This is the author-created version of the following work:

Beaumont, Linda, Graham, Erin, Duursma, Daisy Englert, Wilson, Peter D., Cabrelli, Abigail, Baumgartner, John B., Hallgren, Willow, Esperon-Rodriguez, Manuel, Nipperess, David A., Warren, Dan L., Laffan, Shawn W., and VanDerWal, Jeremy (2016) *Which species distribution models are more (or less) likely to project broad-scale, climate-induced shifts in species ranges?.* **Ecological Modelling, 342 pp. 135-146.**

Access to this file is available from:

Please refer to the original source for the final version of this work:

Title: Which species distribution models are more (or less) likely to project broad-scale, climate-induced shifts in species ranges?

Running title: Biases in species distribution models

Linda J. Beaumont[1], Erin Graham[2], Daisy Englert Duursma[1], Peter D. Wilson[1], Abigail Cabrelli[1], John B. Baumgartner[1], Willow Hallgren[3], Manuel Esperón-Rodríguez[1], David A. Nipperess[1], Dan L. Warren[1], Shawn W. Laffan[4], Jeremy VanDerWal[2]

[1]*Department of Biological Sciences, Macquarie University, NSW, 2109, Australia*

[2]*Centre for Tropical Biodiversity and Climate Change, School of Marine and Tropical Biology, James Cook University, Townsville, QLD, 4811, Australia*

[3]*Griffith University, Gold Coast Campus, Parklands Drive, Southport, QLD, 4215, Australia*

[4]*Centre for Ecosystem Science, School of Biological, Earth and Environmental Sciences, UNSW, Sydney, NSW, 2052, Australia*

Corresponding Author: Linda Beaumont

Department of Biological Sciences, Macquarie University, NSW, 2109, Australia

Phone: (612) 9850 8157

Fax: (612) 9850 8245

Email: linda.beaumont@mq.edu.au

Keywords: biomod, climate change, dismo, Maxent, range changes, species distribution models

Type of paper: Primary Research Article

Abstract

Species distribution models (SDMs) frequently project substantial declines in the spatial extent of climatically suitable habitat in response to scenarios of future climate change. Such projections are highly disconcerting. Yet, considerable variation can occur in the direction and magnitude of range changes projected by different SDM methods, even when predictive performance is similar. In this study, we assessed whether particular methods have a tendency to predict substantial loss or gain of suitable habitat. In particular, we asked, "are 14 SDM methods equally likely to predict extreme changes to the future extent of suitable habitat for 220 Australian mammal species?". We defined five non-mutually exclusive categories of 'extreme' change, based on stability or loss of current habitat, or the dislocation of current and future habitat; a) no future habitat (range extinction), b) low stability of current habitat (≤ 10% remains), c) no gain of habitat in new locations, d) all future habitat is in new locations (i.e. completely displaced from current habitat) and e) substantial increase in size of habitat (future habitat is ≥ 100% larger than current). We found that some SDM methods were significantly more likely than others to predict extreme changes. In particular, distance-based models were significantly less likely than other methods to predict substantial increases in habitat size; Random Forest models and Surface Range Envelopes were significantly more likely to predict a complete loss of current habitat, and future range extinction. Generalised Additive Models and Generalised Linear Models rarely predicted range extinction; future

habitat completely disjunct from current habitat was predicted more frequently than expected by Classification Tree Analysis and less frequently by Maxent. Random Forest generally predicted extreme range changes more frequently than other SDM methods. Our results identify trends among different methods with respect to tendency to predict extreme range changes. These are of significance for climate-impact assessments, with implications for transferability of models to novel environments. Our findings emphasise the need to explore and justify the use of different models and their parameterisations, and to develop approaches to assist with optimisation of models.

Highlights

- Species distribution models frequently predict substantial range shifts due to climate change
- Frequency of predictions of extreme range changes varies across SDM methods
- Random Forest was more likely than other methods to predict extreme range changes
- Distance-based models never predicted substantial increases in range size

Introduction

Contemporary anthropogenic climate change has already resulted in shifts in the range margins of numerous, diverse taxa (Melles *et al.*, 2010; Chen *et al.*, 2011; Garroway *et al.*, 2011; Rubidge *et al.*, 2011), and paleoecological data suggest that range shifts were the norm during previous episodes of climate change (Birks, 1989; Huntley, 1990; Graham, 1992; Willis & MacDonald, 2011). Understanding the extent to which species distributions may shift in response to climate change over the course of this century and beyond may assist in identifying species vulnerable to climate change, prioritising conservation efforts, and developing optimal adaptation and land management plans.

Correlative species distribution models (SDMs) are frequently used to examine the potential for changes to the distribution and quality of habitat under scenarios of future climate (Franklin, 2010). This approach is based on the assumption that the location of populations reflects the environmental preferences and tolerances of a species (Guisan & Thuiller, 2005). Models of this relationship, based on the observed subset of those populations, can then be used to assess the suitability of a region, for a particular species, under current, past or future climate. Such climate change experiments assume that relationships inferred from historical data remain consistent under a changing climate (Maraun, 2015).

Predictions of broad-scale changes in species' distributions in response to climate change have contributed to substantial concern regarding the fate of

biodiversity in a warming world (e.g. Thomas *et al.*, 2004; Ben Ras Lasram *et al.*, 2010; Sauer *et al.*, 2011; Ihlow *et al.*, 2012; Ochoa-Ochoa *et al.*, 2012; Warren *et al.*, 2013). Numerous studies indicate that, under climate change, a sizeable proportion of species will be faced with an absence of suitable habitat (range extinction), or future habitat that is completely disjunct from current habitat. For example, ~24% of Mexican amphibian species are at risk of losing all habitat by 2080 (Ochoa-Ochoa *et al.*, 2012), while habitat for nearly 12% of the world's Chelonian species may lie in areas currently unsuitable (Ihlow *et al.*, 2012). Depending upon the scenario of greenhouse gas emissions, 11–19% of 1,541 seed plants in Alberta, Canada, were predicted to lose > 90% of current habitat by 2080 (Zhang *et al.*, 2015). Similarly, an analysis of 2,954 species across North and South America predicted the loss of all existing habitat for at least 10% of species by 2071–2100 (Lawler *et al.*, 2009). In one of the largest studies to date, Warren *et al.* (2013) estimated that after

accounting for dispersal, 2   –6% of 5,382 animal species would lose

$\geqslant$ 90% of current habitat. Hence, predictions by SDMs indicate that

the rate of extinction in the 21st century may exceed estimates of historical extinction events from the fossil record. However, there is on-going debate as to the accuracy of SDMs and whether they will over- or under-estimate range changes and extinction (see review by Bellard *et al.*, 2012).

