al.* 2006; Loo *et al.* 2007; Wilson *et al.* 2007; Roura-Pascual *et al.* 2008). A species that has recently been introduced to a new region is unlikely to be at equilibrium with the environment as insufficient time will have elapsed to allow it to disperse to suitable areas (Welk 2004; Wilson *et al.* 2007). The occurrence records of this species would be unlikely to capture its full environmental potential and therefore underestimate the species' potential distribution. In terms of introduction history, we suggest that a set of occurrence records for a species that has recently been introduced to a single locality would likely underestimate the potential distribution of that species to a

greater extent than if the species had been introduced to several different localities simultaneously or when a species has been introduced long ago and is now considered to be well spread (well established) in the new range.

Finally, models can be calibrated with the third type of input data; a combination of records obtained from both the native range and the introduced range of a species (Broennimann & Guisan 2008; Steiner *et al.* 2008). It has been proposed that a combination of these records would describe a wider range of environmental variation in which the species can persist, since native range records describe the environmental conditions to which the species is adapted to in its region of origin and the introduced range records describes environmental conditions in which a species can persist when released from biological constraints (Broennimann & Guisan 2008; Roura-Pascual *et al.* 2008). We suggest that the geographical extent of a species fundamental niche would most likely be best described when this approach is followed and thus that models calibrated in this way would be most successful at accurately predicting a species' potential distribution.

The aim of this study was to compare the performance of ecological niche models when calibrated with (1) native range records (2) introduced range records and (3) a combination of both, for invasive alien plant species with different simulated introduction histories.

METHODS AND MATERIALS

Species selection and distribution records

We selected seven Australian plant species that have invaded South Africa and which are considered to be well established, i.e. have spread to all the environmentally suitable sites in the introduced range (Table 1). Native range occurrence records were extracted from the Australian Virtual Herbarium public access database (AVH; <http://www.anbg.gov.au/avh/>). This is an online botanical information resource, representing a collection of plant specimen data for all the major Australian herbaria. Additional occurrence records were requested from the Queensland herbarium (Paul Robins *pers. comm.*) and the Western Australia Herbarium (FloraBase database). As a result of an extensive search carried out to find all the available records, records were assumed to be representative of a species entire native distribution. Occurrence records from the introduced range were derived from two sources. First, records from the Southern African Plant Invaders Atlas (SAPIA; <http://www.agis.agric.za/>), which contains the most comprehensive set of distribution records for alien invasive plant species in southern Africa (Henderson 1998; 2006b; 2007) were used specifically for model calibration. Second, records collected from the clearing activities undertaken by the Working for Water program (WfW; <http://www.dwaf.gov.za/wfw/>) were used solely for model evaluation. Due to the precision of the occurrence records, models were calibrated and evaluated using a 15 minute grid. The coordinates of the occurrence records were used to assign each record to a grid cell in the map region. Only one record per grid cell was used when several occurrence records were present in a grid cell.

Environmental predictor variables

We used seven bioclimatic variables derived from the Worldclim database and resampled to 15 minute grids (<http://www.worldclim.org>; Hijmans *et al.* 2005). These variables included annual mean temperature (Bio 1), mean diurnal temperature range (Bio 2), maximum temperature for the warmest month (Bio 5), minimum temperature for the coldest month (Bio 6), annual precipitation (Bio 12), precipitation for the wettest month (Bio 13) and precipitation for the driest month (Bio 14).

Ecological niche modelling

Niche-based models were produced using Maxent ver. 3.1.0 (<http://www.cs.princeton.edu/~schapire/maxent/>). Maxent estimates a species distribution by finding the probability distribution of maximum entropy (i.e. closest to uniform) subject to the constraint that the mean environmental conditions predicted by the model should be close to the empirical average of the conditions at the known presence localities (Phillips *et al.* 2006; Phillips & Dudik 2008). The Maxent modelling technique has the ability to incorporate presence-only records (Elith *et al.* 2006; Gibson *et al.* 2007; Pearson *et al.* 2007 Roura-Pacual *et al.* 2008), which is an advantage when working with invasive species since true absence data are rarely available and if they are, they are usually unreliable as the species may have been recorded as absent merely because insufficient time had elapsed to allow for invasion (Wilson *et al.* 2007; Peterson *et al.* 2008). Default algorithm parameters were set to a maximum number of 500 iterations, a regularization multiplier of one, and a convergence threshold of 0.00001. In Maxent, predictor variables are transformed into feature types (linear, quadratic, product,

threshold or discrete) and may be the mean of variables, their square, and product with other variables, thresholds or binarizations of categorical variables. The auto features option was selected allowing the set of feature types used to depend on the number of presence records for that particular species. We selected the recommended logistic probability response which is a continuous variable translating to an indicator of environmental suitability (Phillips & Dudik 2008; Roura-Pascual *et al.* 2008)

Datasets representing different types of input data and simulating introduction histories

Calibration datasets were divided into three main categories to represent the three different types of input data (see Table 2). In the first category (N), only one calibration dataset was created for each species, which comprised of all the native range records available for that particular species.

In the second category (I), three calibration datasets were created for each species from introduced range records derived from the SAPIA database to simulate three different introduction scenarios. The first dataset simulated a single introduction scenario (S) and comprised the 30 occurrence records of the species that were closest to a single geographic location selected in the map region. This dataset represented the introduced range occurrence records that would be available for a species that has only recently been introduced into a new region. The second dataset within this category simulated a multiple introductions scenario (M) and comprised of occurrence records adding up to 30, which were extracted from three different geographic locations in the map region. Locations were selected as points nearest to large cities assuming that these areas had

higher propagule pressure (e.g. airports and harbours; Colautti *et al.* 2006; Herborg *et al.* 2007). This dataset represented the introduced range occurrence records that would be available for a species that has recently been introduced to several locations simultaneously in the new region. The third dataset simulated a situation where a species was considered to be well established (W) in its introduced range and comprised of all the occurrence records available from the SAPIA database for that particular species. This dataset represented the introduced range occurrence records that would be available for a species that has been introduced to a new region long-ago.

In the third category (B), three datasets were created for each species comprising of a combination of both the native range occurrence records obtained from the native dataset and introduced range occurrence records obtained from either the single introduction, multiple introductions or well established dataset.

Models

To ensure that predictions were reliable and accurate, native range predictions were produced prior to making projections to the introduced range. For each species all the native range occurrence records obtained were partitioned into a calibration set (70% of the records) and testing set (30% of the records) using a k-fold method (Pearson *et al.* 2006; Phillips *et al.* 2006) where $k=5$. For the testing set, pseudo-absences were generated by randomly selecting grid cells from the map region that did not contain presence records. For each species the number of pseudo-absence records was equal to the number of presence records in the testing set (Jiminez-Valverde & Lobo 2007; McPherson & Jetz 2007; Roura-Pascual *et al.* 2008).

Models were calibrated with the different samples of occurrence records in each category together with environmental variable values extracted at the corresponding grid cells of the calibration set (Table 2). Models were projected to the introduced range and evaluated with the independent test set that comprised of presence records obtained from the WfW data and pseudo-absence records that were generated in a similar fashion as described above.

Model evaluation

Model performance was assessed using the Area Under Curve (AUC) from receiver operating characteristic (ROC) plots (Fielding & Bell 1997; Lobo *et al.* 2008; Peterson *et al.* 2008). The ROC curve is a threshold independent measure that plots sensitivity values (true positive fraction) as a function of 1-specificity values (false positive fraction) across a series of thresholds (Fielding & Bell 1997). At every possible threshold the continuous probability score is converted to a binary value. If the predicted occurrence is above the particular threshold then it is considered to be a presence, if below the threshold then it is considered an absence. These binary transformed probabilities are then compared to the actual occurrence records from the test set, allowing for the production of a confusion matrix (Fielding & Bell 1997). A confusion matrix calculates the number of true presences (sensitivity), false positives (commission error), false negatives (omission error) and true absences (specificity) at each particular threshold (Fielding & Bell 1997). All AUC values range between 0-1, where values of 0.5 indicate that predictions are no better than random, poor when values range between 0.5-0.7, and useful if values fall between 0.7-0.9. Predictions that obtain values greater than 0.9, are

considered to be excellent (Fielding & Bell 1997; Broennimann *et al.* 2007). Additionally, we used the True Skill Statistic (TSS) instead of the popular Kappa statistic, as Kappa is influenced by prevalence (Allouche *et al.* 2006). TSS values range from -1 to +1, where +1 indicates a perfect fit and values of zero or less indicate a performance no better than random (Allouche *et al.* 2006). Since the TSS value was calculated for all possible thresholds ranging from 0-1 with set increments of 0.01, only the maximum TSS (mTSS) value for each species prediction was reported. Omission and commission error rates were calculated at the threshold that obtained this maximum TSS value (mTSS).

