The art of modelling range-shifting species

Jane Elith[1]*, Michael Kearney[2] and Steven Phillips[3]

[1]*School of Botany, The University of Melbourne, Parkville 3010, Australia;* [2]*Department of Zoology, The University of Melbourne, Parkville 3010, Australia* and [3]*AT&T Labs – Research, 180 Park Avenue, Florham Park, NJ 07932, USA*

Summary

1. Species are shifting their ranges at an unprecedented rate through human transportation and environmental change. Correlative species distribution models (SDMs) are frequently applied for predicting potential future distributions of range-shifting species, despite these models' assumptions that species are at equilibrium with the environments used to train (fit) the models, and that the training data are representative of conditions to which the models are predicted. Here we explore modelling approaches that aim to minimize extrapolation errors and assess predictions against prior biological knowledge. Our aim was to promote methods appropriate to range-shifting species.

2. We use an invasive species, the cane toad in Australia, as an example, predicting potential distributions under both current and climate change scenarios. We use four SDM methods, and trial weighting schemes and choice of background samples appropriate for species in a state of spread. We also test two methods for including information from a mechanistic model. Throughout, we explore graphical techniques for understanding model behaviour and reliability, including the extent of extrapolation.

3. Predictions varied with modelling method and data treatment, particularly with regard to the use and treatment of absence data. Models that performed similarly under current climatic conditions deviated widely when transferred to a novel climatic scenario.

4. The results highlight problems with using SDMs for extrapolation, and demonstrate the need for methods and tools to understand models and predictions. We have made progress in this direction and have implemented exploratory techniques as new options in the free modelling software, MAXENT. Our results also show that deliberately controlling the fit of models and integrating information from mechanistic models can enhance the reliability of correlative predictions of species in non-equilibrium and novel settings.

5. *Implications.* The biodiversity of many regions in the world is experiencing novel threats created by species invasions and climate change. Predictions of future species distributions are required for management, but there are acknowledged problems with many current methods, and relatively few advances in techniques for understanding or overcoming these. The methods presented in this manuscript and made accessible in MAXENT provide a forward step.

Key-words: cane toad, changing correlations, climate change, extrapolation, invasive species, niche models, novel environments, species distribution models

Introduction

An increasing number of taxa are undergoing significant range shifts in response to human-assisted dispersal and changes in environmental factors, notably climate (Parmesan 2006). Often these range shifts are into novel environmental space, from both biotic and abiotic perspectives. Correlative occur-

rence-based approaches are most commonly applied to the problem of species distribution modelling (Thuiller *et al.* 2008; Elith & Leathwick 2009), but range-shifting species create two main problems for them: (1) the species records no longer reflect stable relationships with environment, and (2) environmental combinations in future scenarios will not have been adequately sampled (Menke *et al.* 2009). Thus while range-shifting taxa are often the species for which predictions of potential distributions are needed most, they most seriously violate the equilibrium assumption and often require some

*Correspondence author. E-mail: j.elith@unimelb.edu.au
Correspondence site: http://www.respond2articles.com/MEE/

degree of model extrapolation. Clearly, such species represent serious challenges to the field of species distribution modelling (Araújo & Pearson 2005; Thuiller *et al*. 2005b; Dormann 2007; De Marco, Diniz-Filho & Bini 2008).

This begs the question, should correlative models be used at all for range-shifting species? Alternative approaches based explicitly on known mechanisms (Kearney & Porter 2009) are likely to be robust under new environmental combinations in new locations but are limited by the availability of data for model parameterization and because their success in predicting range limits relies on the identification of key limiting processes. By contrast, data required to fit correlative models are widely available at different scales and the models can implicitly capture many complex ecological responses. Because of this, we anticipate ongoing use of correlative models for range-shifting species.

There is no doubt that in using species distribution models (SDMs) for extrapolation we are using them in risky ways; so, our approach is to determine the safest way to proceed. Others have considered the same general problem (e.g. Heikkinen *et al*. 2006; Hijmans & Graham 2006; Pearson *et al*. 2006; Ficetola, Thuiller & Miaud 2007; Buisson *et al*. 2009). One popular technique is to generate an ensemble of predictions based on the standard application of several different modelling methods (Thuiller 2004; Araújo & New 2007; Marmion *et al*. 2008; Roura-Pascual *et al*. 2009), so that the final prediction emphasizes agreement of predictions, and model-based uncertainty can be quantified. However, these are not problem free. Unless the candidate set of models are carefully constructed and evaluated, some lack of congruence may be more due to model error (i.e. specification of an unrealistic model) than uncertainty about the correct model. Alternatively, all models can be wrong in the same way, for example, because the species is not in equilibrium; so, agreement of models does not guarantee correctness. Moreover, there may be *a priori* knowledge of the biology of the organism or the nature of the data that render a particular modelling strategy preferable to others. In this study we instead explore a strategy of interrogating models to assess their behaviour under different data treatments and judging performance based on biological legitimacy.

We develop the approach using the case of an invasive species, the cane toad (*Bufo marinus*) which is spreading rapidly across Australia since its introduction in 1935 (Phillips, Chipperfield & Kearney 2008). The cane toad in Australia provides an informative case for exploring these issues, in part because it has previously been modelled in several different ways with qualitatively different outcomes. These include a bioclimatic envelope approach (van Beurden 1981) and a logistic regression model (Urban *et al*. 2007) based on the current range, a hybrid ecophysiological∕correlative method CLIMEX (Sutherst, Floyd & Maywald 1995) based on the native range and Australian occurrences, and a mechanistic model (Kearney *et al*. 2008) based on physiological tolerances. The potential distributions under current climates predicted by these models broadly coincide across eastern and northern Australia but differ in their predictions for southern areas. These differences are problematic for monitoring and management and raise the

question: what differences in the model or data drive the differences in prediction? How do such models behave for prediction to changed climates? In exploring these issues with the cane toad, we develop tools and techniques that are generally relevant to modelling range-shifting species with correlative approaches. We propose that model uncertainty can be reduced substantially by using ecological and physiological knowledge coupled with model exploration tools to guide model development and evaluation.

## Materials and methods

SPECIES DATA

We are interested in the general problem of modelling species not at equilibrium; so, we focus on the invaded range, where the cane toad has not yet reached all suitable environments. The species data were identical to those used in Urban *et al*. (2007) except they included 270 additional records of occurrence collected in 2006 (Fig. 1a and b), and we reduced locally dense sampling by thinning the records to one per 5-km-by-5-km grid cell. In total there were 1183 presence records and 451 absence records.

PREDICTOR VARIABLES

Eight predictor variables were chosen that had some postulated connection to the ecological requirements of the cane toad, and for which pairwise Pearson correlations between variables was less than 0·85 (Booth, Niccolucci & Schuster 1994; Elith *et al*. 2006): annual mean temperature (clim1), temperature isothermality (clim3), temperature seasonality (clim4), maximum temperature of the warmest month (clim5), mean temperature of the wettest quarter (clim8), annual precipitation (clim12), precipitation of the warmest quarter (clim18) and mean humidity of the warmest quarter (humidity). These were derived at 0·05º (∼5 km) resolution from the ANUCLIM (ANU 2009) software package, with the humidity layer being based on dry- and wet-bulb temperatures (Kearney *et al*. 2008) with a linear 4-week interpolation. We used the moderate climate change scenario presented in Kearney *et al*. (2008) (SRES marker scenario B1mid, CSIRO mk2), obtained from the software package OZCLIM (http://www.csiro.au/ozclim, last accessed January 2010). A more extreme scenario was obtained by linearly extrapolating the predicted changes for each variable at each grid cell by inflating the change threefold, leading to increases in annual mean temperate above current of 2·8–5·4 °C, and a scenario we will call 20xx.