In recent years the number of methods for fitting SDMs has increased considerably (Elith & Graham, 2009; Elith & Leathwick, 2009; Franklin, 2010), and

now includes distance-based or profile models (e.g. BIOCLIM, Nix, 1986; Booth *et al.*, 2014), statistical models that can fit complex non-linear relationships (such as Generalised Additive Models) and machine-learning techniques (Elith *et al.*, 2008; Elith & Leathwick, 2009). These methods differ in complexity, data requirements and ease of use, and their characteristics have been well-described elsewhere (e.g. Syphard & Franklin, 2009; Franklin, 2010; Xinhai & Wang, 2013).

SDM method choice is influenced by access to software, data availability (e.g. presence only or presence/absence records), user ability and the particular goals of the study (Segurado & Araújo, 2004; Elith & Leathwick, 2009; Ahmed *et al.*, 2015), although different disciplines and geographic regions have traditionally utilized different techniques (Elith & Leathwick, 2009). However, identification of the most appropriate SDM method is complicated by a number of factors, and multi-model comparisons have repeatedly concluded that there is no single 'best' method (Elith *et al.*, 2006; Diniz Filho *et al.*, 2010).

Several studies have suggested that SDM methods with high flexibility in modelling complex species-environment relationships may outperform simpler methods (Elith *et al.*, 2006; Tsoar *et al.*, 2007; Li & Wang, 2013). For instance, machine-learning algorithms (such as Random Forests, Maxent and Boosted Regression Trees) frequently outperform regression-based approaches (e.g. Multivariate Adaptive Regression Splines and Generalised Linear Models) (Bucklin *et al.*, 2015, but see Guillera-Arroita *et al.*, 2015). Others caution the use of complex methods: while these may predict observed occurrence patterns well, they may

result in overfitting, and hence may not necessarily predict an acceptable representation of the species' potential distribution (Diniz-Filho *et al.*, 2010; Li & Wang, 2013). Models with high predictive accuracy for the data used in their calibration may also demonstrate poor transferability (or generality); that is, their performance may decline when projected onto different geographic regions (Heikkinen *et al.*, 2012) or time periods (Fronzek *et al.*, 2011). The Random Forest model, for example, generally has high predictive capacity as determined by AUC (the area under the receiver-operating characteristic curve) calculated on data similar to those used for model-fitting (Coetzee *et al.*, 2009; Virkkala *et al.*, 2010; Yen *et al.*, 2011). However, this model has been shown to have lower transferability than other SDM methods, such as Generalised Linear Models (Heikkinen *et al.*, 2012; Crimmins *et al.*, 2013).

Different SDM methods may also have similar predictive performance (e.g. as quantified by the AUC or True Skill Statistic [TSS]; Allouche *et al.*, 2006) yet generate very different predictions of suitable habitat (Beaumont *et al.*, 2009; Parviainen *et al.*, 2009; Syphard & Franklin, 2009). These differences can be magnified when models are projected onto alternative climate scenarios (for which there is also no 'best' choice), with different SDMs varying in the magnitude and direction of predicted changes (Pearson *et al.*, 2006; Beaumont *et al.*, 2007; Beaumont *et al.*, 2008). Model performance can also be influenced by a species' characteristics and its distribution data (Syphard & Franklin, 2009; Dobrowski *et al.*, 2011; García-Callejas & Araújo, 2015), selection of predictor variables (Barbet-

Massin & Jetz, 2014) and geographic location (Engler *et al.*, 2011). As such, the choice of SDM method constitutes the primary source of variation in predictions of species' future distributions (Buisson *et al.*, 2010; Garcia *et al.*, 2012).

These factors have led to the current trend to combine multiple models into 'ensembles' (Araújo & New, 2007, Beaumont *et al.*, 2009, Marmion *et al.*, 2009b; Grenouillet *et al.*, 2011; Garcia *et al.*, 2012; Xinhai & Wang, 2013). Ensemble modelling experiments can then be used to distinguish regions of model consensus (i.e. where most models agree that the environment is suitable/unsuitable) from regions where there is disagreement. A number of studies have found that ensembles outperform individual models (Marmion *et al.*, 2009a; Yen *et al.*, 2011; Grenouillet *et al.*, 2011; Crossman *et al.*, 2012), however, the use of ensembles can still be problematic (Elith *et al.*, 2010; Rapacciuolo *et al.*, 2012; Crimmins *et al.*, 2013) as errors/biases in individual SDMs may lead to incorrect conclusions being drawn from the ensemble.

While previous studies have assessed sources of variation in SDM output (e.g., Dormann *et al.*, 2008; Nenzen & Araujo, 2011; Watling *et al.*, 2015), less attention has been given to whether particular SDM methods are biased towards predicting substantial loss or gain of suitable habitat, and if so, why. Therefore, in this study, we explored 'extreme' range changes predicted by 14 SDM methods incorporated into the R packages *biomod2* (Thuiller *et al.*, 2012) and *dismo* (Hijmans *et al.*, 2011). In particular, we asked whether these SDM methods predicted 'extreme' range changes with equal frequency. We defined five non-mutually

9

exclusive categories of extreme change, based on stability or loss of current habitat or the dislocation of current and future habitat: a) no future habitat (range extinction), b) low stability of current habitat (i.e. ≤ 10% of current habitat remains, although future habitat may be predicted in new locations), c) no gain of habitat in new locations (future habitat is identical to, or a subset of, current habitat), d) all future habitat is in new locations (future habitat does not overlap with current habitat), and e) substantial increase in size of habitat (i.e. future habitat is at least twice the size of current habitat). We modelled future habitat suitability for 220 mammal species across Australia and compared the proportion of predictions from each SDM method that calculated these extreme range changes.

