used for model development and possibly the complexity of the resulting model. However, practitioners are well advised that increased model complexity will not necessarily lead to better or more accurate models.

Modelling resources

Books

Goodenough, K. and McKinion, J. (eds) (1992). Basics of insect modelling. American Society of Agricultural Engineers. ASAE Monograph Number 10.

Grace, J. (2006). Structural equation modelling and natural systems. (Cambridge Press).

Papjorgji, P. and Pardalos, P. (eds) (2009). Advances in modelling agricultural systems. (Springer).

Peart, R. and Shoup, W. (eds) (1998). Agricultural systems modelling and simulation. (CRC Press).

Vohnout, K. (2003). Mathematical modelling for systems analysis in agricultural research. (Elsevier).

Journals

Computers and electronics in agriculture. http://www.elsevier.com/wps/find/journaldescription.cws_home/503304/description#description. (Elsevier).

Ecological modelling. http://www.elsevier.com/wps/find/journaldescription.cws_home/503306/description. (Elsevier).

Environmental modelling and software. http://www.elsevier.com/wps/find/journaldescription.cws_home/422921/description#description. (Elsevier).

Journal of biological dynamics. http://www.tandf.co.uk/journals/titles/17513758.asp. (Taylor and Francis).

Conferences

IASTED International Conference on Modelling and Simulation. Annual. MS 2009: http://www.iasted.org/CONFERENCES/home-670.html.

International Conference on User Modelling, Adaptation, and Personalization. Biennial in odd-numbered years. UMAP 2009: http://umap09.fbk.eu/.

MODSIM – International Conference on Modelling and Simulation. Biennial in odd-numbered years. MODSIM 09: http://www.mssanz.org.au/modsim09/.

Weeds in a warmer world: predicting the impact of climate change on Australia's alien plant species using MaxEnt

Peter D. Wilson[A], **Paul O. Downey**[B], **Michelle Leishman**[A], **Rachael Gallagher**[A], **Lesley Hughes**[A] and **Jessica O'Donnell**[A]

[A] Department of Biological Sciences, Macquarie University, New South Wales 2109, Australia.

[B] Pest Management Unit, Parks and Wildlife, New South Wales Department of Environment and Climate Change, PO Box 1967, Hurstville, New South Wales 1481, Australia.

Summary

Australia is now host to many thousands of introduced plant species, and about 3000 of these have established self-sustaining populations in the wild. Of these, approximately 450 are classified as invasive weeds nationally or regionally, and are being targeted with control measures. Two questions arise in the context of climate change: what changes might occur in the distribution of the 450 species known to be highly invasive, and which species in the pool of 3000 might emerge to become future serious pests. We are applying an advanced modelling tool, MaxEnt, to provide a strategic overview of a large portion of the 450 highly invasive species. Preliminary results suggest differing responses of weed species in northern and southern Australia linked to predicted major shifts in rainfall pattern.

Introduction

Since the earliest European contact with the Australian continent in the late 16th Century, a steady stream of introduced species have been entering Australia's ecological communities with varying impacts on the native biota. Several thousand plant species have been introduced deliberately either as agricultural or horticultural stock, or as ornamental garden plants (Groves *et al.* 2005) or accidentally. Through complex pathways involving genetic change, developmental responses to new environments and chance dispersal events, a portion of these introduced plants have become established as self-sustaining populations in the wild. Current estimates suggest that at least 3000 plant species have become naturalized in Australia (Groves *et al.* 2003) with approximately 450 of these now classified as highly invasive pests. The balance of the 3000 naturalized species represents a pool from which it is likely that new invasive species might emerge, especially given that many of these are garden escapes and still available for sale or grown in gardens (Groves *et al.* 2005).