Analysis

In order to address the question of which type of input data is best to use when predicting the potential distribution of invasive plant species with different introduction histories, we defined treatments to simplify the analyses (see Table 2). Within each introduction scenario we used ANOVA tests to compare the average model performance values across three different treatments. The first treatment comprised of models calibrated with all the native range records available for a species and therefore these model performance values remained consistent within each introduction scenario throughout analyses. The second treatment within an introduction scenario comprised of models calibrated with introduced range records and the third treatment within an introduction scenario comprised of models calibrated with a combination of records obtained from both the native and introduced ranges. Significant differences within an introduction scenario category were identified using the Tukey HSD function in R (R Development Core Team 2004).

Similarly, omission error rate and commission error rate were also compared among treatments within an introduction scenario.

RESULTS

Average AUC values were greater than 0.7 for all species tested in the native range (Appendix A). Model performance assessed with mTSS and AUC analysis was not significantly different across treatments within the single and multiple introduction scenarios respectively (Table 3). Model performance was however significantly different across treatments within the well established scenario (Table 3). Post-hoc analyses revealed that these differences were especially large between the introduced well established (IW) and the both well established (BW) treatment when compared with the mTSS value ($p = 0.017$) and the AUC value ($p = 0.003$). On average for all species, models calibrated with the introduced range records obtained higher mTSS (Fig. 1a) and AUC values (Fig. 1b; see also Table 4) when compared across treatments within the single introduction, multiple introductions and well established scenario. However, there was more variation among the models within the single introduction scenario compared to the amount of variation within the multiple introductions and well established scenarios (Fig. 1). On average, intermediate model performance values (Fig. 1; see also Table 4) were observed for models calibrated with both native and introduced range occurrence records when treatments were compared within a specific introduction scenario.

Omission error rate was not significantly different across treatments for the single introduction scenario ($F = 0.499$, $p = 0.615$). Similar results were obtained for the

multiple introductions scenario ($H = 4.396$, $p = 0.111$) and for the well established scenario ($F = 0.679$, $p = 0.520$). On average, models calibrated with native range occurrence records achieved the lowest omission error rates when treatments were compared within the single introduction scenario (Fig. 2a). Considerable amounts of variation were evident for models of the introduced single (IS) and the both single (BS) treatments (Fig. 2a). Within the multiple introductions and well established scenarios however, lowest omission error rates were observed for models calibrated with both native and introduced range occurrence records (Fig. 2a).

No significant difference was observed when commission error rate was compared across treatments within the single introduction scenario ($F = 1.361$, $p = 0.282$) or the multiple introductions scenario ($F = 0.070$, $p = 0.933$). Commission error rate was however significantly different across treatments within the well established scenario ($F = 5.962$, $p = 0.010$). Post-hoc analyses revealed that commission error rates were on average significantly higher in the BW than the IW or N treatments. On average, models calibrated with introduced range occurrence records achieved lowest commission error rates across all three introduction scenarios (Fig. 2b). The highest commission error rates of all treatments examined was observed when models were calibrated with occurrence records from both the native and introduced ranges (Fig. 2b).

Introduced range predictions for *Acacia mearnsii* (Fig. 3) is an example of a species that showed a gradual increase in the region predicted as suitable when models were consecutively calibrated with introduced range records simulating a single introduction scenario, a multiple introduction scenario and a well established scenario.

In contrast, the region predicted as suitable remained similar when models were calibrated with a combination of both native and introduced range records regardless of the introduction scenario simulated with the introduced range records (Fig. 3). Introduced range predictions for *A. saligna* (Fig. 4) is an example of a species where the region predicted as suitable was similar for all models regardless of the records used for calibration and the introduction scenario simulated. However, the model calibrated with both native range and introduced range occurrence records simulating a well established scenario over predicted the distribution of the species to a great extent (Fig. 4, see also Table 4).

DISCUSSION

On average, we found no significant difference in the performance of models that were calibrated with either native range records, introduced range records or a combination of records from both ranges, for a simulated single introduction scenario. For this scenario, we expected the performance of models that were calibrated with only introduced range records to be weaker than models calibrated with native range records or a combination of records from both ranges. We argued that since these records represented an early stage of invasion that they would only capture part of the environmental variation that the species could tolerate in the introduced range and would therefore underestimate the potential distribution, as opposed to models calibrated with native range records or a combination of records from both ranges that would better describe the potential distribution of the species (Welk 2004; Mau-Crimmins *et al.* 2006; Roura-Pascual *et al.* 2008). This suggests that the datasets representing a single introduction or early stage of

invasion contained sufficient information to describe the niche of these species. It is likely that a larger difference would have been observed if we had used fewer occurrence records to simulate the single and multiple introduction scenarios (e.g. 10 rather than 30). However, for some species (e.g. *A. mearnsii*) model performance was weak (AUC = 0.659; mTSS = 0.211) when models were calibrated with only introduced range records. This is most likely an example of a species with a broad environmental niche in the introduced range where the environmental variation captured with the records from a single introduction was not sufficient to represent all the environmental variation that a species could potentially tolerate. This result supports the notion of Welk (2004) that reliable distribution predictions cannot be made with introduced range records alone for invasive species with wide environmental requirements, for which insufficient time has elapsed since introduction for that species to be at equilibrium in its introduced range.

Similarly, we found no significant difference in model performance between models calibrated with either native range records, introduced range records or a combination of records from both ranges under the multiple introductions scenario that was simulated. This result was however, not totally unexpected since we argued that if a species was introduced to multiple locations in a new region, the proportion of environmental variation within that introduced range would be similarly represented with the introduced range records as with the native range records or a combination of records from both ranges.

Surprisingly, for the well established scenario the performance of models calibrated with introduced range records was significantly higher than those calibrated with a combination of both native and introduced range records. This was contrary to

expectation, since we assumed that if a species was considered to be well established (advanced stage of invasion), the proportion of environmental variation represented with the records from these two datasets would be similar resulting in similar model performance. However, when considering the types of prediction errors we found that the models calibrated with records from both the native and introduced ranges obtained the highest commission error rate, indicating that these models over-predicted the potential distribution of the species relative to those calibrated with records from the introduced range only.

The over-prediction we observed here can possibly be explained by the principle on which the modelling algorithm we used is based. Maxent estimates the probability distribution of a species that is most spread out or closest to uniform (maximum entropy), subject to the constraints imposed by the information regarding the known distribution of a species and the environmental conditions prevailing in this area (Phillips *et al.* 2006; Phillips & Dudik 2008). The problem with this approach is that the Maxent model can then predict high suitability for environmental conditions in the introduced range that are outside the range present in the map region used to calibrate the models, i.e. over-predict the potential distribution in the introduced range (Maxent 2008). The ‘clamping’ feature of Maxent can be employed to solve this problem (Maxent 2008). Clamping is done during the projecting onto the introduced range process and gives rise to more conservative models (less over-prediction), since the environmental conditions predicted as suitable are now ‘clamped’ (restricted) to the range of values on which the model was calibrated (Maxent 2008). At the time of this study we did not select this feature as there was limited guidance in the literature on its application and we were concerned about

making the models too conservative. This probably explains the high commission error rates for models calibrated using native records and those using a combination of native and introduced records.

Commission errors could also occur because predictor variables such as soil characteristics were not incorporated into the models as well as biological interactions with other species that might impose limitations on the geographic extent of a species distribution in its introduced range (Welk 2004; Loo *et al.* 2007). Also, these models predict the potential distribution of the species while the evaluation dataset evaluates the realised distribution of the species (Kadmon *et al.* 2003).