Methods

Species records

For this study, we focused on native terrestrial mammal species in Australia, as their distributions are generally well known and their taxonomy has been resolved. Occurrence records for the 338 extant mammal species recorded on the National Species List (NSL) were downloaded from the Atlas of Living Australia (ALA, www.ala.org.au). The ALA aggregates species records from a wide array of providers including the major museums across the country and maintains the NSL. We applied a number of filters before downloading data, requiring that records be resolved to species-level, observations were made after January 1950, and were

georeferenced. Observations were discarded if the referenced location was classified by ALA as an environmental outlier. After downloading, records were aggregated to a spatial resolution of 5 × 5 km, and reduced to a maximum of one record per species, per cell. We excluded species for which there were fewer than ten occurrence records, since models fit with so few data may be unreliable (Stockwell & Peterson 2002; Hernandez *et al.,* 2006; van Proosdij *et al.* 2016). This left 220 species with between 10 and 7,137 records that, when combined, spanned the breadth of the continent (Supplementary Information Table S1). These species belong to seven orders: Chiroptera (bats, n = 54 species), Dasyuromorphia (carnivorous marsupials, n = 46), Diprotodontia (e.g. kangaroos, wallabies, possums, n = 66), Monotremata (platypus, echidna, n = 2), Notoryctemorphia (marsupial mole, n = 1), Peramelemorphia (omnivorous marsupials such as bandicoots and bilbies, n = 5) and Rodentia (native rats, n = 46) (Supplementary Information Table S1).

Current climate data

Eight bioclimatic variables for the current period (1950-2000) were derived using ANUCLIM 5.1 (Houlder *et al.,* 2000), at a resolution of 5 × 5 km. These variables were selected due to their influence on the physiological and distributional limits of vertebrates, and included annual mean temperature (T), temperature seasonality (TS), maximum temperature of the warmest month (Tmax), minimum temperature of coldest month (Tmin), annual precipitation (Pr), precipitation seasonality (PS), precipitation of the driest quarter (Pdry) and precipitation of wettest quarter (Pwet).

We recognise that there are correlations among these data. We have elected to utilise a common set of variables to provide consistency between models, thereby aiding direct comparisons. For multi-species studies such as ours where, combined, the target species cover a broad geographic region it is unlikely that a common set of predictor variables that are uncorrelated across all species can be identified. Generating distinct subsets for individual species would be worthwhile but is computationally challenging.

Future climate data

We utilised future climate data developed by Reside *et al.* (2013), for the 30-year period centred on 2055 and based on the RCP8.5 greenhouse gas concentration trajectory (the highest of the four Representative Concentration Pathways presented in the Intergovernmental Panel on Climate Change's Fifth Assessment Report, with radiative forcing exceeding 8.5 Wm^{-2} by 2100, see Moss *et al.*, 2010). We chose two global climate models (GCM) for comparison, which represented contrasting futures based on similarity to Australia's current climate in terms of average temperature and precipitation. In doing so, our aim was to take account of the existing variability in future climates and test the SDM methods with the output from two GCMs that represent the range of this variability. Of the 18 GCMs we examined, the National Centre for Atmospheric Research's Community Climate System Model v 3.0 (NCAR-CCSM3; Collins *et al.*, 2004) was the most similar to current temperature and

precipitation (averaged across the continent), whereas the Max Planck Institute for Meteorology's atmospheric general circulation model (MPI-ECHAM5; Roeckner *et al.*, 2006) extrapolated the most for temperature and was among the most different for precipitation.

Species distribution models

To explore variability across alternate SDM methods, we used an ensemble of 14 algorithms included in the R packages *dismo* v1.0-5 (Hijmans *et al.*, 2011) and *biomod2* v3.1-64 (Thuiller *et al.*, 2012). These represent the gamut of approaches used by ecologists and conservation practitioners, and include (i) three regression-based approaches (Generalised Linear Models [GLM, McCullagh and Nelder, 1989], Generalised Additive Models [GAM, Hastie *et al.*, 1990], Multivariate Adaptive Regression Splines [MARS]), (ii) four machine-learning methods (Artificial Neural Networks [ANN, Lek & Guégan, 1999], Random Forests [RF, Breiman, 2001], Boosted Regression Trees [BRT, included in both packages, Friedman, 2001; Elith *et al.*, 2008], and Maxent [Phillips *et al.*, 2006]), (iii) two classification schemes (Classification Tree Analysis [CTA, Brieman *et al.*, 1984], Flexible Discriminant Analysis [FDA, Hastie *et al.*, 1994]), and (iv) four distance-based or profile methods (Surface Range Envelope [SRE], BIOCLIM [Nix, 1986; Busby, 1991], Domain [Carpenter *et al.*, 1993], Mahalanobis [Mahal, Farber & Kadmon, 2003]). Of the above, *biomod2* was used to fit models GLM, GAM, MARS, ANN, RF, CTA, FDA and SRE and the first implementation of BRT (denoted from now as BRT_{bio});

BIOCLIM, Domain, Mahal and the second implementation of BRT (denoted as BRT$_{dis}$) were derived from *dismo*. Maxent was implemented separately via the command line. Characteristics of each algorithm, in the context of species distribution modelling, have been described in detail elsewhere (Elith *et al.*, 2006; Phillips *et al.*, 2006; Phillips & Dudik, 2008; Elith *et al.*, 2011; Hijmans & Elith, 2013; Xinhai & Wang, 2013), and default settings are listed in Supplementary Information Table S2.

We ran most models with their default settings (Supplementary Information Table S2). Maxent was run with threshold features disabled, while "fadebyclamping" was enabled for its predictions. Note also that the two BRTs differed in their implementation, with the *dismo* function using a cross-validation procedure to estimate the optimal number of boosting trees. Further, BRT$_{dis}$ has a default tree complexity of 1 (meaning that it fits only additive terms) while in BRT$_{bio}$ (where complexity is referred to as *interaction.depth*) this value is 7. As such, BRT$_{bio}$ allows up to seven-way interactions, and can thus fit models of far greater complexity. Similarly, SRE (*biomod2*) and BIOCLIM (*dismo*) are different modifications of Nix's (1986) original BIOCLIM. This model is a straight-forward boxcar or parallelepiped approach, whereby the species envelope is bounded by the minimum and maximum values for all presences (or adjusted to exclude outliers) (Booth *et al.*, 2014). The default setting for this model (SRE) in *biomod2* excludes values above the 97.5th percentile, and below the 2.5th percentile for each variable, preventing these from setting the tolerance boundaries for the species (Thuiller *et al.*, 2012). As such,

predictions are binary (suitable/unsuitable). In contrast, in *dismo*'s BIOCLIM a greater range of percentiles are calculated, with the assumption that areas closer to the median are more suitable (Hijmans *et al.*, 2011), and allowing for output that is continuous.