MODELLING METHODS

We chose four modelling methods from those currently used for predicting distributions of species (Table 1). Each has a regression-like structure – i.e. additive terms within a linear predictor, and most are capable of fitting complex surfaces. With the settings we used (Table 1) boosted regression tree (BRT) and MAXENT may fit complex models; with other settings GAMs are also potentially complex. Species at equilibrium tend to be well modelled by complex surfaces (Elith *et al*. 2006), but it is possible that simpler models are more appropriate for range-shifting species. To test this, we fitted the BRT and MAXENT models with settings found reliable for fitting current distributions (Elith, Leathwick & Hastie 2008; Phillips & Dudik 2008; note use of only hinge features for MAXENT), and also fitted smoother models (Hastie, Tibshirani & Friedman 2009) by increasing the
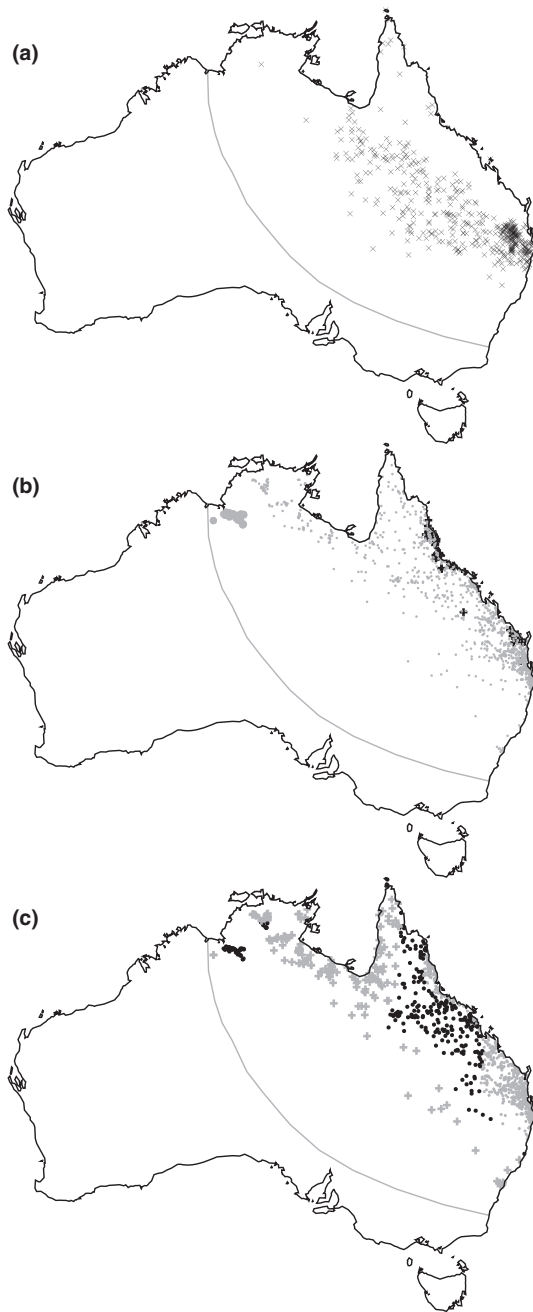
Fig. 1. Absences (a) and presences (b) of the cane toad in Australia. In (b), the presence records highlighted in large grey circles are the 2006 records (towards the northwest corner), and black: early records (1930s and 1940s, near the eastern coast). The line across the continent shows the estimated maximum 'possible spread'. (c) Weights applied to the data [size of circle indicates relative weightings with the smallest (1–5) and largest (15–20) in grey (circle and cross respectively) and the rest, black.

regularization for MAXENT and using only the early trees in the forward stage-wise fit of the BRT (for examples, see Appendix S2). Deciding how smooth to make the fitted functions is not a precise science because we cannot test fit on new unsampled environments; so, we explored a range of settings and visually assessed the effect on the partial dependence plots (Hastie, Tibshirani & Friedman 2009) of the models. We then chose settings that limited locally complex fits.

DATA TREATMENTS

Species distribution models assume that records represent a species at equilibrium with its environment. In this section we consider the cane toad data and ask: how can we adjust them to better mimic equilibrium? The cane toad has been in Australia since 1935 and records are dense along the Queensland coast (particularly around the original release areas; Fig. 1b), and relatively sparse in newly invaded areas. We therefore:

1. Upweight records with few neighbours in geographic space using a bias grid in MAXENT and case weights in regression. Methods for producing these grids and weights are detailed in Appendix S1.
2. Tested three options for representing absence or background, detailed below.

The regression methods (GLM, GAM and BRT) model species data using a binomial error model, and either need true absences or a 'background' sample of the environments in the region against which to compare the presence records (Phillips *et al.* 2009). MAXENT uses a background sample in computing the maximum entropy distribution (Phillips, Anderson & Schapire 2006). One option for the regression methods was to use the recorded absences, as in Urban *et al.* (2007, fig. 1a). Our concern was that these were a snapshot in time and might confound 'absent because unsuitable' with 'absent because beyond the current invasion front'. They also survey only a relatively small proportion of Australia, meaning that models may require substantial extrapolation for prediction across the continent.

We not only used these recorded absence data for the regression models but also tested two alternatives applicable to all four modelling methods. In the first, background was sampled across all of Australia. This represents what will happen if MAXENT or any program that samples background from the prediction extent is used without a mask and allowed to select its own background samples. It implies that the entire region has been available to the species and to those collecting survey records. As an alternative we tested a mask for indicating how far the species could have reached if conditions were suitable for its survival. We used the distance from the early coastal records (NE Australia) to the most recent records (NW Australia) as the maximum distance, and drew a polygon using that distance in all directions on land but reducing it towards the south to allow for the slower hopping speed of the toad in colder conditions (Kearney *et al.* 2008; Fig. 1). Ten thousand samples were randomly placed within the mask for this 'reachable' treatment, and 20 000 in the 'all-Australia' background treatment. The differing numbers reflect a decision to provide approximately equal spatial densities of sampling. They have no impact on the actual probabilities predicted by regression models because weights were applied to the absences; so, the total weight across all presence records equals the weight across all absences. Similarly, the way that MAXENT adjusts the scaling of the logistic output means that the number of background points does not affect the estimated probabilities.

INTEGRATING MECHANISTIC MODELS AND CORRELATIVE MODELS

Mechanistic modelling approaches are presently uncommon for animals but are becoming more feasible (Kearney & Porter 2009), raising the possibility of combining them with correlative models to strengthen predictions (Kearney *et al.* 2008; and for a plant example see Morin & Thuiller 2009). A mechanistic (ecophysiological) model exists for the cane toad (Kearney *et al.* 2008). This quantifies the influence of climate and topography on the thermal potential for above-ground activity in adults, as well as the constraints of pond

Table 1. Details of settings used in fitting the models

Method	Key reference, and model fitting details (representing common approaches)
GLM, generalized linear model	McCullagh & Nelder (1989). Each predictor allowed to be excluded or included as a linear or quadratic term. All possible combinations of the eight or fewer predictors and their linear or nonlinear terms tested, and the model with the lowest AIC selected using code written by JE for searching all combinations. Models run in R v. 2.9.0 (R Development Core Team 2009)
GAM, generalized additive model	Hastie & Tibshirani (1990). Models selected using a both-direction stepwise algorithm relying on AIC to compare models, and allowing either exclusion of a variable or a fit with 1, 2, 3 or 4 degrees of freedom. Models run in R using the 'gam' library, v. 1.0 (Hastie 2008)
BRT, boosted regression trees	Friedman, Hastie & Tibshirani (2000). Tree complexity of five (five nodes), and learning rate set to achieve at least 1000 trees in the model. Models run in R using the 'gbm' library v.1.6.3 (Ridgeway 2007) and custom code written by Leathwick and Elith (Elith, Leathwick & Hastie 2008). Models used as is and also simplified to smoother ones by using only the first 150–200 trees for prediction (Appendix S2)
MaxEnt, maximum entropy model	Phillips, Anderson & Schapire (2006). MaxEnt v 3.3.1 (Phillips & Dudík 2008) used with default settings except: only hinge features allowed, and 10-fold cross-validation; either (a) defaults for regularization, or (b) multiplier of 2·5 to fit more general models (Appendix S2)

duration and temperature on larval development and survival. No distribution data are used in the model; so, the model for current climates can be evaluated against known occurrences of the toad, which it predicts well. ==We explore two proposed methods for integrating these mechanistic predictions with the correlative modelling approach (Kearney *et al.* 2008): (1) deriving absence points from regions predicted as unsuitable by a mechanistic model and (2) using output layers of mechanistic models as inputs for correlative models.== For both we trained models using a GAM fitted as previously (Table 1). For (1) we sampled biophysical predictions of breeding season length (0–12 months; Kearney *et al.* 2008), taking 10 000 samples and using a sampling probability based on the inverse of the squared (season length + 1). This emphasized areas totally unsuitable for breeding and placed 90·5% of records in cells predicted to have 0 months suitable, 4·5% in cells with 1 month, 2·3% in cells with 2 months suitable and dwindling numbers upwards. We used the known presence data, weighted, and the eight climatic predictors described earlier. For (2) we made a new candidate set of predictors using, from the mechanistic model: (i) the potential for adult movement, and (ii) conditions permitting larval development (Kearney *et al.* 2008); then adding those climate variables not highly correlated with the two mechanistic products (i.e. all except clim4). We again used the weighted presence data and background samples within reachable areas.