A high rate of commission errors could result from models seemingly over-predicting the potential distribution of an invasive species simply because the predictions were evaluated with records obtained from the introduced range of the species. Records from the introduced range of a species are not always a true reflection of a species' spreading potential in that region. These records might be an under-representation of a species' distribution if a species is not at equilibrium because insufficient time has elapsed since its introduction into a new region to spread to all the suitable areas (Mau-Crimmins *et al.* 2006; Loo *et al.* 2007; Wilson *et al.* 2007; Broennimann & Guisan 2008). However, we took care to select species that were considered to be well established in South Africa to minimise this problem. Another important step in the design of this study was that we evaluated models with records obtained from an independent source in South Africa. This is preferable to splitting the original dataset into a training and a testing set as this would have tended to favour the models calibrated with introduced records.

When predicting the potential distribution of invasive species, the issue of genetic variation in the source populations relative to the introduced populations should be considered (Mau-Crimmins *et al.* 2006). When calibrating models with all the native range records available for a species, the potential distribution in the introduced range might be over-predicted when only a portion of the genetic variability was introduced from the native population. Although we do not know the introduction history of the species considered in this study, differences in genetic variability between native and introduced populations of this species could play a role. In addition, recent studies have suggested that niche shifts can occur across a species native and introduced range, i.e. environmental requirements are different across ranges (Fitzpatrick *et al.* 2007; Broennimann *et al.* 2007; Steiner *et al.* 2008; but see Peterson & Nakazawa 2008; Beaumont *et al.* 2009). In cases where niche shift is suspected then it would be advisable to calibrate models using records from the introduced range.

In conclusion, we demonstrated that for invasive alien plant species model performance was not significantly different when models were calibrated with native range records, introduced range records or a combination of both under a simulated single and multiple introduction scenario. Under a well established scenario models calibrated with introduced range records and a combination of records from both ranges showed a significant difference in model performance. Our findings suggest that the extent of a species environmental niche is an important determinant of the type of input data used when making range predictions for invasive species. The potential distribution of a species with a restricted environmental niche in the introduced range could more accurately be predicted with records obtained from the introduced range than a species

with a broad environmental niche (McPherson & Jetz 2007). When niche shift occurs across the ranges of an invasive species, the potential distribution predicted from introduced range records are likely to identify the new environmental requirements to a greater extent than native range records. However, when models are calibrated with introduced range records, the potential range might be underestimated since it is difficult to determine if enough time has elapsed since the introduction of the species into the new region for that records to be a representation of all the environmentally suitable sites. Also, since the origin (source population) of an invasive species is mostly unknown (Mau-Crimmins *et al.* 2006; Steiner *et al.* 2008), we therefore suggest, as have others (Welk 2004; Mau-Crimmins *et al.* 2006; Broennimann & Guisan 2008; Steiner *et al.* 2008), that when predicting the potential distribution of invasive species, models should be calibrated with a combination of native range and introduced range records. Native range records could identify areas of introduction and areas that are environmentally similar to the native range where a species is considered to be at equilibrium (Welk 2004; Broennimann & Guisan 2008), and introduced range records can identify environmentally suitable sites that a species can occupy when released from biological constraints and when niche shift has taken place (Broennimann & Guisan 2008; Steiner *et al.* 2008). Our results indicated that when models were calibrated with this type of input data, models over-predicted the potential distribution of the species in the introduced range, but considering all the possible explanations for this finding we are of opinion that an over-estimation of an invasive alien species potential distribution is still less of a concern than an underestimation. We recommend that future studies include the

‘clamping’ feature of the Maxent modelling algorithm when producing introduced range predictions.

ACKNOWLEDGEMENTS

This project was funded by the University of Pretoria and the DST-NRF Centre of Excellence for Invasion Biology. Lesley Henderson assisted with the species selection process and kindly provided the SAPIA database. We thank the Working for Water program for kindly providing records of clearing activities undertaken by them. The Council of Heads of Australian Herbaria and the custodial institutions, the Australian National Herbarium (Canberra), the National Herbarium of New South Wales (New South Wales), Northern Territory Herbarium (Darwin and Alice Springs), Queensland Herbarium (Brisbane), State Herbarium of South Australia (Adelaide), Tasmanian Herbarium (Hobart), National Herbarium of Victoria (Melbourne) and the Western Australian Herbarium (Perth) are acknowledged for providing Australian plant specimen data. Paul Robins, Queensland herbarium, Environmental Protection Agency, in charge of the HERBRECS kindly provided native range distribution data. Many thanks to Steven Phillips and Miroslav Dudik for help and advice with Maxent.

REFERENCES

ALLOUCHE, O., TSOAR, A. & KADMON, R. 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology* **43**: 1223-1232.

BEAUMONT, L.J., GALLAGHER, R.V., THUILLER, W., DOWNEY, P.O.,
LEISHMAN, M.R. & HUGHES, L. 2009. Different climatic envelopes among invasive

populations may lead to underestimations of current and future biological invasions.

Diversity and Distributions **15**: 409-420.

BROENNIMANN, O., TREIER, U.A., MULLER-SCHARER, H., THUILLER, W., PETERSON, A.T. & GUISAN, A. 2007. Evidence of climatic niche shift during biological invasion. *Ecology Letters* **10**: 701-709.

BROENNIMANN, O. & GUISAN, A. 2008. Predicting current and future biological invasions: both native and invaded ranges matter. *Biological Letters* **4**: 585-589.

COLAUTTI, R.I., GRIGOROVICH, I.A. & MACISAAC, H.J. 2006. Propagule pressure: a null model for biological invasions. *Biological Invasions* **8**: 1023-1037.

DE MEYER, M., ROBERTSON, M.P., PETERSON, A.T. & MANSELL, M.W. 2007. Ecological modelling and potential geographic distributions of Mediterranean fruit fly (*Ceratitis capitata*) and Natal fruit fly (*Ceratitis rosa*). *Journal of Biogeography* **35**: 270-281.

DE MEYER, M., ROBERTSON, M.P., MANSELL, M.W., EKESI, S., TSURUTA, K., MWAIK, W., VAYSIERES, J-F. & PETERSON, A.T. 2009. Ecological niche and potential geographic distribution of the invasive fruit fly *Bactrocera invadens* (Diptera, Tephritidae). *Bulletin of Entomological Research*. In press.

ELITH, J., GRAHAM, C.H., ANDERSON, R.P., DUDI, M., FERRIER, S., GUISAN, A., HIJMANS, R.J., HUETTMANN, F., LEATHWICK, J.R., LEHMANN, A., LI, J., LOHMANN, L.G., LOISELLE, B.A., MANION, G., MORITZ, C., NAKAMURA, M., NAKAZAWA, Y., OVERTON, J.M., TOWNSEND-PETERSON, A., PHILLIPS, S., RICHARDSON, K., SCACHETTI-PEREIRA, R., SCHAPIRE, R.E., SOBERON, J., WILLIAMS, S., WISZ, M.S. & ZIMMERMANN, N.E. 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* **29**: 129-151.

FIELDING, A.H. & BELL, J.F. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation* **24**: 38-49.

FITZPATRICK, M.C., WELTZIN, J.F., SANDERS, N.J. & DUNN, R.R. 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**: 24-33.

FITZPATRICK, M.C., DUNN, R.R. & SANDERS, N.J. 2008. Data sets matter, but so do evolution and ecology. *Global Ecology and Biogeography* **17**: 562-565.

GIBSON, L., BARRETT, B. & BURBIDGE, A. 2007. Dealing with uncertain absences in habitat modeling: a case study of a rare ground-dwelling parrot. *Diversity and Distributions* **13**: 704-713.

GUISAN, A. & ZIMMERMAN, N.E. 2000. Predictive habitat distribution models in ecology. *Ecological Modelling* **135**: 147-186.

GUISAN, A. & THUILLER, W. 2005. Predicting species distribution: offering more than simple habitat models. *Ecology Letters* **8**: 993-1009.