The mechanisms by which an introduced plant species makes the transition following introduction to establish self-sustaining wild populations are complex and not fully understood. They appear to involve a combination of genetic changes, phenotypic and developmental changes, and ecological interactions (both positive and negative) within the new environment. It is clear that the climate experienced by populations of a species is a powerful driver of ecological and micro-evolutionary processes. Climate directly influences a species' establishment, growth, reproduction, and survival. Climate also has indirect influences on invasive species via its impact on species within the ecological communities of which an invasive species is a part. Climate data is thus an obvious candidate to use as a surrogate for detailed ecological models of how a species responds to its environment, with the added advantage that modelled climate data is available in GIS coverage for current climate conditions. Future climate models also allow us to make predictions of distributions under certain constraints or caveats.

Given the paucity of detailed genetic, physiological or population data for the majority of the 3000 naturalized species, other forms of inference about likely changes in distribution and abundance are required. Species distribution models (SDMs) represent one tool that may assist our management of invasive plants. Not only can SDMs guide our understanding of current distributions and the response of species to the cumulative influences of past conditions, they offer the prospect of some degree of prediction under novel environmental conditions such as climate change.

The motivation for our project is twofold. First, we wish to investigate through experimental methods the way in which key groups of invasive plant species will respond to climate change, particularly to increased CO_2 concentration. This work will inform later efforts to correlate predictions of shifts in distribution with species traits. Second, we are concerned with providing a firmer foundation for strategic planning at broad spatial scales. That is, we aim to provide information for strategists and decision-makers on the likely changes in threat levels from the large pool of naturalized plant species in Australia. We are using SDMs coupled with predicted climate, supplemented with other environmental data as appropriate, to identify species likely to undergo large shifts in distribution and potentially in abundance. Our work is collaboration between Macquarie University and the New South Wales Department of Environment and Climate Change supported by an ARC Linkage Grant.

The purpose of this paper is to describe our approach to species distribution modelling using MaxEnt, and to provide an overview of emerging patterns in the predicted change in species distributions.

Modelling using MaxEnt

A large armoury is now available to anyone wishing to develop SDMs (Elith *et al.* 2006). MaxEnt (Phillips *et al.* 2006, Phillips *et al.* 2009) is a recent addition that is being widely adopted around the world to model species distributions. Its growing popularity is due to its flexibility and ease of use regarding data types (continuous versus categorical), its ability to deal with interactions between variables and to incorporate information on ecological processes and species interactions provided only that these data can be pre-processed into GIS grids for incorporation into the environmental input set.

The term 'MaxEnt' has several meanings and a short outline of the method's history will help clarify the way in which the term is used for species distribution modelling. MaxEnt, which is a contraction of 'maximum entropy', is an approach to finding the most parsimonious probability distribution for the states of entities in a complex ensemble of interacting entities given a set of constraints. It began in thermodynamics and statistical physics in the 1950s, but has since been adapted and applied to a vast array of systems including language processing and linguistics, image analysis and reconstruction, complex telecommunication networks, and fitting complex models in many branches of biology, chemistry, economics and psychology. In the context of SDM, it has been applied to finding the probability distribution of the occurrence of a species given a set of environmental conditions that has the maximum support from the environmental evidence and the minimum difference from a uniform probability distribution. MaxEnt is also the name given to the software package written by Steven Phillips and collaborators to implement the MaxEnt method for SDMs. From now on whenever the term MaxEnt is used in this paper, I will be referring to the MaxEnt software developed by Phillips and co-workers.

In large comparative trials, MaxEnt has been shown to be a consistently good modelling tool (Elith *et al.* 2006, Graham and Hijmans 2006, Hernandez *et al.* 2006) and ranks alongside other newer methods such as boosted regression trees (BRTs) (De'Ath 2007, Elith *et al.* 2008) and multi-variate adaptive regression splines (MARS) (Leathwick *et al.* 2006). MaxEnt has been shown to perform well in the face of spatial uncertainty in species location records, and with as few as four occurrence records, making it a valuable tool for modelling large numbers of species with widely varying spatial accuracy and numbers of occurrence records. Naturally, however, a model based on highly inaccurate location data or few location records must be considered with scepticism, but it is still possible extract some indication of the relationship between a species and its environment using MaxEnt when other methods would struggle.