Background data: Absence data were unavailable for this study, and it is highly likely that occurrence records sourced from natural history collections contain sampling biases. One approach to reducing the impact of these biases is to utilise background data that have the same sampling distribution as the occurrence records (i.e. target-group background) (Elith & Leathwick, 2007; Phillips & Dudik, 2008). Hence, background data comprised the occurrence records for all mammals. We note, however, that this approach may also inflate accuracy measures for species with narrow distributions, due to high specificity (see Lobo *et al.*, 2008).

Model calibration and performance: The predictive performance of models was assessed by the True Skill Statistic, TSS. In the context of SDMs, TSS is a threshold-dependent measure based on sensitivity (Se) and specificity (Sp), or the probability that the model correctly predicts true presences and true absences, respectively (Allouche *et al.*, 2006; Liu *et al.*, 2009). Specifically, TSS = Se + Sp − 1. Values therefore range from -1 to 1, with values below 0.4 indicating poor model discrimination between occurrence and background locations (i.e., Se and Sp are, on average, less than 0.7) and values close to 1 indicating excellent discrimination.

We used ten-fold cross-validation to estimate out-of-sample predictive performance. For each species, this entailed splitting model-fitting data into ten subsets of approximately equal size (i.e., folds), fitting each model to nine of the ten folds, and predicting to the tenth. This process was repeated until each fold had been used nine times for fitting and once for testing. TSSwas calculated for the predictions to each test fold, with the average of the ten TSSs values used to estimate the performance of the final model (see below).

Final models: Models were calibrated with all available occurrence data and projected onto current climate and two climate futures for 2055 (described above). Continuous predictions were converted to binary layers indicating suitable/unsuitable habitat, with threshold values chosen to maximise the sum of sensitivity and specificity (equivalent to maximising TSS) as this threshold tends to well reflect the modelled species' prevalence (Jimenez-Valverde & Lobo, 2007). Further, this threshold is less sensitive to modelling method, and has been demonstrated to produce less variation across the output of multiple models than most other thresholds (Nenzen & Araujo, 2011). Models with an average cross-validated TSS below 0.4 were removed from further analyses, since these are considered to have low accuracy (see BIOMOD manual from Thuiller *et al*. 2012).

Statistical Analyses

Changes in habitat suitability predicted by SDM methods can be summarised as the proportion of current habitat (cH) that remains suitable in the future (i.e. is 'stable'), and the proportion of future habitat (fH) that is located in areas currently unsuitable (i.e., 'gain' of new habitat). Based on these definitions of stability and gain, we assessed five non-mutually exclusive categories of extreme changes to the spatial configuration of suitable habitat. When graphed along these two axes models may predict a number of extreme outcomes (Fig. 1). For the purposes of this study, we categorised extreme to include the following, although we note that 'extreme' may be context dependent:

a) No fH (i.e. range extinction);

b) Low stability of cH (< 10% of cH remains stable, although gains may occur elsewhere);

c) No gain of new habitat (i.e. fH is entirely contained within cH; this excludes predictions that meet criterion (a) of no fH);

d) All fH is gain (i.e. all fH is in previously unsuitable areas); and

e) fH is at least double the size of cH.

For each of the above, we tested the null hypothesis that all 14 SDM methods used in this study predict that extreme scenario with equal frequency. Under the null hypothesis, the expected frequency of extreme$_i$, predicted by SDM$_j$, was calculated as the number of predictions made by SDM$_j$ as a fraction of the total number of predictions (i.e. made by all 14 SDMs), multiplied by the total number of instances of

extreme$_i$. Chi-squared Goodness-of-fit tests were used to assess the null hypothesis. When p < 0.05, we calculated the standardised residuals (z) of the contingency table to identify for which SDM(s) the observed frequency (f_o) deviated significantly from the expected (f_e).   A larger z means a greater difference between f_o and f_e, and provides an indication of the contribution of the SDM method to the significance of the Chi-squared test. For any SDM method with $|z| > 2$, f_o differs from f_e at a significance level of p < 0.05. Negative and positive z indicate that f_o is less than or greater than f_e, respectively (Sheskin, 2004).

Scatterplots of stable habitat versus habitat gain were created to reveal any patterns amongst this relationship across either SDM method or taxonomic order. All statistics were calculated in R v. 3.1.2 (R Development Core Team, 2014).

Results

SDM predictive power

We calibrated 3,080 models (220 species × 14 SDM methods), of which 160 returned errors and were excluded from further analysis. These were primarily CTA (133 models) and ANN (28). In each case, the model failed due to a single value being predicted (i.e. all presences or all absences), resulting from the model being unable to discriminate between these. For the remaining 2,920 models, predictive performance measured using TSS averaged 0.784 (SD = 0.158), although this varied considerably across the 14 algorithms (Fig. 2a). Classification methods

18

generally had poorer performance, in terms of TSS, than machine learning techniques (ranging from mean TSS for CTA 0.537 ± S.D. 0.240 to mean TSS for RF 0.997 ± S.D. 0.008).

We excluded from further analyses 226 models (7.7%) that did not meet the TSS threshold of 0.4. Of these, 68 were modelled by Mahal and 24 by CTA. In contrast, all RF models, and all but one GBA and Maxent model exceeded the threshold. Hence, the number of projections in our final analysis was 2,694 (Table 2).

Among the five mammal taxonomic orders, mean TSSvalues were lowest for Notoryctemorphia, which is represented here by a single species (0.634 ± 0.183), and highest for Diprotodontia (0.822 ± 0.137) and Peramelemorphia (0.829 ± 0.107) (Supplementary Information Fig. S1).

In general, distance-based methods tended to predict a larger area of current habitat than other SDM methods, with projections by Mahal being, on average, two orders of magnitude greater ($3.03 \times 10^6 \pm 2.27 \times 10^6$ km^2), than those from RF ($3.55 \times 10^4 \pm 7.16 \times 10^4$ km^2) (Fig. 2b).

Predictions of future distributions - overview

SDM methods differed in how frequently they projected extreme range changes. Across all species, SDM methods and climate futures, changes to the overall area of suitable habitat were right skewed, with 66.3% of models predicting declines to the area of *fH* relative to *cH* (median change = -25.3% [10:90th percentiles: -90 to 121.8%]). Only 6.3% of models predicted that the area of *fH* would be similar to that

of *cH* (i.e. within 5%) while 4.2% predicted there to be no suitable habitat in the future (extreme *a*). At the other end of our spectrum of extremes, 11.5% of models predicted that the area of *fH* would be at least double that of *cH* (extreme *e*), and more than half of these (7.7%) predicted the area of suitable habitat to triple.