HENDERSON, L. 1998. Southern Africa plant invaders atlas (SAPIA). *Applied Plant Sciences* **12**: 31-32.

HENDERSON, L. 2006b. Comparisons of invasive plants in southern Africa originating from southern temperate, northern temperate and tropical regions. *Bothalia* **36**: 210-222.

HENDERSON, L. 2007. Invasive, naturalized and casual alien plants in southern Africa: a summary based on the Southern African Plant Invaders Atlas (SAPIA). *Bothalia* **37**: 215-248.

HERBORG, L.M., JERDE, C.L., LODGE, D.M, RUIZ, G.M. & MACISAAC, H.J. 2007. Predicting invasion risk using measures of introduction effort and environmental niche models. *Ecological Applications* **17**: 663-674.

HIJMANS, R.J., CAMERON, S.E., PARRA, J.L., JONES, P.G. & JARVIS, A. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* **25**: 1965-1978.

JIMINEZ-VALVERDE, A. & LOBO, J.M. 2007. Threshold criteria for conversion of probability of species presence to either-or presence-absence. *Acta Oecologia* **31**: 361-369.

KADMON, R., FARBER, O. & DANIN, A. 2003. A systematic analysis of factors affecting the performance of climatic envelope models. *Ecological Applications* **13**: 853-867.

LOBO, J.M., JIMENEZ-VALVERDE, A. & REAL, R. 2008. AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography* **17**: 145-151.

LOO, A.E., MACNALLY, R. & LAKE, P.S. 2007. Forecasting New Zealand mudsnail invasion range: model comparison using native and invaded ranges. *Ecological Applications* **17**: 181-19.

MAU-CRIMMINS, T., SCHUSSMAN, H.R. & GEIGER, H. 2006. Can the invaded range of a species be predicted sufficiently using only native-range data? Lehmann lovegrass (*Eragrostis lehmanniana*) in the southwestern United States. *Ecological Modelling* **19**: 736-746.

MAXENT 2008. Help file ver. 3.1.0(<http://www.cs.princeton.edu/~schapire/maxent/>).

McPHERSON, J.M., JETZ, W. & ROGERS, D.J. 2004. The effects of species' range sizes on the accuracy of distribution models: ecological phenomenon or statistical artifact? *Journal of Applied Ecology* **41**: 811-823.

McPHERSON, J.M. & JETZ, W. 2007. Effects of species' ecology on the accuracy of distribution models. *Ecography* **30**: 135-151.

MGIDI, T., LE MAITRE, D.C., SCHONEGEVEL, L., NEL, J.L., ROUGET, M. & RICHARDSON, D.M. 2007. Alien plant invasions – incorporating emerging invaders in regional prioritization: A pragmatic approach for Southern Africa. *Journal of Environmental Management* **84**: 173-187.

PEARMAN, P.B., GUIBAN, A., BROENNIMANN, O. & RANDIN, C.F. 2008. Niche dynamics in space and time. *Trends in Ecology and Evolution* **23**: 149-158.

PEARSON, R.G. & DAWSON, T.P. 2003. Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Global Ecology & Biogeography* **12**: 361-371.

PEARSON, R.G., THUILLER, W., ARAUJO, M.B., MARTINEZ-MEYER, E., BROTONS, L., McCLEAN, C., MILES, L., SEGURADO, P., DAWSON, P. & LEES, D.C. 2006. Model-based uncertainty in species range prediction. *Journal of Biogeography* **33**: 1704-1711.

PEARSON, R.G., RAXWORTHY, C.J., NAKAMURA, M. & PETERSON, A.T. 2007. Predicting species distribution from small occurrence records: a test case using cryptic geckos in Madagascar. *Journal of Biogeography* **34**: 102-117.

PETERSON, A.T. 2001. Predicting species' geographic distributions based on ecological niche modelling. *The Condor* **103**: 599-605.

PETERSON, A.T. 2003. Predicting the geography of species' invasions via ecological niche modeling. *The Quarterly Review of Biology* **78**: 419-433.

PETERSON, A.T. & NAKAZAWA, Y. 2008. Environmental data sets matter in ecological niche modelling: an example with *Solenopsis invicta* and *Solenopsis richteri*. *Global Ecology and Biogeography* **17**: 135-144.

PETERSON, A.T., PAPES, M. & SOBERON, J. 2008. Rethinking receiver operating characteristic analysis applications in ecological niche modeling. *Ecological Modelling* **213**: 63-72.

PHILLIPS, S.J., ANDERSON, R.P. & SCHAPIRE, R.E. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* **190**: 231-259.

PHILLIPS, S.J. & DUDIJK, M. 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* **31**: 161-175.

R Development Core Team 2004. 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>.

ROBERTSON, M.P., CAITHNESS, N. & VILLET, M.H. 2001. A PCA-based modelling technique for predicting environmental suitability for organisms from presence records. *Diversity and Distributions* **7**: 15-27.

ROBERTSON, M.P., PETER, C.I., VILLET, M.H. & RIPLEY, B.S. 2003. Comparing models for predicting species potential distribution: a case study using correlative and mechanistic predictive modelling techniques. *Ecological Modelling* **164**: 153-167.

ROUGET, M., RICHARDSON, D.M., NEL, J.L., LE MAITRE, D.C., EGOH, B. & MGIDI, T. 2004. Mapping the potential ranges of major plant invaders in South Africa,

Lesotho and Swaziland using climatic suitability. *Diversity and Distributions* **10**: 475-484.

ROURA-PASCUAL, N., SUAREZ, A.V., GOMEZ, C., PONS, P., TOUYAMA, Y., WILD, A.L. & PETERSON, A.T. 2004. Geographical potential of Argentine ants (*Linepithema humile* Mayr) in the face of global climate change. *Proceedings of the Royal Society of London B Biological Sciences* **271**: 2517-2534.

ROURA-PASCUAL, N., BROTONS, L., PETERSON, A.T. & THUILLER, W. 2008. Consensual predictions of potential distributional areas for invasive species: a case study of Argentine ants in the Iberian Peninsula. *Biological Invasions* **11**: 1017-1031.

SCHUSSMAN, H., GEIGER, E., MAU-CRIMMINS, T. & WARD, J. 2006. Spread of current and potential distribution of an alien grass, *Eragrostis lehmanniana* Nees, in the southwestern USA: comparing historical data and ecological niche models. *Diversity and Distributions* **12**: 582-592.

SOBERON, J. & PETERSON, A.T. 2005. Interpretation of models of fundamental ecological niches and species' distributional areas. *Biodiversity Informatics* **2**: 1-10.

STEINER, F.M., SCHLICK-STEINER, B.C., VANDERWAL, J., REUTHER, K.D., CHRISTIAN, E., STAUFFER, C., SUAREZ, A.V., WILLIAMS, S.E. & CROZIER, R.H. 2008. Combined modeling of distribution and niche in invasion biology: a case study of two invasive *Tetramorium* ant species. *Diversity and Distributions* **14**: 538-545.

WELK, E. 2004. Constraints in range predictions of invasive plant species due to non equilibrium distribution pattern: Purple loosestrife (*Lythrum salicaria*) in North America. *Ecological Modelling* **179**: 551-567.

WIENS, J.J. & GRAHAM, C.H. 2005. Niche conservatism: Integrating evolution, ecology, and conservation biology. *Annual Review of Ecology, Evolution and Systematics* **36**: 519-539.



WILSON, J.R.U., RICHARDSON, D.M., ROUGET, M., PROCHEs, S., AMIS, M.A.,
HENDERSON, L. & THUILLER, W. 2007. Residence time and potential range: crucial
considerations in modelling plant invasions. *Diversity and Distributions* **13**: 11-22.



Table 1. Species selected for analysis. The number of native range records and the number of introduced range records in the well established dataset (SAPIA) used for calibration and the number of introduced range records (WfW) used for evaluation.