One of the distinct advantages of MaxEnt is that the mathematics implemented in the software package has been subjected to rigorous analysis (Dudík *et al.* 2004, Dudík *et al.* 2005, Dudík *et al.* 2007) which is informed by a vast literature on the technical and practical aspects of implementing the maximum entropy method in many areas as noted earlier. It shares this property with only a few other approaches to species distribution modelling (e.g. generalized linear modelling or GLM, generalized additive modelling or GAM).

Additional benefits of the MaxEnt package are it is free, it has a friendly user interface, it is not platform-specific (exactly the same software will run on any computer supporting Sun Microsystems' free Java Runtime environment (JRE) which includes Windows, Mac, Unix and Linux systems), and it has a very powerful batch-mode interface allowing large ensembles of species, perhaps using alternate environmental data sets, to be modelled efficiently. This means that, provided species distribution data and environmental data are mutually accessible and of good quality, allowing a number of people to easily cross-check or validate existing SDMs. Such an open and transparent approach to SDM promises a great deal with respect to the quality of predictions being made for invasive plant species.

It is extremely important when viewing the results of SDMs to remember that these are only models. They are always inaccurate simply because they are an approximation to complex reality, but may be found useful or informative in relation to some question regarding the way a species is responding to environmental conditions (Box 1979). Maximizing informativeness or usefulness is a key challenge for any modelling task including SDM. The models will only be as good as the quality of the data fed into the modelling tool, and of the characteristics of the methodology used to build a model.

The MaxEnt software has a number of features that provide measures of model usefulness. It has a function to run multiple replications of a model leaving out random location records to provide sensitivity or error maps. It also provides three methods of assessing model quality: regularized gain, area under the curve (AUC) of a receiver operating characteristic plot, and clamping maps. Regularized gain is a measure of the goodness-of-fit of a model adjusted for the number of features (derived from the raw environmental variables). It is therefore equivalent in nature to the Akaike's Information Criterion or AIC figure of merit for generalized linear or generalized additive models (Quinn and Keough 2002). The second measure, the AUC statistic, measures the quality of a fitted model when calculated for the training data set, and a measure of the quality of prediction for novel environments.

An additional feature of the MaxEnt software that is vitally important for its use in predicting distributions under climate change is the clamping feature. When asked to 'project' a model fitted on current environmental data (a process referred to as 'training' the model) into novel environments, the software highlights those parts of the predicted distribution that represent environmental conditions not experienced in training the model. Predictions of occurrence in geographic regions with high levels of clamping are naturally to be treated with considerable caution.

MaxEnt was explicitly developed for, and is routinely used with, presence only data. That is, records of species occurrence collected or collated in ways where there is not information on species absence. So-called 'presence-absence' data (more correctly it is 'present-not present' data) comes from designed sampling projects using standardized sampling protocols. However, the motivation for using presence-only tools such as MaxEnt is to tap into the large body of data held by museums, herbaria, informal literature and so forth. To build SDMs in these circumstances requires the provision of information plausibly associated with the general environment of the species but in which the species has not be recorded. The terms 'pseudo-absences' or 'background' points are used interchangeably but background points has become the dominant term and is used in the context of MaxEnt.

Selecting the geographical scope of the region within which random background points are selected to represent the environmental background can affect model quality. If the range of environments sampled to create the background is too broad, then the model may be overfitted. If the range of background environments is not broad enough then the predicted distribution may be uninterpretable. Guidelines for the appropriate selection of background points are still a matter of active research for both MaxEnt, and other tools such as GAMs when they are applied to presence-only data. For MaxEnt guidelines for background selection are an active area of research with results now appearing in the literature (VanDerWal *et al.* 2009).