MODEL EVALUATION

A substantial difficulty in modelling species not at equilibrium is that model evaluation tends to the subjective. What is the relevant measure? 'Truth' (the final, future distribution of the species) is generally not available, except in retrospective studies (Araújo *et al.* 2005) or simulations (Scheller & Mladenoff 2005; De Marco, Diniz-Filho & Bini 2008; Zurrell *et al.* 2009). Is fit to the training data, prediction of current data held out for evaluation, or agreement between predictions relevant? (Araújo, Thuiller & Pearson 2006; Broennimann & Guisan 2008). A model may fit the current distribution well, as assessed through the fit to training or testing data, yet have properties that lead to poor predictions in other times or places. We focus here on creating multiple lines of evidence for assessing the models.

(a) *Predictive performance on known data*. This is a common first step for assessing how well the model predicts the current distribution of the species, and uses withheld parts of whatever data were available

for modelling. We used 10-fold cross-validation and assessed predictive performance on the held-out folds using the area under the receiver operating characteristic curve (AUC; Hanley & McNeil 1982), a measure of the ability of the predictions to discriminate presence from absence (or background). This test mimics typical evaluation of these models. It will not be a consistent measure across models fitted on different data sets because these will have different data available to them.

(b) *Variable importance, fitted functions and maps*. Simply fitting a model and making predictions does not provide enough information to understand the underlying basis of the prediction. We interrogated our models in a number of ways. Variable importance was assessed for all models by dropping terms and noting the change in deviance or gain, and additionally for BRT and MaxEnt using measures supplied with the software and described by Elith, Leathwick & Hastie (2008) and Phillips (see online tutorial for MaxEnt at http://www.cs.princeton.edu/˜schapire/MaxEnt/). All models and mapped predictions were visually assessed for features that might indicate causes for concern. Fitted functions that were very complex or that increased or decreased sharply at the edges of the sampled environmental range, or produced complex patchy mapped distributions in regions where models were extrapolating (see below) were viewed as potential problems. In the absence of tools to interrogate the causes behind predictions at any given site we improvised. Particular sites were targeted, the environments sampled, position on the fitted functions assessed and information on the relative importance of variables used to gain an understanding of what was driving predictions. As this proved useful and provided insights, we programmed into MaxEnt two new features: one that enables easy access to site-based information on these components of the prediction and another that maps the most limiting variable in each predicted grid cell (for details, see Appendix S3).

(c) *Analysis of extent of extrapolation using multivariate environmental similarity surfaces*. Our models were required to predict to places and times not sampled in the training data, motivating the need for a measure of the similarity between the new environments and those in the training sample. We programmed a method that measures the similarity of any given point to a reference set of points, with respect to the chosen predictor variables. It reports the closeness of the point to the distribution of reference points, gives negative values for dissimilar points and maps these values across the whole prediction region (for more details, see Appendix S3). The calculation is
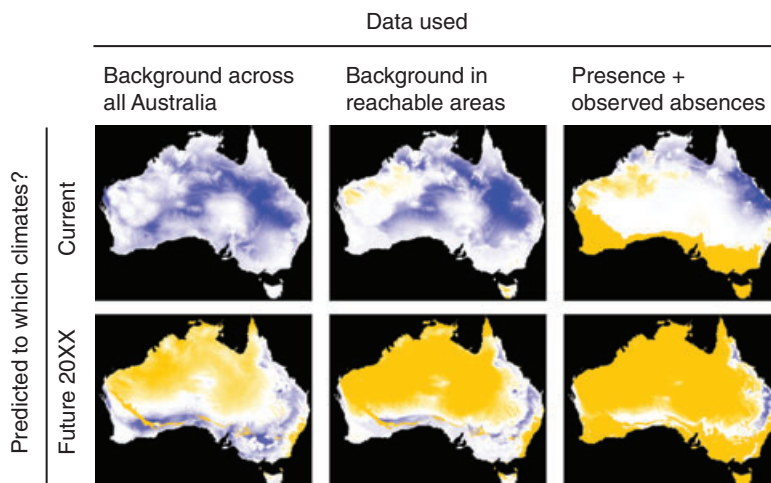
similar to that in a BIOCLIM model (Busby 1991) but extended to differentiate levels of dissimilarity when outside the range of the reference points. An accompanying map shows the variable that drives the multivariate environmental similarity surface (MESS) value in each grid cell. These methods are incorporated into the most recent version of MaxEnt (version 3.3.2, and for code, see Appendix S3). We supplied the model training data (presences plus absence, or background samples) as input points and estimated similarities across all Australia for current and changed climates. We also explored changes in correlations between variables using Pearson correlation coefficients and other new methods described in Appendix S4.

(d) *Comparison with output from a mechanistic model.* The physiologically based mechanistic model of Kearney *et al.* (2008) described earlier provides an independent set of predictions against which to compare the SDMs. Whilst the predictions of the mechanistic model are not verifiable for future times, they have a strong physiological basis, are unlikely to overestimate the potential range and are likely to give particularly strong inference about unsuitable locations. For this study we used the mechanistic model to predict to both climate scenarios (2050 and 20xx) and compared these with predictions of the correlative models, calculating three statistics to quantify goodness of fit. A Pearson correlation coefficient (COR) measures the strength of the relationship between a pair of mechanistic and other predictions; Kulczynski's coefficient (KUL) is a measure of dissimilarity that pays more attention to agreement of high values than to agreement of zeros – we used the symmetric form (Faith, Minchin & Belbin 1987) and calculated it in R (R Development Core Team 2009) using the 'gdist' function in the 'mvpart' library v.1.2.6, and standardizing all predictions to the same maximum value; AUC (see point a, above) was calculated using thresholded mechanistic predictions as the observation of presence or absence (AUC_{mech}), using a threshold of 3 months suitable for breeding (Kearney *et al.* 2008) as the minimum suitable for species persistence.
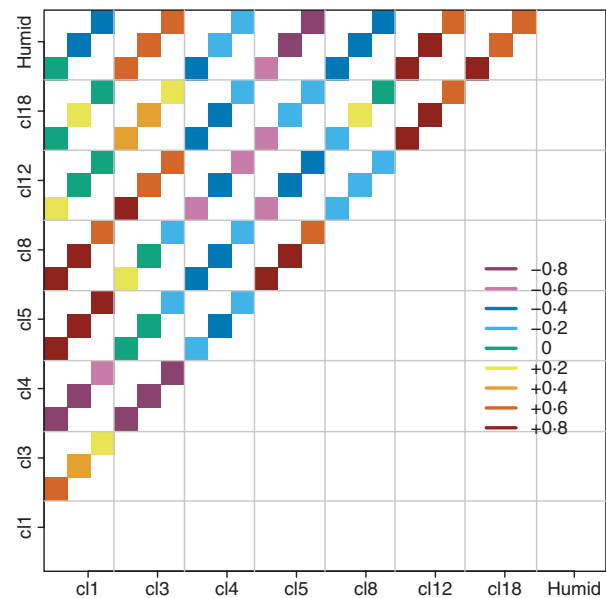
Results

EXTRAPOLATION AND MULTIVARIATE ENVIRONMENTAL SIMILARITY SURFACE

All predictions required extrapolation, with the most extreme cases being associated with the most geographically restricted data sets (those with the observed absences), and predicting into changed climates (Fig. 2). As the results for 2050 were

Fig. 3. Pairwise correlations (Pearsons *r*) between variables in space and time. Within each grid cell (i.e. the cell bounded by grey lines) the lower left correlation is for current climates, within the reachable areas; the central tile is for current climates but across all Australia, and the top right, for 20xx climates across the continent. The legend shows the midpoint value for each colour.

muted versions of 20xx, and those for 20xx most instructive, we only report current and 20xx results. Correlations between pairs of variables did not change substantially over time or region when assessed by a summary statistic (Fig. 3), although some did vary spatially (for spatially explicit investigations into correlations between pairs and suites of variables, see Appendix S4).