Botanical name	Common name	Number of native records	Number of introduced records (well established)	Number of introduced records (evaluation)
<i>Acacia cyclops</i>	Red-eye	100	150	38
<i>Acacia dealbata</i>	Silver wattle	200	285	73
<i>Acacia decurrens</i>	Green wattle	101	110	37
<i>Acacia longifolia</i>	Sydney golden wattle	220	86	35
<i>Acacia mearnsii</i>	Black wattle	135	430	246
<i>Acacia melanoxylon</i>	Australian blackwood	287	138	41
<i>Acacia saligna</i>	Port Jackson willow	268	146	38



Table 2. Summary of the treatments that comprised of datasets generated with occurrence records from a specific type of input data within a specific introduction scenario. The first treatment comprised of all the native range records available for a species and therefore remained consistent within each introduction scenario.

Input data	Introduction scenario	Dataset records	Treatment
Native (N)		All native records available	N
Introduced (I)	Single (S)	30 records from a single location	IS
Both (B)	Single (S)	All native and 30 introduced records from a single location	BS
Introduced (I)	Multiple (M)	Records from three different locations adding up to 30	IM
Both (B)	Multiple (M)	All native and 30 introduced records from multiple locations	BM
Introduced (I)	Well (W)	All introduced records available	IW
Both (B)	Well (W)	All native and introduced records available	BW



Table 3. Model performance (mTSS and AUC values) compared across treatments within a specific introduction scenario. $F(2,18)$ is the proportion of variation explained between the three different treatments ($df = 2$), within subjects ($df = n-2$), $n = 21$. Significant difference $p < 0.05$.

Treatment	mTSS value		AUC value	
	F (2,18)	p - value	F (2,18)	p - value
N, IS, BS	0.226	0.800	0.536	0.594
N, IM, BM	0.256	0.777	0.559	0.581
N, IW, BW	4.746	0.022	7.863	0.004

Table 4. Model performance values (mTSS and AUC values) for models calibrated with the different datasets generated. Models calibrated with all the available native range records (Native). Models calibrated within the single introduction scenario with introduced range records (IS) and both native and introduced range records (BS). Within the multiple introductions scenario, models calibrated with introduced range records (IM) and records from both ranges (BM). Within the well established scenario, models calibrated with introduced range records (IW) and records from both ranges (BW). *Aca cyc* = *Acacia cyclops*, *Aca dea* = *A. dealbata*, *Aca dec* = *A. decurrens*, *Aca lon* = *A. longifolia*, *Aca mea* = *A. mearnsii*, *Aca mel* = *A. melanoxylon*, *Aca sal* = *A. saligna*.

Species	mTSS value							AUC value						
	Native	IS	BS	IM	BM	IW	BW	Native	IS	BS	IM	BM	IW	BW
<i>Aca cyc</i>	0.632	0.711	0.368	0.868	0.763	0.842	0.342	0.854	0.918	0.715	0.96	0.877	0.952	0.552
<i>Aca dea</i>	0.411	0.616	0.548	0.548	0.507	0.575	0.548	0.725	0.867	0.788	0.841	0.761	0.870	0.780
<i>Aca dec</i>	0.405	0.459	0.568	0.514	0.541	0.595	0.568	0.705	0.804	0.831	0.798	0.796	0.844	0.786
<i>Aca lon</i>	0.800	0.743	0.686	0.800	0.800	0.829	0.571	0.934	0.937	0.915	0.945	0.935	0.949	0.792
<i>Aca mea</i>	0.524	0.211	0.407	0.370	0.541	0.581	0.508	0.761	0.659	0.718	0.753	0.818	0.856	0.781
<i>Aca mel</i>	0.439	0.537	0.341	0.488	0.341	0.512	0.268	0.758	0.795	0.722	0.750	0.723	0.813	0.649
<i>Aca sal</i>	0.711	0.763	0.711	0.789	0.684	0.868	0.263	0.907	0.933	0.888	0.938	0.889	0.943	0.530

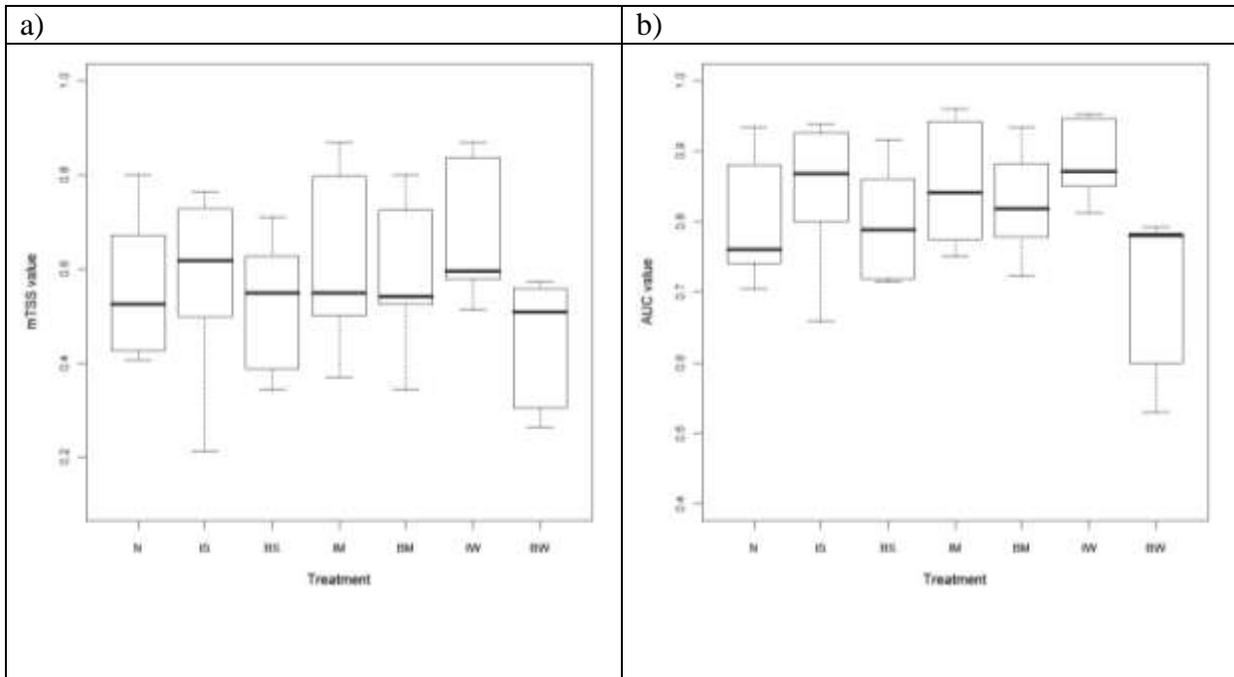


Figure 1. Model performance compared across treatments. Model performance expressed as a) maximum TSS (mTSS) value and b) AUC value. Treatment codes explained in Table 2. The box represents the inter-quartile range of the data and the whiskers show the minimum and maximum values. The horizontal line represents the median value.