Of course, range changes comprise a mix of contraction of *cH* and expansion into new areas (i.e. 'gain'). Total loss of *cH* was projected for ~9% of models, with an additional 10% predicting that $\leqslant$ 10% of *cH* would remain by 2050 (extreme *b*). Approximately 5.6% of projections reported no gain in *fH* (extreme *c*) while 5.4% predicted that all *fH* would lie in currently unsuitable areas (extreme *d*).

Projections onto ECHAM5 (which was the least similar to current conditions) predicted greater and more variable changes in range size than projections onto the CCSM3 climate (ECHAM5: median -38.7% [10-90th percentile; -94.8%, 167.4%]; CCSM3: median -11.6% [10-90th percentile; -80.5%, 84.1%]).

Four mammal orders were represented by > 5 species (Chiroptera, Dasyuromorphia, Diprotodontia, Rodentia). For Chiroptera, < 1% and 2% of projections predicted no *fH* (CCSM3 and ECHAM5) while 12.8% and 21.6% predicted range area to more than double. Across the models for Diprotodontia, 6% predicted no *fH*, while 7.1% predicted *fH* to double (ECHAM5). Current suitable habitat was predicted to decline substantially (> 90%) for 27% of Rodentia models while 17% predicted *fH* to double (ECHAM5).

Do SDM methods project extreme spatial changes with equal frequency?

We assessed the hypothesis that the 14 SDM methods were equally likely to project particular range changes, i.e. f_e (expected frequency) of a given extreme equals the number of projections of that extreme by SDM$_i$ as a fraction of all SDM projections. In total, 5,388 projections of future distributions were analysed (2,694 for each climate future). The total number of models with sufficient predictive capacity (TSS $\geqslant$ 0.4) varied across algorithms, resulting in f_e ranging from 0.023 (CTA) to 0.082 (RF) (Table 1).

Chi-squared test statistics ranged from 49.3 to 297.6 for the five categories of extreme change and were significant to at least p = 0.0001 (Table 3). This demonstrates that the 14 SDM methods are unequally represented in each category (Supplementary Information Table S2). Across the five categories, RF was frequently the most under- or over-represented method.

a)   No future habitat (fH)

Of the 2,694 models projected onto both climate futures (ECHAM5 and CCSM3), 144 (5.3%) and 80 (3.0%) predicted total loss of *fH*, respectively. RF accounted for approximately half of these, making it significantly more likely than the other methods to return this result ($\square$ = 12.85, 16.13 for ECHAM5 and CCSM3, respectively), although SRE was also significantly over-represented in the ECHAM5 climate future ($\square$ = 5.04). In contrast, MARS, Mahal and BRT$_{dis}$ were significantly under-represented ($\square$ -2.187 to -3.13) with 0-3 projections each within this category.

Observed frequencies for ANN, BIOCLIM, CTA, Domain and BRT$_{bio}$ did not differ significantly from expected ($|\hat{\chi}| < 2$) (Fig. 3).

b)   *Low stability of current habitat (cH)*

Current suitable habitat was projected to decline in size by at least 90% for 575 (21.3%) and 318 (11.8%) models under the ECHAM5 and CCSM3 future climates, respectively. Again, RF was significantly over-represented in both future climates, comprising 17.7 and 30.5% of these projections ($\hat{\chi} = 8.38$; 14.54). Indeed, over 40% of RF projections fell into this category. CTA was also over-represented, with 33.3% ($\hat{\chi} = 3.46$; CCSM3) and 41.3% ($\hat{\chi} = 5.03$; ECHAM5) of the 63 CTA models predicting extensive loss of *cH*. In comparison, BRT$_{dis}$, GLM, MARS and Maxent were significantly under-represented, with between 2.9-13.4% of their projections falling into this category (with $\hat{\chi}$ ranging from -2.52 to -4.051). Observed frequencies of both FDA and GAM did not differ significantly from expected for either climate future.

c)   *No gain of new habitat*

This category, in which ~2% of projections fell, denotes models for which *fH* is entirely contained within areas currently classified as suitable. Significantly under-represented in this category were projections from BIOCLIM ($\hat{\chi} = -2.32$; -2.13, ECHAM5 and CCSM3 respectively) and SRE, ($\hat{\chi}$ ranging from -2.30 to -2.11), while GLM was over-represented ($\hat{\chi} = 3.44$, 3.61, ECHAM5 and CCSM3, respectively). Under the ECHAM5 climate future, ANN and Maxent were also over-represented, while FDA was for CCSM3. BRT$_{dis}$, CTA, GAM, BRT$_{bio}$ and Mahal did not differ from expected for either climate future.

d)   *All* fH *is gain*

For ECHAM5 and CCSM3, respectively, 231 (8.6%) and 105 (3.9%) models projected that all *fH* would lie in areas currently classified as unsuitable, that is, that suitable habitat by 2055 would be completely disjunct from current habitat. Again, RF was significantly over-represented in both climate futures ($\hat{z}$ = 5.56, 9.06), as was CTA ($\hat{z}$ = 6.36, 6.16) and, to a lesser extent, BRT$_{bio}$ ($\hat{z}$ = 2.20, 3.74). These three algorithms combined accounted for 38% (CCSM3) and 62% (ECHAM5) of the projections that fell into this category. Mahal, MARS and Maxent were significantly under-represented (with $\hat{z}$ ranging from -2.29 to -3.80), while the frequencies for ANN, BIOCLIM, Domain, and GAM did not differ from expected.

e)   fH *is, at least, double the size of* cH

This category consisted of 13.9% (ECHAM5) and 9.1% (CCSM3) of projections. There was substantial variation across SDMs in the frequency with which their projections fell within this category. For instance, none of the models fitted using the four distance-based methods (BIOCLIM, Domain, Mahal, SRE) were included in this category for either climate future ($\hat{z}$ between -3.82 and -5.61). However, BRT$_{bio}$ ($\hat{z}$ = 6.36, 4.96) and RF ($\hat{z}$ = 6.13, 6.10), and to a lesser extent FDA and GAM, were significantly over-represented. Indeed, for both BRT$_{bio}$ and RF, 29% of their projections under the ECHAM5 climate future predicted *cH* to at least double in area. BRT$_{dis}$, CTA and Maxent were the only models for which f_o did not differ significantly from f_e. These patterns are clear in scatterplots of the proportion of *cH* remaining stable versus the proportion of *fH* in new habitat: distance-based methods have few

23

points in the upper-right quadrant in contrast to RF (Supplementary Information Fig. S2).