Sampling biases in space and time can lead to considerable errors in predicted distributions no matter what modelling method is used (Dudík *et al.* 2005, Phillips *et al.* 2009, MacKenzie *et al.* 2006). MaxEnt provides two pathways to guide the training of a model in the presence of sampling bias. First, a bias density map can be provided to MaxEnt giving a weighting to grid cells according to the degree of sampling bias expected in each (Dudík *et al.* 2005). The second is to use target group selection of background points (Phillips *et al.* 2009). Here spatially well-distributed sample locations of ecologically or behaviourally similar species (the 'target group') at which the model species was not found may be used as background points on which the model is trained. The target group should show similar sampling biases to the species being modelled. This method of bias adjustment is, of course, restricted to those situations where reasoned arguments can be made to identify target species and where there are sufficient sample locations.

Finally, the *only* constraints on the data used in the MaxEnt package are that it is available in a GIS raster or grid format, and all data layers are on the same grid over the same geographical region. Environmental data can therefore be climate, presence of an obligate species (e.g. pollinator), soil type, an index of adjacency to watercourses, topographic ruggedness, and so forth, at any spatial scale that is appropriate to the questions being asked. The statistical distribution of the data is not relevant in most circumstances because the software can accept any combination of continuous or categorical variables measured on any scale.

Example applications

As an illustration of the results being produced by our SDM work using MaxEnt, we present a summary of our models for two invasive grasses in Australia: (1) a tropical species, *Andropogon gayanus* Kunth (Gamba grass) and (2) *Nassella neesiana* Trin. & Rupr.) Barkworth (Chilean needle grass). The current and future distribution of these species was modelled using the full set of 19 bioclim variables (Busby 1991, Nix 1986). Current climate data were obtained from the WorldClim website (www.worldclim.org) where 'current climate' refers to average climate conditions from 1960 to 2000. Future climate data was the CSIRO Mk3 Global Circulation Model used in the IPCC Third Assessment Report, also available at the WordClim website. Location data was obtained from Australia's Virtual Herbarium (www.anbg.gov.au/avh) and the Global Biodiversity Information Facility or GBIF (www.gbif.org).

MaxEnt models for each species indicate contrasting responses to climate change with the tropical species increasing its distribution southwards but also increasing the area of highly suitable climate within its limits of distribution (Figure 1). These results supports calls to have *Andropogon gayanus* acknowledged as a significant weed in the tropical savannas of Australia (Rossiter *et al.* 2003), especially given its interaction with fire and the likely increase in fire frequency under climate change (Pitman *et al.* 2007).

The temperate grass species, on the other hand, shows a southward trend but an overall decrease in highly suitable habitat (Figure 2). These results highlight (1) that not all weed species are going to increase in distribution as a result of climate change; (2) that targeted strategic control in the southern part of its range now will be extremely useful at reducing future impacts; and (3) surveillance for new infestations in Tasmania should be given high priority.

Conclusions

MaxEnt has performed well, enabling many species to be modelled efficiently using ecologically meaningful environmental data. Our preliminary work, involving modelling the distribution of a number of invasive grasses, indicates that predicted changes under climate change show a general trend towards increased suitability of environments for already highly invasive species in northern Australia. In contrast, established species in southern Australia may face reduced environmental suitability and are predicted to contract in distribution. Given the satisfactory results from the first phase of modelling, we are in the process of extending modelling to a much wider list of invasive plants to test the validity of these preliminary findings.

References

Box, G.E.P. (1979). Robustness in the strategy of scientific model building. *In* 'Robustness in statistics', eds R.L. Launer and G.N. Wilkinson. (Academic Press, New York).

Busby, J.R. (1991). BIOCLIM – A bioclimate analysis and prediction system. *In* 'Nature conservation: cost effective biological surveys and data analysis', eds C.R.