CURRENT CLIMATE

Whilst modelling method did affect predicted potential distributions for current climate, differences between methods were usually relatively minor (Fig. 4, left column of maps, rows 2–5 and Appendix S5A) and were mostly characterized by differing spatial patterns of high and low predicted probabilities of

Fig. 2. Comparison of climates, using all variables and the multivariate environmental similarity surface (MESS) methods programmed in MaxEnt (for method details, see Appendix S3). Legend: blue, positive; white, around zero; orange, negative; with more intense colours indicating more extreme values.
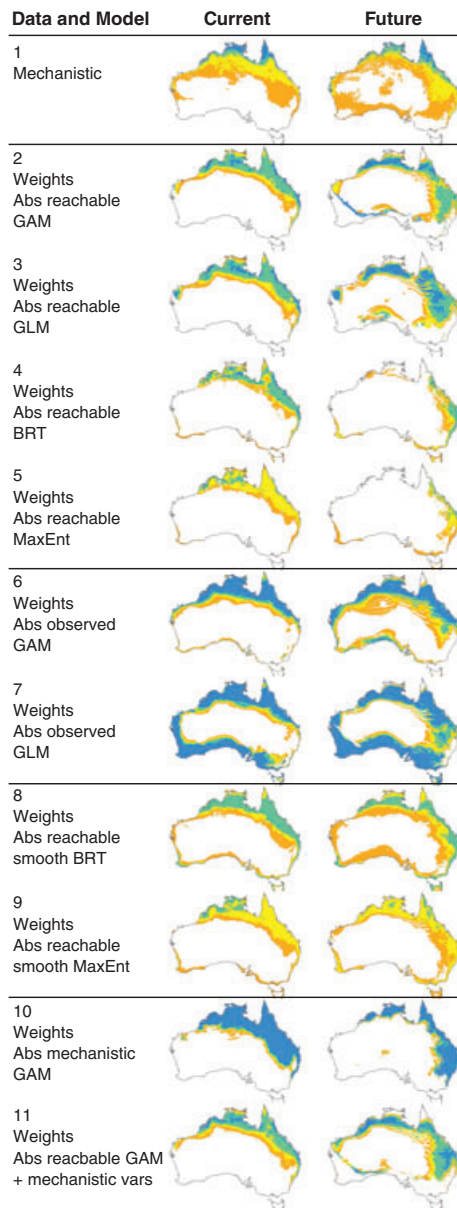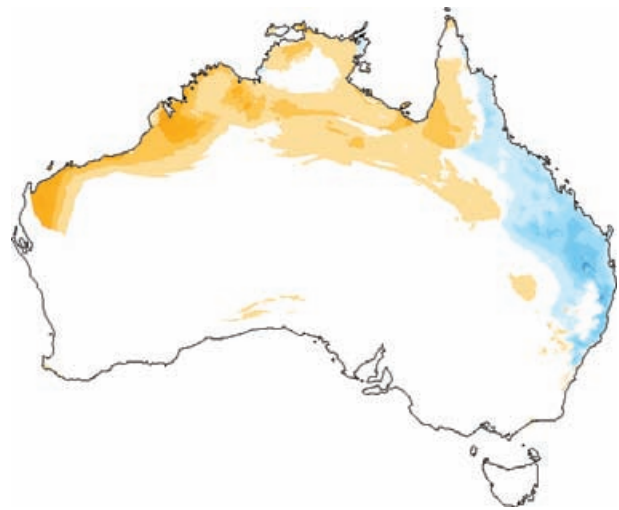
Fig. 4. Current and future predictions of the distribution of the cane toad, for various model types and data treatments. Predictions are coded white (low) to orange–yellow–green–blue (high), with breaks of 0–2, 3–5, 6–8, 9–10, 11–12 months suitable for the mechanistic model, and equal classes from 0 to 1 for the relative suitabilities predicted from the correlative models.

Fig. 5. Predictions from a weighted GAM with background in reachable areas minus those from an unweighted GAM with background across all of Australia, summarizing the overall effect of weights and background choice. Blue indicates negative values and orange, positive, with stronger colours showing more extreme differences.

occurrence within the predicted ranges rather than by substantially different locations of predicted unsuitable conditions. The results for one of the data treatments (using observed absences) is the main exception and reported in the next paragraph. These general trends can be confirmed numerically by comparing predictions across methods within data treatments using Pearson correlations (i.e. for all predicted grid cells in Australia): all pairwise correlations between methods using background samples were greater than 0·9. Variable importance varied across models (and with method of measuring it), although most models consistently identified humidity as one of the most important predictors. GLMs and GAMs then

tended to emphasize any of the temperature-related predictors, whereas BRT and MaxEnt either identified a rainfall predictor or temperature seasonality (clim4) as important (Appendix S6).

Data treatment caused more disparate predictions, with the largest effect related to the use of observed absences (Fig 4, left column of maps, rows 6 and 7; and Appendix S5A). These produced some models (most notably the GLMs) that predicted suitable locations along southern coastlines, whereas background samples throughout all Australia or in reachable areas constrained predictions to more northern areas. Background across all of Australia tended to predict more locations with high values on the east coast, whereas restriction to reachable areas gave more emphasis westward (e.g. Fig. 5). Use of weights created a more even spread of high predictions east, north and west trending compared with the more easterly emphasis without weights (Fig. 5; Appendix S5A).

Comparisons with predictions from the eco-physiological model of Kearney *et al.* (2008) show stronger correspondence for methods with relatively simple or smooth fitted functions (all but the standard BRT; for example fitted functions for standard and smooth BRTs, see Appendix S2); weights were more often useful than not, and the effect of absence source varied with the modelling method (Table 2). Cross-validated estimates of AUC (AUC$_{cv}$, Table 2) on the current data were not good predictors of the agreement of predictions with the ecophysiological model, with correlations between the AUC$_{cv}$ and the other three statistics (calculated across all method/treatment combinations) of −0·2, 0·3 and 0·2 with COR, KUL and AUC$_{mech}$ respectively.

CLIMATE CHANGE SCENARIOS

Future predictions revealed new features of the fitted models. No modelling method was clearly best: the eight models most similar to the future mechanistic predictions as judged by

Table 2. Evaluation statistics for all models, sorted by decreasing correlations (COR) with the mechanistic predictions for 20xx

Model	Wt	Abs	Current				Future: 20xx		
			AUC$_{cv}$	COR	KUL	AUC$_{mech}$	COR	KUL	AUC$_{mech}$
maxent.simple	y	reach	0·79	0·82	0·26	0·90	*0·78*	0·24	*0·87*
brt.simple	y	obs	0·92	0·82	0·27	0·87	*0·71*	0·25	0·73
gam	y	mk	*0·99*	*0·86*	0·28	*0·96*	0·70	0·43	0·80
gam	y	obs	0·90	0·85	0·27	0·92	0·70	0·28	0·80
brt	y	obs	0·91	0·75	0·32	0·84	0·68	0·28	0·80
brt.simple	y	reach	0·91	0·74	0·28	0·87	0·67	*0·23*	0·79
brt.simple	–	aus	0·94	0·74	0·29	0·85	0·67	*0·23*	0·77
gam.mk	y	reach	0·88	0·85	*0·23*	*0·97*	0·64	0·33	*0·87*
gam	y	reach	0·88	0·84	*0·25*	0·94	0·62	0·36	0·78
maxent.simple	y	aus	0·89	*0·85*	0·25	0·94	0·61	0·41	0·86
glm	y	aus	0·93	0·83	0·26	0·94	0·59	0·38	0·75
gam	y	aus	0·94	0·83	0·26	0·95	0·58	0·44	0·79
glm	y	reach	0·87	0·82	*0·25*	0·94	0·53	0·39	0·72
gam	–	aus	0·95	0·75	0·32	0·94	0·52	0·45	0·77
brt.simple	y	aus	0·94	0·80	0·28	0·88	0·50	0·27	0·74
glm	–	aus	0·95	0·74	0·32	0·93	0·45	0·45	0·71
maxent.simple	–	aus	0·95	0·72	0·34	0·94	0·44	0·50	0·86
glm	y	obs	0·89	0·38	0·43	0·68	0·43	0·39	0·73
brt	y	reach	0·91	0·73	0·34	0·86	0·41	0·46	0·79
brt	y	aus	0·95	0·75	0·32	0·91	0·35	0·51	0·72
brt	–	aus	*0·96*	0·63	0·39	0·90	0·33	0·54	0·79
maxent	y	aus	0·90	0·82	0·26	0·92	0·31	0·53	0·71
maxent	y	reach	0·82	0·81	0·28	0·88	0·30	0·48	0·68
maxent	–	aus	0·95	0·66	0·37	0·93	0·27	0·56	0·70

The top two (or more, given ties) results for each statistic are given in bold italics. The models and weights (Wt) are explained in the text; 'Abs' describes the source of the absences or background samples. AUC, area under the receiver operating characteristic curve; KUL, Kulczinski's coefficient; COR, Pearson correlation coefficient; obs, observed; reach, background sample in reachable areas; aus, background sample across all Australia; mk, from the mechanistic model of Kearney *et al.* (2008).

correlations include four smoothed models (three BRT and one MAXENT), one standard BRT, and three GAMs. Two of these models were informed by the mechanistic model (Table 2, rows 3 and 8). The two machine learning methods, MAXENT and BRT, produced similar and restricted climate change distributions when fitted using the methods commonly applied for modelling species' current distributions. However, when constrained to smoother fits (retaining key trends but ignoring finer detail) they predict future distributions amongst those most consistent with the ecophysiological model (Fig. 4 right column of maps, rows 8 and 9; and Appendix S5B). The GAMs and GLMs showed some similarities to the smooth BRTs and MAXENT models, but for some data treatments predicted unlikely high values in central-southern areas (Fig. 4; Appendix S5B).