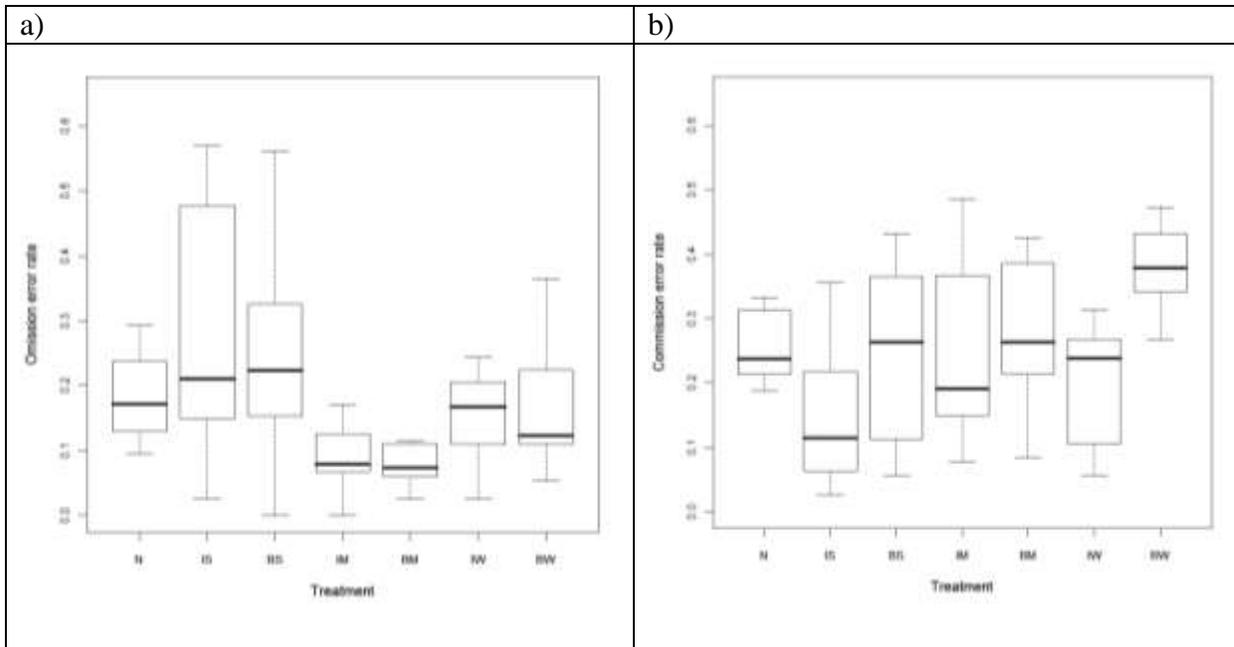


Figure 2. Prediction error rates calculated at the mTSS threshold compared across treatments. Prediction error rate expressed as a) omission error rate and b) commission error rate. Treatment codes explained in Table 2. The box represents the inter-quartile range of the data and the whiskers show the minimum and maximum values. The horizontal line represents the median value.

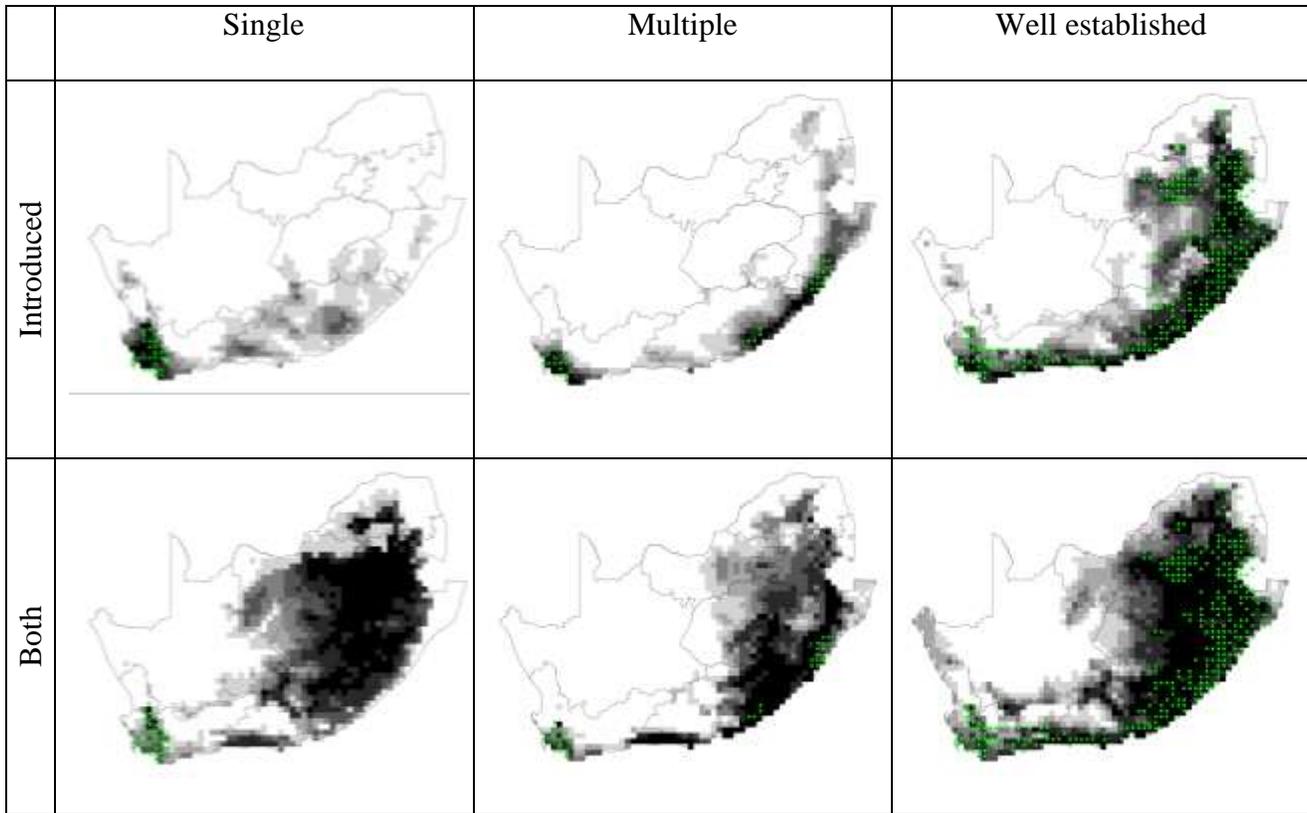


Figure 3. Introduced range predictions for *A. mearnsii*. Top panel figures represents models that were calibrated with introduced range occurrence records and lower panel figures represent models calibrated with occurrence records from both the native and introduced ranges. Each column contains the ranges predicted under a specific introduction scenario, where the green dots symbolize the introduced range records used to calibrate the models.

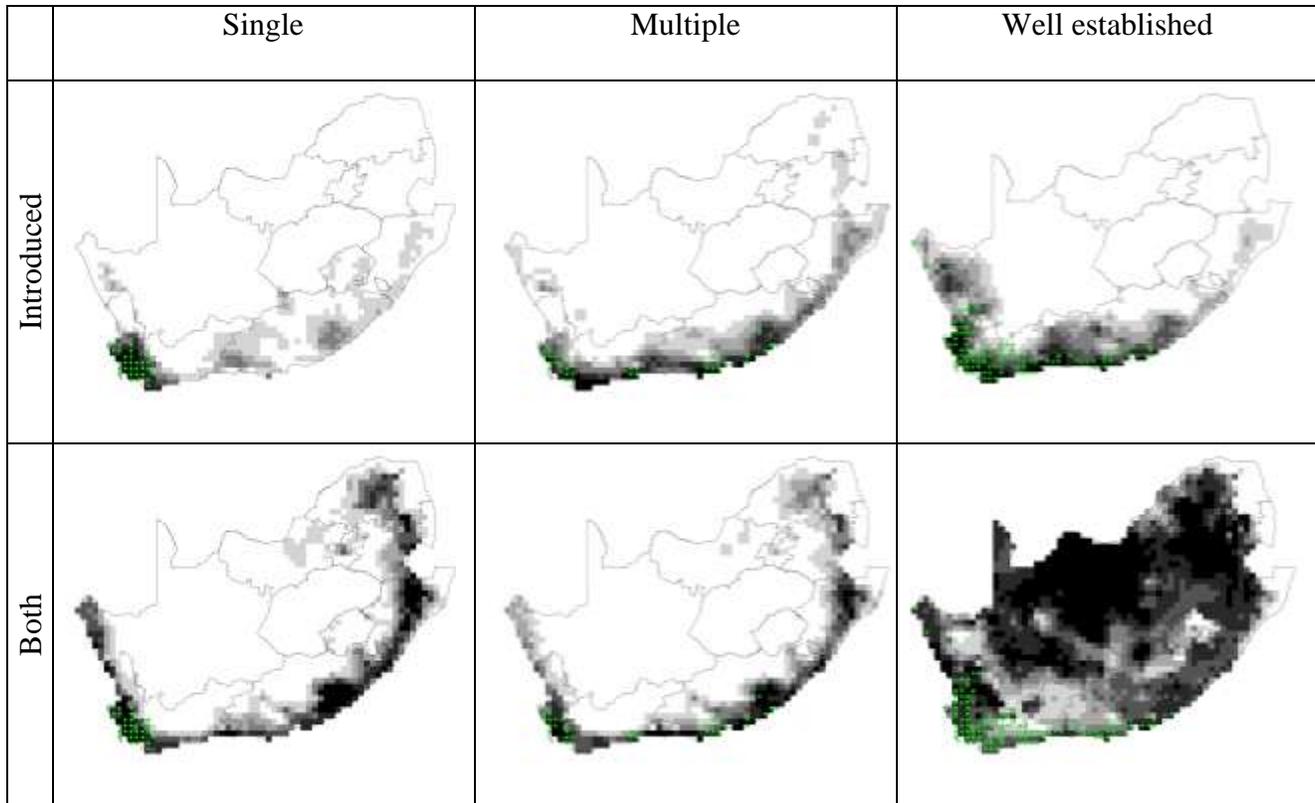


Figure 4. Introduced range predictions for *A. saligna*. Top panel figures represents models that were calibrated with introduced range occurrence records and lower panel figures represent models calibrated with occurrence records from both the native and introduced ranges. Each column contains the ranges predicted under a specific introduction scenario, where the green dots symbolize the introduced range records used to calibrate the models.