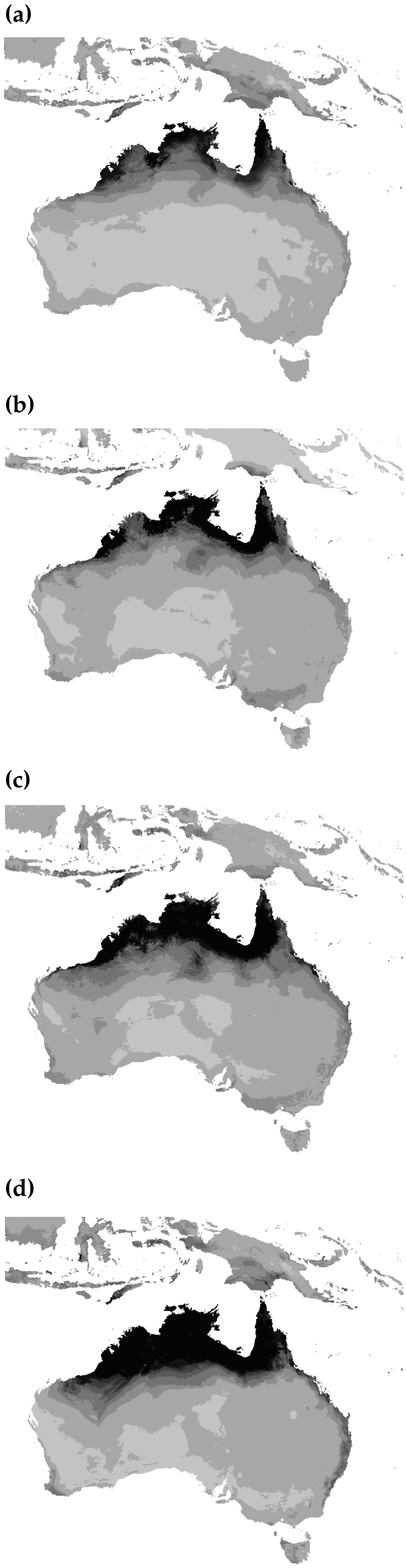

Figure 1. MaxEnt models of the distribution of Gamba grass *Andropogon gayanus*. (a) Current climate, (b) Climate in 2020, (c) 2050, and (d) 2080. Darkest shading indicates highest levels of suitable conditions for the species.

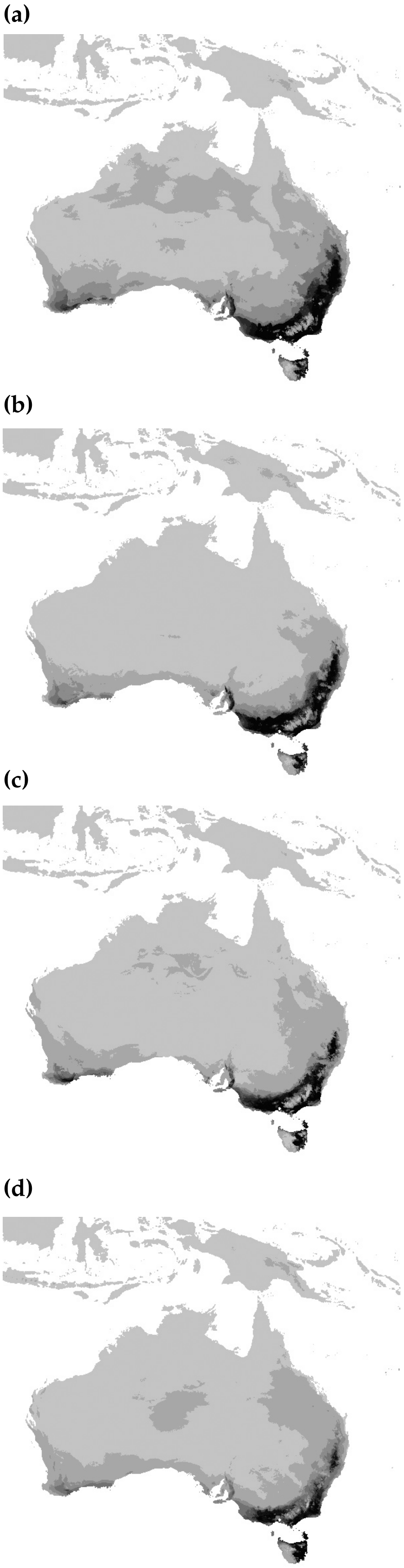

Figure 2. MaxEnt models of the distribution of Chilean needle grass *Nassella neesiana*. (a) Current climate, (b) Climate in 2020, (c) 2050, and (d) 2080. Darkest shading indicates highest levels of suitable conditions for the species.