Consistency of patterns across taxonomic orders

For each SDM, we visually assessed consistency in projections of stability and gain of new habitat across the various mammal orders. Within SDMs, the relationship between stability and gain was generally consistent across taxonomic orders (Supplementary Information Fig. S2). For some SDMs, there was an apparent negative relationship between these variables, which was most pronounced for distance-based methods (BIOCLIM, Domain, Mahal, SRE). When comparing GCMs, there were no clearly discernible differences in these relationships, though retention of existing suitable habitat in some cases appeared to be greater for CCSM3 than for ECHAM5 (which presented a climate future more different to current conditions than CCSM3).

Discussion

Species distribution models differ significantly from one another in the frequency with which they project extreme changes to the future size and configuration of suitable habitat. These changes (such as the total loss of current suitable habitat, future habitat that is completely disjunct from current habitat, and substantial increases in the future size of habitat) are extreme from a spatial perspective only – this does not

mean that such outcomes are rarely projected. To the contrary, approximately one-quarter of our model predictions fell into at least one category of extreme.

Which SDM methods projected extreme range changes more (or less) frequently than others, and why?

Such dramatic changes to the size of species' climatically suitable habitat, due to climate change projected to occur in coming decades, is disconcerting. Yet, this outcome has been projected for numerous species across a multitude of studies (e.g. Ihlow *et al.*, 2012; Ochoa-Ochoa *et al.*, 2012; Warren *et al.*, 2013; Krause *et al.*, 2015). However, our results demonstrate that the frequency with which such extreme range changes are projected may be related to the choice of SDM. For instance, we found that RF projected substantial range loss for almost half of the 220 species studied - a frequency significantly at odds with other SDM methods. Similarly, CTA and SRE were significantly more likely to project loss of current habitat (*cH*), while CTA and RF frequently projected future habitat (*fH*) to be located in areas currently unsuitable. In stark contrast to the other methods, the four distance-based approaches (BIOCLIM, Domain, Mahal, SRE) never projected extreme increases in the area of suitable habitat.

These results are a function of how a model fits the species-environment relationship and selects from competing, and potentially correlated, predictors. It may also be a function of model parameterisation and the details of extrapolation to environments beyond those represented by the training data.

25

The goal of SDM studies is to describe accurately and precisely the species-environment relationship and, frequently, to project this relationship onto data for other time periods or geographic regions. Given the same input data, the outputs of competing methods will differ because their fundamental approaches to describing the species-environment relationship and capturing interactions among variables, as well as their parameterisation and ability to interpolate/extrapolate, also differ (Merow *et al*., 2014). For example, BIOCLIM, a simple boxcar approach, generally sets a species' environmental tolerance to the most extreme values in the training records, making this method very sensitive to outliers or insufficient sampling of the environment. Interactions between variables are not considered; all variables are weighted equally and extrapolation beyond the environmental envelope is not possible. Further, all areas within the envelope are deemed equally suitable, although note that the *dismo* version of BIOCLIM calculates a greater range of percentiles from the training data, with the assumption that grid cells closer to the 50th percentile are more suitable (Hijmans *et al*., 2011).

In contrast, machine learning methods, such as Maxent and RF, can identify non-linear species-environment relationships and may be more efficient at modelling complex patterns, possibly leading to more accurate predictions than GLMs or GAMs (Thibaud *et al*., 2014). Unless model complexity is managed, however (e.g., with regularisation; Warren et al. 2011), machine learning techniques are prone to describing noise in the training data, leading to poor generality.

Therefore, the extent to which a model can estimate realistic response curves is critical, and errors may result in models that are under- or over-fitted (for a description of simple and complex response curves from different methods, see Table 1 in Merow *et al*., 2014). Simplistic models that fail to capture the underlying relationship are said to under-fit. These models have low variance, in that they can reproduce the result consistently, but are highly biased, i.e. model-fitting data are poorly predicted. Conversely, over-fitting occurs when noise or random error within the data is described by the model (Radosavljevic & Anderson, 2014), generating unrealistically complex response curves. In this instance, the model may describe the occurrence pattern accurately (i.e., it has low bias, and hence high predictive performance, with respect to model-fitting data), but transferability to other locations or time periods may not be equally high. Both under- and over-fitting may lead to poor predictions across space and time (Merow *et al*., 2014).

Of interest is the propensity for RF to report very high TSS or AUC values while also projecting extreme range changes (also see Guo *et al*., 2015). Previous studies have found this model to interpolate accurately but extrapolate or transfer poorly (e.g., see Wenger & Olden, 2012). This model produces very jagged response curves that may fit data well but are biologically unrealistic, and are unlikely to transfer (Wenger & Olden, 2012).

Typically, the predictive performance of a model is measured by splitting data into training and test sets. However, the two subsets are rarely independent (Veloz 2009; Radosavljevic & Anderson, 2014). Hence, performance measures derived

27

from them may be optimistic, and fail to reflect the performance of a model when it is projected onto environmental data from different regions or time periods. Crimmins *et al*. (2013) used five SDM methods to model the distributions of 145 plant species. Models calibrated and tested using cross-validation procedures, where 25% of data were randomly selected and withheld from calibration, were compared to models tested with temporally independent data. They found that, as with other studies (e.g. Grenouillet *et al*., 2011; Bucklin *et al*., 2015; Guo *et al*., 2015), RF returned the highest AUC values when tested with non-independent data (i.e. via cross-validation) (mean ~0.98). However, for all models, AUC values were lower when tested with the independent data, with RF the second lowest (mean ~0.77) indicating poor transferability. This was primarily the result of high rates of false positives (median ~33%), with one-quarter of RF models having an FP rate of 1.0. Similarly, Heikkinen *et al*., (2012) assessed the interpolation and extrapolation ability of 10 SDM methods against spatially independent data. Interpolation accuracy was highest for RF, Maxent and BRT$_{bio}$. Extrapolation proved more difficult for all methods, particularly RF and MARS.

The way in which the fitted functions of a model are constrained during extrapolation will also influence its performance when doing so (Elith & Graham, 2009). For instance, Maxent by default utilises a 'clamping' approach whereby the response to environmental variables beyond their calibration range is kept at a constant value (Elith & Graham, 2009). Similarly, CTA, BRT$_{dis}$ and RF extrapolate at a constant value from the most extreme environmental value (Elith & Graham, 2009).