Appendix A. Model performance values (AUC values) obtained when producing native range predictions.

Species	AUC value
<i>Acacia cyclops</i>	0.854
<i>Acacia dealbata</i>	0.725
<i>Acacia decurrens</i>	0.705
<i>Acacia longifolia</i>	0.934
<i>Acacia mearnsii</i>	0.761
<i>Acacia melanoxylon</i>	0.758
<i>Acacia saligna</i>	0.907



Chapter 4

Summary and Conclusions

Summary and Conclusions

In general, one of three different types of input data is used to calibrate models when making potential range predictions for invasive alien plant species in their introduced range. In the first category, distribution models are calibrated with native range occurrence records; in the second category models are calibrated with introduced range occurrence records and in the third category with a combination of records from both the native and introduced range. The aim of this study was to evaluate how the quality of the occurrence records in these three input data categories influence ecological niche model performance and hence, the accuracy of the potential range predictions produced for invasive alien plant species by these models.

Main findings

My findings showed that model performance is determined by the extent to which the environmental (i.e. ecological) niche of a species is represented with the occurrence records used to calibrate the model, and not necessarily by the quality of the records itself. I compared model performance achieved when calibrated with ‘good’ quality data and that achieved when calibrated with ‘bad’ quality data.

Geographical bias was simulated with native range occurrence records to represent two different bias scenarios that might be encountered when calibrating models with native range records obtained from data collections. When occurrence records are geographically biased, the underlying environmental gradients in which a species can persist will most likely also not be fully sampled, which could result in environmental



bias (Raes & ter Steege 2007; Hortal *et al.* 2008). Occurrence records suffering from either one of these two forms of bias were considered to be examples of ‘bad’ quality data and were expected to negatively influence model performance and the accuracy of the potential distribution predicted for the species. The introduced range of a species was expected to be underestimated when models were calibrated with these data since it was assumed that some of the geographic areas that were excluded with the bias simulation might contain portions of the environmental gradients in which a species can persist that were not represented with the records used for the model calibration (Welk 2004; Guisan & Thuiller 2005; Hortal *et al.* 2008; Beaumont *et al.* 2009).

When producing introduced range predictions for invasive species, the success of the models depend on how the extent of a species’ environmental niche occupied in the native range overlap with the extent of the environmental niche occupied in the introduced range (i.e. niche conservatism; Wiens & Graham 2005). My results showed that even though the geographical bias that was simulated in the occurrence records did lead to significant environmental bias, model performance was not significantly different when compared between models that were calibrated with geographically biased records and models calibrated with unbiased records. However, I show that the geographic bias simulated in the native range occurrence records did not reside in this relevant region of overlap for the majority of the species studied. Therefore that the environmental niche occupied by the species in the native range was still sufficiently represented with the biased occurrence records to not significantly influence model performance.

Furthermore, three different introduction history scenarios (single introduction, multiple introductions and well established) were simulated with introduced range



occurrence records to represent different qualities of introduced range records available for invasive plant species. Models calibrated with records that represented a species that have only recently been introduced to a single locality was considered as the models calibrated with ‘bad’ quality data and was expected to underestimate the potential distribution of a species to a greater extent than occurrence records that represented a species that has been introduced to several different localities simultaneously (‘better’ quality) or occurrence records that represented a species that has been introduced long ago and considered to be well established in the introduced range (‘good’ quality). It was argued that the records that represented a single introduction (an early stage of invasion) would only capture part of the environmental variation that the species could tolerate in the introduced range and would therefore underestimate the potential distribution (Welk 2004; Mau-Crimmins *et al.* 2006; Roura-Pascual *et al.* 2008). Model performance was compared between models that were calibrated with these three different qualities of introduced range records and models calibrated with native range records or a combination of records from both the native and introduced range. Models calibrated with native range records or a combination of records from both the native and introduced ranges were expected to perform better under the single and multiple introductions scenarios as these records represented a greater extent of a species environmental niche than the models calibrated with only introduced range records. However, no significant difference was found in model performance, indicating that all the records used to calibrate these models represented sufficient proportions of the environmental niches of the species considered. In addition, no significant difference in model performance was expected under the well established scenario since it was argued

that the native range occurrence records, introduced range occurrence records and a combination of records from both ranges would all capture similar proportions of the species' environmental niche. I did however, find a significant difference in model performance between models calibrated with introduced range records and a combination of records from both the native and introduced range. When evaluating the predictions errors obtained with these models however, I found that the models calibrated with the combinations of records from both the native and introduced range obtained high commission error rates (over-prediction).

Concluding remarks

I recommend that ecological niche models for invasive alien plant species be calibrated with a combination of both native range and introduced range occurrence records (third type of input data) since I demonstrated that the quality of the occurrence records used to calibrate distribution models only play a role if the extent of the environmental niche represented by these records is influenced. Native range records could identify areas of introduction and areas that are environmentally similar to the native range where a species is considered to be at equilibrium (Welk 2004; Broennimann & Guisan 2008), and introduced range records can identify environmentally suitable sites that a species can occupy when released from biological constraints and when niche shift has taken place (Broennimann & Guisan 2008; Steiner *et al.* 2008). When niche shift occurs across the ranges of an invasive species, the potential distribution predicted from introduced range records are likely to identify the new environmental requirements to a greater extent than native range records. However, when models are calibrated with introduced



range records, the potential range might be underestimated since it is difficult to determine if enough time has elapsed since the introduction of the species into the new region for that records to be a representation of all the environmentally suitable sites. Also, since the origin (source population) of an invasive species is mostly unknown (Mau-Crimmins *et al.* 2006; Steiner *et al.* 2008) the potential distribution in the introduced range might be over-predicted when only a portion of the genetic variability was introduced from the native population. My results indicated that this approach over-predicted the potential distribution of the species in the introduced range, but when working with invasive alien plant species an over-estimation of the potential distribution is still less of a concern than an underestimation.

Strengths and weaknesses

The presence of niche shift between the native and introduced range of the species, could potentially have complicated my analyses since it could be argued that the geographical bias that was simulated in the native range occurrence records could be part of the environmental space that was actually not occupied by the species in the introduced range, hence that niche shift could mask the influence of the resulting environmental bias on model performance. However, for most species the native range records excluded from the dataset (biased records) did not reside in the region of the environmental space occupied by the introduced range records that was used for model evaluation.

The evaluation of model performance is reliant on the quality of the testing set. We evaluated models with test sets that consisted of presence records obtained from the introduced range and pseudo-absence records that were generated. Pseudo-absences



were randomly selected grid cells from the introduced map region that were not occupied by a presence record. A weakness of this approach is the implicit assumption that the records available are a true reflection of a species' distribution in the introduced range, therefore that a species is at equilibrium with its environment and that sampling effort has been thorough. Absence from a particular site might just be that not enough time has passed since the introduction of the species into a new region to spread to all the suitable areas and is not necessarily due to low environmental suitability (Mau-Crimmins *et al.* 2006; Loo *et al.* 2007; Wilson *et al.* 2007; Broennimann & Guisan 2008). This prohibits thorough testing of commission errors (over prediction) when evaluating model performance (Peterson *et al.* 2008). However, although I specifically selected species that were considered to be well established and well sampled in their introduced ranges (based on expert opinion), it is possible that some of the species may have been under sampled. Furthermore, since only reliable information with regards to the presence of an invasive species is typically available (Welk 2004; Wilson *et al.* 2007) the Maxent modelling algorithm was selected, as this is a highly regarded approach to presence-only distribution modelling (Elith *et al.* 2006; Phillips *et al.* 2006; Loiselle *et al.* 2008).