A key trend emerging from our analyses was that the use of background data across all of Australia tended to produce more eastward predictions for all models, compared with the use of data restricted to reachable areas (namely background in reachable areas or observed absences; Appendix 5B). Again, the models based on observed absences tended to predict high future suitability along the southern coastline (e.g. Fig. 4, row 7).

The maps (Fig. 4; Appendix S5B) and statistics (Table 2) comparing these climate change predictions with those of the ecophysiological model provide several perspectives on the results. The Kulczynski coefficients emphasize agreement of

high predictions, and ranked two smooth BRTs best because they most completely predicted the northern and north-eastern areas predicted by the ecophysiological model. Correlation penalizes deviation from 'truth' and viewed a smooth BRT and MAXENT as most similar, and most of the standard-fit BRTs and MAXENT as most different, to the mechanistic predictions. AUC targets the ability of the predictions to distinguish presence from absence and – given our choice of threshold for the mechanistic prediction – tended to favour those predictions without high values in the areas mapped absent in the future ecophysiological predictions (Fig. 4). More generally, the maps showed that the models emphasize different future 'hotspots' for the toad (Fig. 4; Appendix S5B). There appears no unifying feature of the eight climate change predictions most similar to the mechanistic prediction in terms of variables selected as most important in the models (Appendix S6). Again, the cross-validated estimates of AUC on the current data (AUC$_{cv}$) were poor predictors of the agreement of future correlative model predictions with the future ecophysiological predictions, with correlations between the AUC$_{cv}$ and the three test statistics of −0·2, 0·3 and −0·2 for COR, KUL and AUC$_{mech}$ respectively.

Whilst variable importance (Appendix S6) and partial plots (Appendix S7) are useful overall summaries, the question of what drives a high or low prediction at any given site in geographic space remains unanswered. The latter requires specific information on the climate and on the components of the
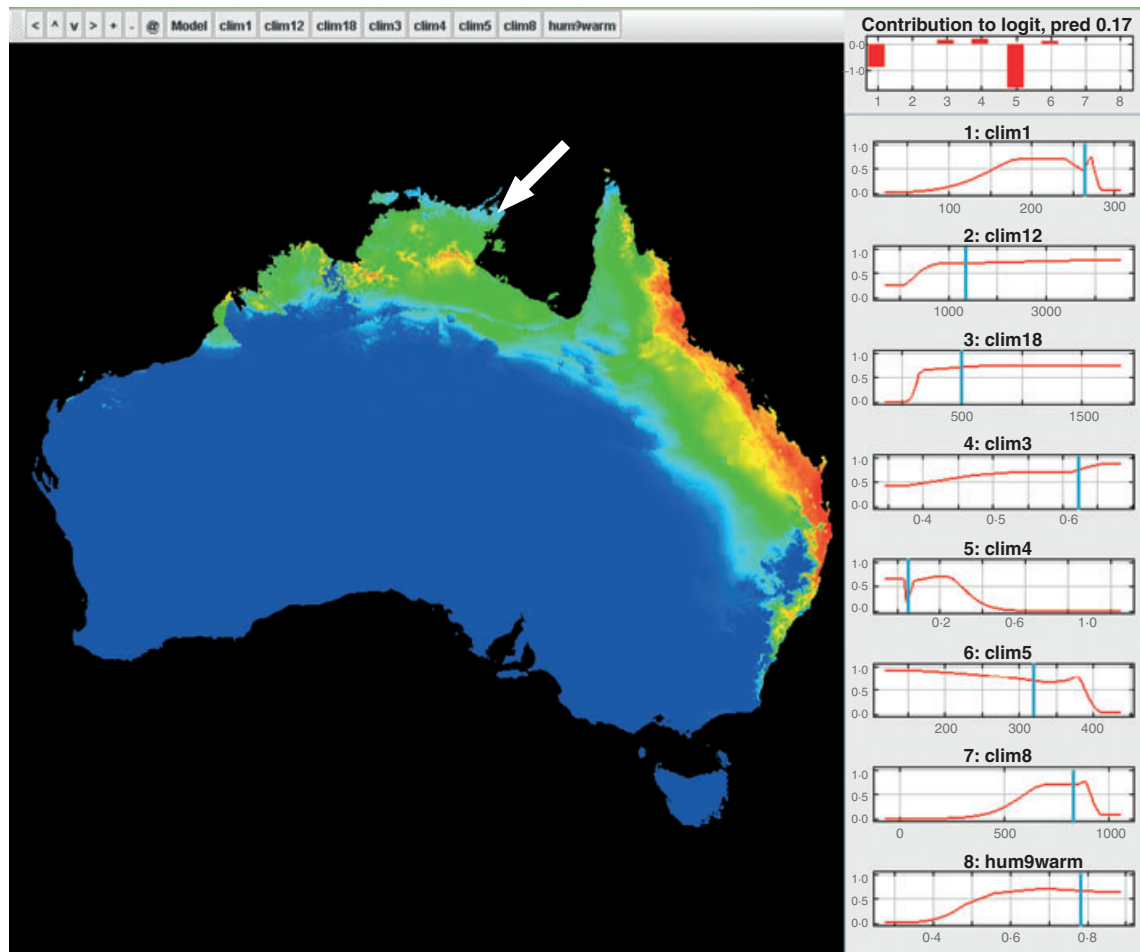
Fig. 6. Exploration of the components of a prediction at a site in northern Australia (indicated on the map with an arrow), showing relative influence of each variable at that site (top right) and fitted functions (right, other panels) with vertical blue lines showing conditions at the selected site.

prediction at the site. We have programmed new capabilities into MAXENT for exploring what drives predictions at any selected site in terms of the underlying model and its fitted functions (Fig. 6 and Appendix S3). In the illustrated example, a site at the very north of Australia is predicted as relatively low suitability (see arrow Fig. 6), largely driven by the response to clim1 and clim4 (upper right panel, Fig. 6). The partial plots to the right of the map show that the fitted functions for these influential variables both fit a local 'trough' in the response at these environmental values (see vertical blue lines indicating conditions at the selected site) because the data set contains close to zero presences in such conditions despite these conditions being available in the background. Figure 7 presents another spatially explicit exploration of a model and data, indicating how the variables most influencing predictions vary across Australia.

INTEGRATING CORRELATIVE AND MECHANISTIC MODELS

Absences drawn from mechanistic predictions led to models that successfully predicted suitable habitat in the far north of Australia (Fig. 4, row 10). Proximal predictors drawn from the
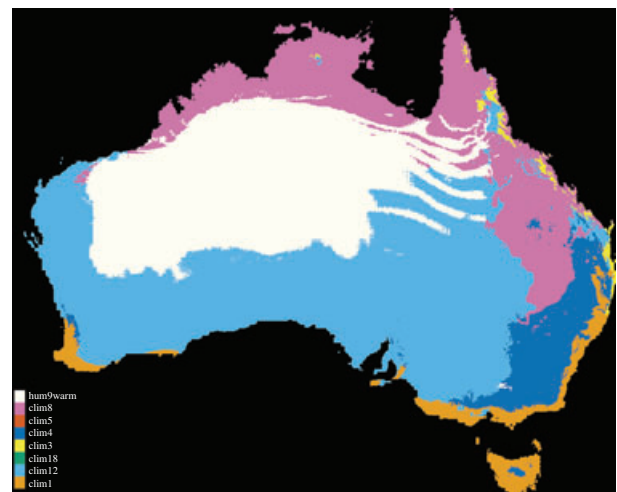
Fig. 7. Limiting factors based on the smoothed MAXENT model using weights and background in reachable areas – for any point, the limiting factor is the variable whose value at that point most influences the model prediction. For details, See Appendix S4.

mechanistic model were selected as important (Appendix S6) but caused only subtle changes in the predictions (Fig. 4, rows 2 and 11).

Discussion

Our study reinforces the fact that model predictions can be very sensitive in the context of range-shifting species. When predicting potential invasion extent for the cane toad under the current climate, the largest differences among models were related to treatment of absence points. Even more dramatic was the widespread disagreement of models when pushed to an extreme climate change scenario (Fig. 4). This sensitivity of outcome to method has been demonstrated elsewhere, especially in the context of climate change (e.g. Pearson *et al.* 2006), and some authors have argued that consensus (ensemble) methods are an appropriate way to deal with the issue (Thuiller 2004; Araújo, Thuiller & Pearson 2006; Araújo & New 2007; Broennimann & Guisan 2008). A substantial and acknowledged problem with this is that there are usually no future data for testing the relative performance of methods and selecting the best performing ones for the consensus predictions. Testing on withheld portions on current data is sometimes used (e.g. Broennimann & Guisan 2008), but in our case study this did not correlate at all with performance under climate change scenarios (Table 2). Whilst it is possible that this lack of agreement is exacerbated by the species being invasive and not yet filling its potential distribution, we expect the same problems for species at equilibrium when projecting their distributions to novel climates.