GLMs with cubic and quadratic functions continue the fitted trend beyond the most extreme environmental value, which can cause unrealistic results, although options do exist to control extrapolation (Elith & Graham, 2009). Stohlgren *et al*. (2011) note that restricting extrapolation by CART and Maxent can greatly reduce gain of habitat in new locations, and suggest that bounding techniques be included in best practices for SDMs. In contrast to the above models, distance-based measures tend not to extrapolate to conditions outside their training data, instead assuming these conditions are unsuitable (e.g. see Pearson *et al*., 2006).

An additional consideration is multicollinearity. We used the same eight variables for each SDM but the direction and magnitude of collinearity among these varies spatially and temporally. For instance, across northern Australia, the driest months are generally in winter while to the south they are in summer. The consequences of collinearity depend on the method, but in general, the accuracy and precision of estimated coefficients (and thus partial response curves) are lower for correlated predictors than for uncorrelated predictors. Further, small changes in values of the input data can result in large changes in the model. This can lead to unreliable projections onto alternate climate scenarios, particularly if correlations change in direction (Elith *et al*., 2010).

Implications for projections of potential range shifts

How sensitive are projections of range changes to model selection and parameterisation? Our results demonstrate that the frequency with which extreme

range changes are predicted can vary significantly across SDM methods. To an (unknown) extent, this may be due to the choice of model settings as can be seen by comparing the results for BRT_{bio} and BRT_{dis}. The frequency with which these methods projected extreme range changes differed substantially. For instance, while BRT_{bio} was significantly more likely than expected to project substantial declines to *cH* and that *fH* would be mostly disjunct from *cH*, BRT_{dis} had the opposite response, being significantly less likely than other methods to project these extremes. Further, BRT_{bio} frequently projected *fH* to be at least double the size of *cH*, while BRT_{dis} was unlikely to project no *fH*.

To date, there is little guidance on approaches to parameterising SDMs (although see Merow *et al*., 2014), and for some studies parameterisation of each model for each species may be prohibitively time-consuming or, in the case of small or biased datasets, may be unreliable (Phillips & Dudik, 2008). Hence, default settings for numerical model parameters are typically relied upon, with the assumption that they are appropriate as they were previously tested on large datasets and were found to perform well (e.g., see Elith *et al*., 2006; Phillips & Dudik, 2008).

Many of the settings used in the R packages for SDMs are the program's default values (e.g., as with FDA within *biomod2*; pers.comm. W. Thuiller), while others may be the result of trial and error by package authors. For example, within *biomod2,* the default value for the maximum number of iterations used by ANN was 200. This value was derived experimentally and is the result of a trade-off between

computing time and efficiency (pers.comm. W. Thuiller). Similarly, default settings for Maxent were selected based on assessments across diverse datasets (Phillips & Dudik, 2008). Default settings, however, should be used with caution, since they do not always result in optimal models (e.g., Moreno-Amat *et al.*, 2015). Further, experimentally derived settings are usually optimised according to their ability to accurately predict independent test data rather than their accuracy when projecting to other locations or time periods. However, the purpose that model settings are tuned for is not necessarily the purpose for which the models are being used. That is, a model with high predictive performance based on current distributional data may not be particularly effective at predicting the effects of climate change, and so default values, or "best practices", may not be applicable depending on the application of the model.

Systematic evaluations of alternative values of settings for individual algorithms are rarely undertaken. In their review of the SDM literature, Hallgren and Mackey (2014) found that the majority of SDM studies do not provide values for algorithm configuration options, and when available the justifications for these values were usually minimal. The overall lack of documentation of the rationale for values for model configuration options has the potential to be of concern. Without knowing how sensitive a model - and model output - is to these values, then arbitrarily assigned values and the lack of a tested range of suitable values could lead, in an unpredictable manner, to erroneous and/or unrealistic model results (Hallgren and Mackey, 2014). Optimisation of most models is possible, however, comparisons

within and between models undergoing optimisation may be complex (Merow *et al*.,

2014). Further, optimisation has generally not been automated within SDM packages

(but see BRT$_{dis}$) hence, large multi-species studies may be limited to default settings.

Conclusion

Crucially, our study shows that model selection and parameterisation will influence

inferences of the likelihood of species facing extreme changes to the distribution of

suitable habitat as climate changes. To summarise, we found that compared to other

models, distance-based measures are significantly less likely to project massive

increases to the size of suitable habitat, while SRE is highly likely to project no future

habitat; CTA projects substantial loss of current habitat and that future habitat will be

disjunct from current habitat more frequently than expected, while Maxent predicts

these outcomes less frequently than expected; RF frequently projects extreme range

changes – indeed this model is the least likely to project stability of current habitat

under alternative climate scenarios.

For some models, these patterns may be influenced by parameterisation, the

impact of predictor collinearity on the model, and the way in which models

extrapolate beyond the training data. A comprehensive sensitivity analysis of the

predictions of SDM methods to their settings would provide valuable guidance on

setting sensible parameter values for the models. Future developments in coding to

automate optimisation procedures (similar to BRT$_{dis}$) will greatly assist with fine-

tuning models used for multi-species studies rather than relying on a 'one-size-fits-all' approach.

Acknowledgments

References

Ahmed, S.E., McInerny, G., O'Hara, K., Harper, R., Salido, L., Emmott, S. & Joppa, L.N. (2015) Scientists and software–surveying the species distribution modelling community. *Diversity and Distributions*, **21**, 258-267.

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

Barbet-Massin, M. & Jetz, W. (2014) A 40-year, continent-wide, multispecies assessment of relevant climate predictors for species distribution modelling. *Diversity and Distributions*, **20**, 1285-1295.

Beaumont, L.J., Hughes, L. & Pitman, A.J. (2008) Why is the choice of future climate scenarios for species distribution modelling important? *Ecology Letters*, **11**, 1135-1146.

Beaumont, L.J., Pitman, A.J., Poulsen, M. & Hughes, L. (2007) Where will species go? Incorporating new advances in climate modelling into projections of species distributions. *Global Change Biology*, **13**, 1368-1385.

Beaumont, L.J., Gallagher, R.V., Thuiller, W., Downey, P.O., Leishman, M.R. & Hughes, L. (2009) Different climatic envelopes among invasive populations may lead to underestimations of current and future biological invasions *Diversity and Distributions*, **15**, 409-420.

Bellard, C., Bertelsmeier, C., Leadley, P., Thuiller, W. & Courchamp, F. (2012) Impacts of climate change on the future of biodiversity. *Ecology Letters*, **15**, 365-377.