Finally, models were evaluated using three different techniques. Firstly, using the conventional Area Under Curve (AUC) from receiver operating characteristic (ROC) plots (Fielding & Bell 1997; Lobo *et al.* 2008; Peterson *et al.* 2008). Secondly, with the True Skill Statistic (TSS) instead of the popular Kappa statistic, as Kappa is influenced by prevalence (Allouche *et al.* 2006). Lastly, with a modified AUC approach proposed by Peterson *et al.* (2008), which eliminates the reliance on commission error where areas



might be classified as unsuitable based on uncertain absences i.e. pseudo-absences (Peterson *et al.* 2008).

Recommendations for future research

Since I found considerable variation among species, I recommend that future studies include and analyse more species to address the questions investigated in this thesis. Studies like these would have to rely on finding regions where species have been introduced and for which good sources of data are available. However, it may be difficult to find a sufficiently large number of species that meet the requirement of being considered well established.

It would be interesting to investigate the effect of introduction history and different stages of invasion on model performance by using actual case studies from around the world where good records of the invasion history and how it has progressed through time have been collected.

This study was conducted on a relatively coarse scale (15 minute) and finer scale studies may yield different results since different factors influence species' distribution at finer scales e.g. at very broad scales climate conditions are important driving forces of species' distributions but at finer scales topography and biotic interaction becomes increasingly important (Pearson *et al.* 2002; Welk *et al.* 2002; Pearson & Dawson 2003; Guisan & Thuiller 2005; Soberon & Peterson 2005; Phillips *et al.* 2006).

A wide variety of modelling techniques are currently available (Guisan & Zimmermann 2000; Elith *et al.* 2006) and comparing the outcome of correlative models with techniques likely to be less sensitive to bias, e.g. CLIMEX (Climex 2007) would



yield valuable results. Understanding how the clamping feature of MAXENT influences commission error rates is a vital question that needs to be addressed.

Finally, future studies should also evaluate how model performance for invasive alien species is influenced when ecological niche models are calibrated with introduced range occurrence records obtained from more than one introduced range.

References:

ALLOUCHE, O., TSOAR, A. & KADMON, R. 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology* **43**: 1223-1232.

BEAUMONT, L.J., GALLAGHER, R.V., THUILLER, W., DOWNEY, P.O., LEISHMAN, M.R. & HUGHES, L. 2009. Different climatic envelopes among invasive populations may lead to underestimations of current and future biological invasions. *Diversity and Distributions* **15**: 409-420.

BROENNIMANN, O. & GUISAN, A. 2008. Predicting current and future biological invasions: both native and invaded ranges matter. *Biological Letters* **4**: 585-589.

CLIMEX 2007. CLIMEX version 3 User's Guide. Sutherst, R.W., Maywald, G.F. & Kriticos, D. CSIRO, Hearne Scientific Software Pty Ltd.

ELITH, J., GRAHAM, C.H., ANDERSON, R.P., DUDI, M., FERRIER, S., GUISAN, A., HIJMANS, R.J., HUETTMANN, F., LEATHWICK, J.R., LEHMANN, A., LI, J., LOHMANN, L.G., LOISELLE, B.A., MANION, G., MORITZ, C., NAKAMURA, M., NAKAZAWA, Y., OVERTON, J.M., TOWNSEND-PETERSON, A., PHILLIPS, S., RICHARDSON, K., SCACHETTI-PEREIRA, R., SCHAPIRE, R.E., SOBERON, J., WILLIAMS, S., WISZ, M.S. & ZIMMERMANN, N.E. 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* **29**: 129-151.

- FIELDING, A.H. & BELL, J.F. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation* **24**: 38-49.
- GUISAN, A. & ZIMMERMAN, N.E. 2000. Predictive habitat distribution models in ecology. *Ecological Modelling* **135**: 147-186.
- GUISAN, A. & THUILLER, W. 2005. Predicting species distribution: offering more than simple habitat models. *Ecology Letters* **8**: 993-1009.
- HORTAL, J., JIMINEZ-VALVERDE, A., GOMEZ, J.F., LOBO, J.M. & BASELGA, A. 2008. Historical bias in biodiversity inventories affects the observed environmental Niche of the species. *Oikos* **117**: 847-858.
- LOBO, J.M., JIMENEZ-VALVERDE, A. & REAL, R. 2008. AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography* **17**: 145-151.
- LOISELLE, B.A., JORGENSON, P.M., CONSIGLIO, T., JIMINEZ, I., BLAKE, J.G., LOHMANN, L.G. & MONTIEL, O.M. 2008. Predicting species distributions from herbarium collections: does climate bias in collection sampling influence model outcomes? *Journal of Biogeography* **35**: 105-116.
- LOO, A.E., MACNALLY, R. & LAKE, P.S. 2007. Forecasting New Zealand mudsnail invasion range: model comparison using native and invaded ranges. *Ecological Applications* **17**: 181-19.
- MAU-CRIMMINS, T., SCHUSSMAN, H.R. & GEIGER, H. 2006. Can the invaded range of a species be predicted sufficiently using only native-range data? Lehmann lovegrass (*Eragrostis lehmanniana*) in the southwestern United States. *Ecological Modelling* **19**: 736-746.

PEARSON, R.G., DAWSON, T.P., BERRY, P.M. & HARRISON, P.A. 2002. SPECIES: A spatial evaluation of climate impact on the envelope of species. *Ecological Modelling* **154**: 289-300.

PEARSON, R.G. & DAWSON, T.P. 2003. Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Global Ecology & Biogeography* **12**: 361-371.

PETERSON, A.T., PAPES, M. & SOBERON, J. 2008. Rethinking receiver operating characteristic analysis applications in ecological niche modeling. *Ecological Modelling* **213**: 63-72.

PHILLIPS, S.J., ANDERSON, R.P. & SCHAPIRE, R.E. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* **190**: 231-259.

RAES, N. & TER STEEGE, H. 2007. A null-model for significance testing of presence only species distribution models. *Ecography* **30**: 727-736.

ROURA-PASCUAL, N., BROTONS, L., PETERSON, A.T. & THUILLER, W. 2008. Consensual predictions of potential distributional areas for invasive species: a case study of Argentine ants in the Iberian Peninsula. *Biological Invasions* **11**: 1017-1031.

SOBERON, J. & PETERSON, A.T. 2005. Interpretation of models of fundamental ecological niches and species' distributional areas. *Biodiversity Informatics* **2**: 1-10.

STEINER, F.M., SCHLICK-STEINER, B.C., VANDERWAL, J., REUTHER, K.D., CHRISTIAN, E., STAUFFER, C., SUAREZ, A.V., WILLIAMS, S.E. & CROZIER, R.H. 2008. Combined modeling of distribution and niche in invasion biology: a case study of two invasive *Tetramorium* ant species. *Diversity and Distributions* **14**: 538-545.



WELK, E., SCHUBERT, K. & HOFFMANN, M.H. 2002. Present and potential distribution of invasive garlic mustard (*Alliaria petiolata*) in North America. *Diversity and Distributions* **8**: 219-233.

WELK, E. 2004. Constraints in range predictions of invasive plant species due to non equilibrium distribution pattern: Purple loosestrife (*Lythrum salicaria*) in North America. *Ecological Modelling* **179**: 551-567.

WIENS, J.J. & GRAHAM, C.H. 2005. Niche conservatism: Integrating evolution, ecology, and conservation biology. *Annual Review of Ecology, Evolution and Systematics*, **36**: 519-539.

WILSON, J.R.U., RICHARDSON, D.M., ROUGET, M., PROCHE, S., AMIS, M.A., HENDERSON, L. & THUILLER, W. 2007. Residence time and potential range: crucial considerations in modelling plant invasions. *Diversity and Distributions* **13**: 11-22.