Other research fields use ensembles, often by taking averages over all available models that satisfy certain prespecified criteria – e.g. tests of predictive skill, or of model independence (Tebaldi & Knutti 2007; Jose & Winkler 2008; Abramowitz 2010). Ensembles may produce robust forecasts, but the component models must be realistic and well understood. In species modelling, we do not think there is sufficient information to automatically select models for consensus (e.g. using measures of predictive performance on current observations), and instead prefer to emphasize the importance of understanding the data, the model and its predictions when assessing predictions. In our analyses we have developed several constructive ways forward that involve: (1) exploring data weighting schemes and absence delineation; (2) assessing environmental novelty in the projected space; (3) exploring modelled responses and predictor weightings in contentious regions; (4) enforcing 'smooth' responses; and (5) integrating mechanistic predictions with correlative models. We discuss each of these in turn.

EXPLORING DATA TREATMENTS: ABSENCES, BACKGROUNDS, WEIGHTINGS AND NATIVE RANGES

Previous modelled predictions of the potential distribution of the cane toad have varied (Phillips, Chipperfield & Kearney 2008), and most notably Urban *et al.* (2007) predicted suitable environments in southern Australia. Whilst we have not tried to exactly reproduce the Urban model (a model averaged GLM), we obtained similar results when using the absence data with GLMs (Fig. 4; Appendix S5) suggesting that this set of absences has the potential to drive southern predictions. The MESS maps (Fig. 2) show that, even for current climate, the Urban *et al.*

(2007) data require models to extrapolate into novel climates, including ones with changed correlations between variables (e.g. Fig. S4.2). In such situations, the interaction between the data set, the fitted model (Appendix S7) and the predictive task become extremely important. If a model fits responses that extrapolate in ecologically unrealistic ways, predictions into novel spaces will have poor foundations. Our use of background data in reachable areas informed the model about the sampled space but also reduced the required degree of extrapolation for current climates, providing a less risky prediction space. Similarly, our use of weights helped to emphasize environments that were possibly under-represented in the data. Our general aim with these data treatments was to adjust the data to account for deviations from the equilibrium distribution of the species, simultaneously staying aware of the novelty of the prediction space and the behaviour of our models in that space.

Other methods have been suggested for dealing with non-equilibrium records for invasive species, including some that include dynamics of dispersal and other population processes (Hooten *et al.* 2007; Prasad *et al.* 2010; Smolik *et al.* 2010). Amongst those that more simply focus on static models, a common approach is to use native range data for model building, or data covering both native and invaded ranges (Roura-Pascual *et al.* 2004, 2006; Mau-Crimmins, Schussman & Geiger 2006; Fitzpatrick *et al.* 2007). These can be useful, and we know of current research integrating our methods presented here with models based on both native range and invaded range data for the cane toad (R. Tingley, personal communication). Our research focussed on invaded range data because there were sufficient data in this case for meaningful model development, results were comparable with published models, and attention was directed towards the lack of equilibrium, which is our main interest here. The methods are transferable to other data configurations. For instance, the MESS maps, the methods for visualizing changes in correlations between variables (Figure 3; Appendix S4), and those for understanding what is driving the predictions could all be used for models based on native range data. Weights and/or judicious choice of background data can be used for the invaded range component of models combining invaded and native data.

In our study, weighted models were amongst the better ones (Table 2, Fig. 5, Appendix S5), although the effect was not clear-cut. We expect that a combination of factors contribute to the limited effect: the extent to which the toad has already invaded much of the environmental space in northern Australia; the relatively large number of presence records; the use of backgrounds consistent with invasion patterns; and the happenstance that the regression models did not predict erratically outside the sampled range of the data. The idea of using weights or otherwise adjusting the data or models to better represent an equilibrium distribution deserves further attention.

ASSESSING ENVIRONMENTAL NOVELTY: CORRELATIONS AND MESS

It is important to know *where* climates are novel. Few modellers have paid attention to this, although exceptions include

Williams, Jackson & Kutzbac (2007), Platts *et al.* (2008) and Fitzpatrick & Hargrove (2009). The necessity is an acknowledged one, and other researchers provide interesting comparisons of climate spaces (e.g. using principal components analyses and metrics summarizing differences between niches; Thuiller, Lavorel & Araujo 2005a; Broennimann *et al.* 2007; Warren, Glor & Turelli 2008; Medley 2010) but without mapping results into geographic space. Here, we used the understanding of the similarity/novelty of climates to indicate where the models are most uninformed, guiding us to the locations where we needed to interrogate our predictions and aiding our interpretation of model differences. This helps to identify and reject models with fitted functions that extrapolate in ways that are biologically implausible. Alternatively, novelty could be used as a mask to warn against the use of predictions in certain areas, or as a quantitative measure of prediction uncertainty. We expect that the new programmed capabilities in MAXENT for estimating and mapping MESS will aid such investigations. Associated maps that identify which climate variable is driving the MESS value in each grid cell are also provided in the software (Appendix S3).

The MESS maps will not identify changes in correlations between variables, and tests for these are also critical because the model parameters are estimated on the correlation structure between predictors in the training data. For most models, predictions to areas with substantially different correlations between important variables will be unreliable (Harrell 2001; but see Zadrozny 2004). This is particularly problematic when the available predictors are only indirectly related to the species' distribution (Austin 2002). The selected set might together represent the unmeasured directly influential variable reasonably well, but if correlations between them change in new areas, prediction will be compromised. We have provided some suggestions (Fig. 3, Appendix S4), but there remains much scope for providing relevant tools for visualizing changes in correlations. The climate variables used here, and frequently in many analyses, include several based on the warmest months, the driest quarters and so on. As these change seasonally across many continents (e.g. in Australia the driest months are in winter in the north, and summer in the south), the correlations between the full suite of climate variables can be particularly prone to strong spatial variation. Figure S4.2 illustrates the problem. A useful strategy could include testing for changes in correlation before modelling and choosing candidate variables in the light of that information.

EXPLORING MODELLED RESPONSES AND PREDICTOR WEIGHTING

Ecologists have surprisingly few tools available to explore the spatial patterns in modelled predictions; yet, these are particularly important to understanding predicted distributions. Even though methods are available for summarizing overall model features, it is the spatially varying make-up of the predictions that can give particular insights. For instance, the ideas presented in Fig. 6 and programmed into MAXENT link predictions, the model terms and environmental conditions at the site, enabling the modeller to understand what is driving predictions. We intend to extend this programmed capability to be applicable to any modelling method that can output appropriate data summaries because this information is particularly important for understanding model performance. The limiting factors in Fig. 7 are another interesting insight into the drivers of predictions, and provide a powerful basis for comparisons with physiological knowledge.

More generally, if models are to be used for extrapolation, we believe it is important to know how the fitted responses are acting at the extremes of the sampled environments, and control them appropriately. The GLMs and GAMs performed reasonably well for many of the data treatments tested here, but the potential for fitted responses that extrapolate unrealistically (e.g. the GLM fitted to observed absences, Fig. 4; Appendix S7) is likely to lead to poor performance in other situations, if left uncontrolled. A solution is to use nonlinear terms with known behaviour beyond the range of the data – for instance, natural splines (which are linear beyond their boundary knots) or splines that become constant beyond the boundaries (Trevor Hastie, personal communication). Another possible solution is to use clamping as implemented in MAXENT – i.e. by making the response constant outside of the range of the training data. These are promising avenues for research.

ENFORCE 'SMOOTH' RESPONSES

One marked result new to this study is the change in performance of the statistical (machine) learning methods when smoothness was enforced. The general concept that models which fit complex responses to environment may not predict well for species not at equilibrium is not new (Elith *et al.* 2006), but so far has not been explored. Many methods are capable of fitting complexity – for instance, GAMs can be allowed many degrees of freedom for the smooth functions (Hastie, Tibshirani & Friedman 2009), but here we only tested relatively simple GAM fits. We allowed complexity for MAXENT and BRT. The important issue is whether a method capable of complexity is fitting patterns pertinent to the species, or fitting unwanted features of the data sample. In this study, there were no records for the toad in the far north of the Northern Territory (centre north on the map). Background samples throughout all this region implied that this hot and humid area had been surveyed, whereas in reality surveys were probably missing (the area is remote and local knowledge suggests occurrences; NT Government 2006). If the species had been at equilibrium, multiple examples of occurrence in these hot climates would probably have existed in the data. The data were correctly modelled by MAXENT and BRT but were misleading. Enforcing smoothness meant that the models focussed on the strongest trends but did not model detail included in the more complex fits. Most importantly, response to annual temperature (clim1) no longer declined at high temperatures (Appendix S2). This meant that when predicting to hotter temperatures of the change scenarios, the smoother models predicted the north of Australia as suitable for the toads, consistent with the mechanistic model.