Margules and M.P. Austin. (CSIRO, Canberra).

De'Ath, G. (2007). Boosted regression trees for ecological modelling and prediction. *Ecology* 88, 243-51.

Dudík, M., Phillips, S.J. and Schapire, R.E. (2004). Performance guarantees for regularized maximum entropy density estimation. Proceedings of the 17th Annual Conference on Computational Learning Theory.

Dudík, M., Phillips, S.J. and Schapire, R.E. (2007). Maximum entropy density estimation with generalized regularization and an application to species distribution modelling. *Journal of Machine Learning Research* 8, 1217-60.

Dudík, M., Schapire, R.E. and Phillips, S.J. (2005). Correcting sample selection bias in maximum entropy density estimation. *Advances in Neural Information Processing Systems* 18.

Elith, J., Graham, C.H., Anderson, R.P., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A.L.J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M., Peterson, A.T. and Phillips, S.J. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29, 129-51.

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

Graham, C.H. and Hijmans, R.J. (2006). A comparison of methods for mapping species ranges and species richness. *Global Ecology and Biogeography* 15, 578-87.

Groves, R.H., Boden, R. and Lonsdale, W.M. (2005). Jumping the garden fence: invasive garden plants in Australia and their environmental and agricultural impacts. WWF-Australia, Sydney.

Groves, R.H., Hosking, J.R., Batianoff, G.N., Cooke, D.A., Cowie, I.D., Johnson, R. W., Keighery, G.J., Lepschi, B.J., Mitchell, A.A., Moerkerk, M., Randall, R.P., Rozefelds, A.C., Walsh, N.G. and Waterhouse, B.M. (2003) Weed categories for natural and agricultural ecosystem management. Bureau of Rural Sciences, Canberra.

Hernandez, P.A., Graham, C.H., Master, L.L. and Albert, D.L. (2006). The effect of sample size and species characteristics on performance of different species distribution modelling methods. *Ecography* 29, 773-85.

Leathwick, J.R., Elith, J. and Hastie, T. (2006). Comparative performance of generalized additive models and multivariate adaptive regression splines for statistical modelling of species distributions. *Ecological Modelling*.

MacKenzie, D.I., Nichols, J.D., Royle, J.A., Pollock, K.H., Bailey, L.L. and Hines, J.E. (2006). 'Occupancy estimation and modelling', 1st edition. (Academic Press, Burlington, Massachusetts).

Nix, H.A. (1986). A biogeographic analysis of the Australian elapid snakes. *In* 'Atlas of Elapid snakes', ed. R. Longmore, pp. 4-15. (AGPS, Canberra).

Phillips, S.J., Anderson, R.P. and Schapire, R.E. (2006). Maximum entropy modelling of species geographic distributions. *Ecological Modelling* 190, 231-59.

Phillips, S.J., Dudík, M., Elith, J., Graham, C.H., Lehmann, A., Leathwick, J. and Ferrier, S. (2009). Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications* 19, 181-97.

Pitman, A.J., Narisma, G.T. and McAneney, J. (2007). The impact of climate change on the risk of forest and grassland fires in Australia. *Climatic Change* 84, 383-401.

Quinn, G.P. and Keough, M.J. (2002). 'Experimental design and data analysis for biologists', 1st edition. (Cambridge University Press, Cambridge).

Rossiter, N.A., Setterfield, S.A., Douglas, M.M. and Hutley, L.B. (2003). Testing the grass-fire cycle: alien grass invasion in the tropical savannas of northern Australia. *Diversity and Distributions* 9, 169-76.

VanDerWal, J., Shoo, L.P., Graham, C. and Williams, S.E. (2009). Selecting pseudo-absence data for presence-only distribution modelling: How far should you stray from what you know? *Ecological Modelling* 220, 589-94.