We do not view this result as an argument against complexity but more as a warning that inaccuracies due to unrepresentative data can be strongly amplified when extrapolating. Making over-smooth models will unrealistically predict equal probabilities in differing environments (Barry & Elith 2006); so, smoothness is not a panacea. Further, responses that imply ability of the organism to survive across any range of future temperatures are clearly implausible and need to be used with due care. The problem is lack of suitable training data. There are likely to be other ways to control the fits of these potentially complex models to data with biases, and further research will be instructive. We only used hinge features for MaxEnt, but linear and quadratic features would also produce smooth models, and products would allow simple interactions. More information on how uncertainty in the predictions varies spatially (e.g. from running multiple models on perturbed data) should also help to inform this general area of the effect of unrepresentative data on model predictions.

INTEGRATING MECHANISTIC AND CORRELATIVE PREDICTIONS

Given the difficulties in robustly modelling species not at equilibrium, it is essential to think about the biology of the species, interrogate the fitted models and do as much as possible to ensure that unwanted effects are not incorporated. It may be practically difficult to run a full physiological model for most species but even small amounts of information can help. For instance, physiologically relevant predictors can be selected (Rödder *et al.* 2009), the fitted functions can be assessed for plausibility, the way they extrapolate can be controlled to be consistent with available knowledge, and predictions can be tested with experiments or compared with existing physiological data. If physiological models are available, new opportunities open up (for recent examples, see Morin & Thuiller 2009; Kearney, Wintle & Porter in press). In our modelling of the cane toad, using the mechanistic predictions to establish likely absences resulted in the first models that successfully predicted highly suitable conditions to the northernmost parts of Australia. Rather than background absences that implied that the unsampled north was unsuitable, or observed absences so restricted in distribution that they require extrapolation of models even for current conditions, the mechanistic model provided absence data from unsuitable habitats across Australia. The physiological models provide strong inference on absences because they emphasize processes that define the fundamental niche of the species. The models could also be viewed as a source of presence data, but we consider this less well supported. In focussing on a few key processes, mechanistic models may miss some variables influencing abundance of the species; so, the way that modelled months of suitability translates into observed toad frequencies across the landscape is probably less well prescribed. Use of mechanistically derived predictions as a source of absence information is conceptually most appealing.

Alternatively, use of information from the mechanistic model in the form of proximal predictors has the potential to provide important missing information to a correlative model. In our modelling, additional predictors quantifying the potential for adult movement and conditions permitting larval development were selected in the model and were amongst the three most important predictors (Appendix S6) but did not substantially alter the modelled predictions. Apparently the available climatic predictors already did a reasonable job in defining suitable conditions for the toad.

Conclusion

Modelling of any sort is an art, but modelling range-shifting species is a particularly delicate one. The toad was a useful case for exploring how to do this, but is it representative of invasive species more generally? The success of the ecophysiological model in capturing the toad's distribution limits probably means that there are few strong biotic interactions influencing its range at present. This means that the toad's final range will be close to its fundamental niche. However, this may not be a lone example: it is likely that the absence of strong predators, pathogens or competitors is an important reason for the success of invasive species in general (Lockwood, Hoopes & Marchetti 2007). It is also clear that climate match is an important predictor of invasion success (Hayes & Barry 2008; Bomford *et al.* 2009). Hence, the methods we advocate may help considerably in the study of invasive species. Species whose ranges are a reflection of environmentally contingent biotic interactions (the majority of species) are likely to respond in more complex ways in novel environmental space. These will be much harder to model successfully, and will require even greater caution in modelling future ranges.

Acknowledgements

References

Abramowitz, G. (2010) Model independence in multi-model ensemble prediction. *Australian Meteorological and Oceanographic Journal*, **59**, 3–6.

ANU (2009) ANUCLIM, http://fennerschool.anu.edu.au/publications/software/anuclim.php#contacts.

Araújo, M.B. & New, M. (2007) Ensemble forecasting of species distributions. *Trends in Ecology and Evolution*, **22**, 42–47.

Araújo, M. & Pearson, R.G. (2005) Equilibrium of species' distributions with climate. *Ecography*, **28**, 693–695.

Araújo, M.B., Thuiller, W. & Pearson, R.G. (2006) Climate warming and the decline of amphibians and reptiles in Europe. *Journal of Biogeography*, **33**, 1712–1728.

Araújo, M.B., Pearson, R.G., Thuiller, W. & Erhard, M. (2005) Validation of species-climate impact models under climate change. *Global Change Biology*, **11**, 1504–1513.

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

Barry, S.C. & Elith, J. (2006) Error and uncertainty in habitat models. *Journal of Applied Ecology*, **43**, 413–423.

van Beurden, E.K. (1981) Bioclimatic limits to the spread of *Bufo marinus* in Australia: a baseline. *Proceedings of the Ecological Society of Australia*, **11**, 143–149.

Bomford, M., Kraus, F., Barry, S.C. & Lawrence, E. (2009) Predicting establishment success for alien reptiles and amphibians: a role for climate matching. *Biological Invasions*, **11**, 713–724.

Booth, G.D., Niccolucci, M.J. & Schuster, E.G. (1994) Identifying proxy sets in multiple linear regression: an aid to better coefficient interpretation. Intermountain Research Station, USDA Forest Service, Ogden, Utah, USA.

Broennimann, O. & Guisan, A. (2008) Predicting current and future biological invasions: both native and invaded ranges matter. *Biology Letters*, **4**, 585–589.

Broennimann, O., Treier, U.A., Müller-Schärer, H., Thuiller, W., Peterson, A.T. & Guisan, A. (2007) Evidence of climatic niche shift during biological invasion. *Ecology Letters*, **10**, 701–709.

Buisson, L., Thuiller, W., Casajus, N., Lek, S. & Grenouillet, G. (2009) Uncertainty in ensemble forecasting of species distribution. *Global Change Biology*, **16**, 1145–1157.

Busby, J.R. (1991) BIOCLIM – a bioclimate analysis and prediction system. *Nature Conservation: Cost Effective Biological Surveys and Data Analysis* (eds C.R. Margules & M.P. Austin), pp. 64–68. CSIRO, Canberra.

De Marco, P., Diniz-Filho, J.A.F. & Bini, L.M. (2008) Spatial analysis improves species distribution modelling during range expansion. *Biology Letters*, **4**, 577–580.

Dormann, C.F. (2007) Promising the future? Global change projections of species distributions. *Basic and Applied Ecology*, **8**, 387–397.

Elith, J. & Leathwick, J.R. (2009) Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*, **40**, 677–697.

Elith, J., Leathwick, J.R. & Hastie, T. (2008) A working guide to boosted regression trees. *Journal of Animal Ecology*, **77**, 802–813.

Elith, J., Graham, C.H., Anderson, R.P., Dudík, M., Ferrier, S., Guisan, A. *et al.* (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129–151.

Faith, D.P., Minchin, P.R. & Belbin, L. (1987) Compositional dissimilarity as a robust measure of ecological distance. *Vegetatio*, **69**, 57–68.

Ficetola, G.F., Thuiller, W. & Miaud, C. (2007) Prediction and validation of the potential global distribution of a problematic alien invasive species the American bullfrog. *Diversity and Distributions*, **13**, 476.

Fitzpatrick, M.C. & Hargrove, W. (2009) The projection of species distribution models and the problem of non-analog climate. *Biodiversity and Conservation*, **18**, 2255–2261.

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.

Friedman, J.H., Hastie, T. & Tibshirani, R. (2000) Additive logistic regression: a statistical view of boosting. *The Annals of Statistics*, **28**, 337–407.

Hanley, J.A. & McNeil, B.J. (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, **143**, 29–36.

Harrell, F.E. (2001) *Regression Modeling Strategies with Applications to Linear Models, Logistic Regression and Survival Analysis*. Springer Verlag, New York.

Hastie, T. (2008) *GAM: Generalized Additive Models*. R package version 1.0. http://CRAN.R-project.org/package = gam.

Hastie, T. & Tibshirani, R. (1990) *Generalized Additive Models*. Chapman & Hall, London.

Hastie, T., Tibshirani, R. & Friedman, J.H. (2009) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd edn. Springer-Verlag, New York.

Hayes, K.R. & Barry, S. (2008) Are there any consistent predictors of invasion success? *Biological Invasions*, **10**, 483–506.

Heikkinen, R.K., Luoto, M., Araujo, M.B., Virkkala, R., Thuiller, W. & Sykes, M.T. (2006) Methods and uncertainties in bioclimatic envelope modelling under climate change. *Progress in Physical Geography*, **30**, 751–777.

Hijmans, R.J. & Graham, C.H. (2006) The ability of climate envelope models to predict the effect of climate change on species distributions. *Global Change Biology*, **12**, 2272.

Hooten, M.B., Wikle, C.K., Dorazio, R.M. & Royle, J.A. (2007) Hierarchical spatiotemporal matrix models for characterizing invasions. *Biometrics*, **63**, 558–567.

Jose, V.R.R. & Winkler, R.L. (2008) Simple robust averages of forecasts: some empirical results. *International Journal of Forecasting*, **24**, 163–169.

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

Kearney, M.R., Wintle, B.A. & Porter, W.P. (2010) Correlative and mechanistic models of species distribution provide congruent forecasts under climate change. *Conservation Letters*, in press.

Kearney, M., Phillips, B.L., Tracy, C.R., Christian, K.A., Betts, G. & Porter, W.P. (2008) Modelling species distributions without using species distributions: the cane toad in Australia under current and future climates. *Ecography*, **31**, 423–434.

Lockwood, J.L., Hoopes, M.A. & Marchetti, M.P. (2007) *Invasion Ecology*. Blackwell Publishing, Malden, Massachusetts, USA.

Marmion, M., Parviainen, M., Luoto, M., Heikkinen, R.K. & Thuiller, W. (2008) Evaluation of consensus methods in predictive species distribution modelling. *Diversity and Distributions*, **15**, 59–69.

Mau-Crimmins, T.M., Schussman, H.R. & Geiger, E.L. (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*, **193**, 736–746.

McCullagh, P. & Nelder, J.A. (1989) *Generalized Linear Models*, 2nd edn. Chapman & Hall, London.

Medley, K.A. (2010) Niche shifts during the global invasion of the Asian tiger mosquito, *Aedes albopictus* Skuse (Culicidae), revealed by reciprocal distribution models. *Global Ecology and Biogeography*, **19**, 122–133.

Menke, S.B., Holway, D.A., Fisher, R.N. & Jetz, W. (2009) Characterizing and predicting species distributions across environments and scales: Argentine ant occurrences in the eye of the beholder. *Global Ecology and Biogeography*, **18**, 50–63.

Morin, X. & Thuiller, W. (2009) Comparing niche- and process-based models to reduce prediction uncertainty in species range shifts under climate change. *Ecology*, **90**, 1301–1313.

NT Government (2006) (http://www.nt.gov.au/nreta/wildlife/animals/caneto-ads/pdf/nt_distribution_2006_Cane_Toad.pdf, last accessed March2010. Northern Territory Government, Australia.

Parmesan, C. (2006) Ecological and evolutionary responses to recent climate change. *Annual Review of Ecology, Evolution, and Systematics*, **37**, 637–669.

Pearson, R.G., Thuiller, W., Araújo, M.B., Martinez-Meyer, E., Brotons, L., McClean, C. *et al.* (2006) Model-based uncertainty in species range prediction. *Journal of Biogeography*, **33**, 1704–1711.

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

Phillips, B., Chipperfield, J.D. & Kearney, M.R. (2008) The toad ahead: challenges of modelling the range and spread of an invasive species. *Wildlife Research*, **35**, 222–234.

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

Phillips, S.J., Dudik, M., Elith, J., Graham, C., Lehmann, A., Leathwick, J. *et al.* (2009) Sample selection bias and presence-only models of species distributions. *Ecological Applications*, **19**, 181–197.

Platts, P.J., McClean, C.J., Lovett, J.C. & Marchant, R. (2008) Predicting tree distributions in an East African biodiversity hotspot: model selection, data bias and envelope uncertainty. *Ecological Modelling*, **218**, 121–134.

Prasad, A.M., Iverson, L.R., Peters, M.P., Bossenbroek, J.M., Matthews, S.N., Sydnor, T.D. *et al.* (2010) Modeling the invasive emerald ash borer risk of spread using a spatially explicit cellular model. *Landscape Ecology*, **25**, 353–369.

R Development Core Team (2009) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna. http://www.R-project.org.

Ridgeway, G. (2007) *GBM: Generalized Boosted Regression Models*. R package version 1.6-3. http://www.i-pensieri.com/gregr/gbm.shtml.

Rödder, D., Schmidtlein, S., Veith, M. & Lötters, S. (2009) Alien invasive slider turtle in unpredicted habitat: a matter of niche shift or of predictors studied? *PLoS ONE*, **4**, e7843.

Roura-Pascual, N., Suarez, A., Gómez, C., Pons, P., Touyama, Y., Wild, A.L. *et al.* (2004) Geographic potential of Argentine ants (*Linepithema humile* Mayr) in the face of global climate change. *Proceedings of the Royal Society of London Series B – Biological Sciences*, **271**, 2527–2534.

Roura-Pascual, N., Suarez, A.V., Gómez, C., Pons, P., Touyama, Y., Wild, A.L. *et al.* (2006) Niche differentiation and fine-scale projections for Argentine ants based on remotely-sensed data. *Ecological Applications*, **16**, 1832–1841.

Roura-Pascual, N., Brotons, L., Peterson, A. & Thuiller, W. (2009) Consensual predictions of potential distributional areas for invasive species: a case study of Argentine ants in the Iberian Peninsula. *Biological Invasions*, **11**, 1017–1031.

Scheller, R.M. & Mladenoff, D.J. (2005) A spatially interactive simulation of climate change, harvesting, wind, and tree species migration and projected changes to forest composition and biomass in northern Wisconsin, USA. *Global Change Biology*, **11**, 307–321.

Smolik, M.G., Dullinger, S., Essl, F., Kleinbauer, I., Leitner, M., Peterseil, J. *et al.* (2010) Integrating species distribution models and interacting particle systems to predict the spread of an invasive alien plant. *Journal of Biogeography*, **37**, 411–422.

Sutherst, R.W., Floyd, R.B. & Maywald, G.F. (1995) The potential geographic distribution of the Cane Toad, *Bufo marinus*, in Australia. *Conservation Biology*, **9**, 294–299.

Tebaldi, C. & Knutti, R. (2007) The use of the multimodel ensemble in probabilistic climate projections. *Philosophical Transactions of the Royal Society of London Series A*, **365**, 2053–2075.

Thuiller, W. (2004) Patterns and uncertainties of species' range shifts under climate change. *Global Change Biology*, **10**, 2020–2027.

Thuiller, W., Lavorel, S. & Araujo, M.B. (2005a) Niche properties and geographical extent as predictors of species sensitivity to climate change. *Global Ecology and Biogeography*, **14**, 347–357.

Thuiller, W., Richardson, D.M., Pysek, P., Midgley, G.F., Hughes, G.O. & Rouget, M. (2005b) Niche-based modelling as a tool for predicting the risk of alien plant invasions at a global scale. *Global Change Biology*, **11**, 2234–2250.

Thuiller, W., Albert, C., Araújo, M.B., Berry, P.M., Cabeza, M., Guisan, A. *et al.* (2008) Predicting global change impacts on plant species' distributions: future challenges. *Perspectives in Plant Ecology, Evolution and Systematics*, **9**, 137–152.

Urban, M.C., Phillips, B.L., Skelly, D.K. & Shine, R. (2007) The cane toad's (*Chaunus (Bufo) marinus*) increasing ability to invade Australia is revealed by a dynamically updated range model. *Proceedings of the Royal Society B-Biological Sciences*, **274**, 1413–1419.

Warren, D.L., Glor, R.E. & Turelli, M. (2008) Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. *Evolution*, **62**, 2868–2883.

Williams, J.W., Jackson, S.T. & Kutzbac, J.E. (2007) Projected distributions of novel and disappearing climates by 2100 AD. *Proceedings of the National Academy of Sciences*, **104**, 5738–5742.

Zadrozny, B. 2004. Learning and evaluating classifiers undersample selection bias. *Proceedings of the Twenty-First International Conference on Machine Learning*. Association of Machine Learning, New York, USA, p. 114–121.

Zurrell, D., Jeltsch, F., Dormann, C.F. & Schröder, B. (2009) Static species distribution models in dynamically changing systems: how good can predictions really be? *Ecography*, **31**, 1–12.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Appendix S1. Methods for providing weights for species records.

Appendix S2. Fitted functions for selected models.

Appendix S3. MESS maps and limiting factors.

Appendix S4. Changes in correlations.

Appendix S5. Predictions for current and climate change scenarios.

Appendix S6. Variable importance.

Appendix S7. Fitted functions for some models.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.