Ben Ras Lasram, F., Guilhaumon, F., Albouy, C., Somot, S., Thuiller, W. & Mouillot, D. (2010) The Mediterranean Sea as a 'cul-de-sac' for endemic fishes facing climate change. *Global Change Biology*, **16**, 3233-3245.

Birks, H.J.B. (1989) Holocene isochrone maps and patterns of tree-spreading in the British Isles. *Journal of Biogeography*, **16**, 503-540.

Booth, T. H., Nix, H. A., Busby, J. R. & Hutchinson, M. F. (2014) BIOCLIM: the first species distribution modelling package, its early applications and relevance to most current MaxEnt studies. *Diversity and Distributions*, **20**, 1-9.

Breiman, L. (2001). Random forests. *Machine Learning*, **45**, 5-32.

Breiman L., Friedman J. H., Olshen R. A., and Stone, C. J. (1984) *Classification and Regression Trees.* Wadsworth.

Bucklin, D.N., Basille, M. & Benscoter, A.M. (2015) Comparing species distribution models constructed with different subsets of environmental predictors. *Diversity and Distributions*, **21**, 23-35.

Buisson, L., Thuiller, W., Casajus, N., Lek, S. & Grenouillet, G. (2010) Uncertainty in ensemble forecasting of species distributions. *Global Change Biology*, **16**, 1145-1157.

Busby, J.R. (1991) BIOCLIM - A Bioclimatic 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.

Carpenter, G., Gillison, A. N., & Winter, J. (1993). DOMAIN: a flexible modelling procedure for mapping potential distributions of plants and animals. *Biodiversity & Conservation*, **2**, 667-680.

Chen, I.-C., Hill, J.K., Ohlemüller, R., Roy, D.B. & Thomas, C.D. (2011) Rapid range shifts of species associated with high levels of climate warming. *Science*, **333**, 1024-1026.

Coetzee, B.W.T., Robertson, M.P., Erasmus, B.F.N., Van Rensburg, B.J. & Thuiller, W. (2009) Ensemble models predict Important Bird Areas in southern Africa will become less effective for conserving endemic birds under climate change. *Global Ecology and Biogeography*, **18**, 701-710.

Collins, W.D., Rasch, P.J., Boville, B.A., Hack, J.J., McCaa, J.R., Williamson, D.L., Kiehl, J.T. & Briegleb, B. (2004) Description of the NCAR Community Atmosphere Model (CAM 3.0). In: *NCAR/TN-464+STR NCAR Technical Note*. Climate and Global Dynamics Division, National Center for Atmospheric Research, Boulder, Colorado, USA.

Crimmins, S. M., Dobrowski, S. Z., & Mynsberge, A. R. (2013). Evaluating ensemble forecasts of plant species distributions under climate change. *Ecological Modelling*, **266**, 126-130.

Crossman, N.D., Bryan, B.A. & Summers, D.M. (2012) Identifying priority areas for reducing species vulnerability to climate change. *Diversity and Distributions*, **18**, 60-72.

Diniz-Filho, J.A.F., Ferro, V.G., Santos, T., Nabout, J.C., Dobrovolski, R. & DeMarco Jr, P. (2010) The three phases of the ensemble forecasting of niche models: geographic range and shifts in climatically suitable areas of *Utetheisa ornatrix* (Lepidoptera, Arctiidae) *Revista Brasileira de Entomologia*, **54**, 339-349.

Dobrowski, S.Z., Thorne, J.H., Greenberg, J.A., Safford, H.D., Mynsberge, A.R., Crimmins, S.M. & Swanson, A.K. (2011) Modeling plant ranges over 75 years

of climate change in California, USA: temporal transferability and species traits. *Ecological Monographs*, **81**, 241-257.

Dormann, C. F., Purschke, O., Márquez, J. R. G., Lautenbach, S., & Schröder, B. (2008) Components of uncertainty in species distribution analysis: a case study of the great grey shrike. *Ecology*, **89**, 3371-3386.

Elith, J. & Leathwick, J. (2007) Predicting species distributions from museum and herbarium records using multiresponse models fitted with multivariate adaptive regression splines. *Diversity and Distributions*, **13**, 265-275.

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

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**, 677-97.

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

Elith, J., Phillips, S.J., Hastie, T., Dudik, M., Chee, Y.E. & Yates, C.J. (2011) A statistical explanation for Maxent for ecologists. *Diversity and Distributions*, **17**, 43-57.

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

Farber, O. & Kadmon, R. (2003) Assessment of alternative approaches for bioclimatic modeling with special emphasis on the Mahalanobis distance. *Ecological Modelling*, 160, 115-130.

Franklin, J. (2010) *Mapping species distributions: spatial inference and prediction*. Cambridge University Press.

Friedman J.H. (2001) Greedy Function Approximation: A Gradient Boosting Machine, *Annals of Statistics* **29**, 1189-1232.

Fronzek, S., Carter, T. & Luoto, M. (2011) Evaluating sources of uncertainty in modelling the impact of probabilistic climate change on sub-arctic palsa mires. *Natural Hazards and Earth System Sciences*, **11**, 2981-2995.

García-Callejas, D. & Araújo, M.B. (2015) The effects of model and data complexity on predictions from species distributions models. *Ecological Modelling*,

Garcia, R.A., Burgess, N.D., Cabeza, M., Rahbek, C. & Araujo, M.B. (2012) Exploring consensus in 21st century projections of climatically suitable areas for African vertebrates. *Global Change Biology*, **18**, 1253-1269.

Garroway, C.J., Bowman, J., Holloway, G.L., Malcolm, J.R. & Wilson, P.J. (2011) The genetic signature of rapid range expansion by flying squirrels in response to contemporary climate warming. *Global Change Biology*, **17**, 1760-1769.

Graham, R.W. (1992) Late pleistocene faunal changes as a guide to understanding effects of greenhouse warming on the mammalian fauna of North America. *Global warming and biological diversity* (ed. by R.L. Peters and T. Lovejoy), pp. 76-87. Yale University, New Haven and London.

Grenouillet, G., L. Buisson, N. Casajus, and S. Lek. 2011. Ensemble modelling of species distribution: the effects of geographical and environmental ranges. *Ecography*, **34**, 9-17.

Guillera-Arroita, G., Lahoz-Monfort, J.J., Elith, J., Gordon, A., Kujala, H., Lentini, P.E., McCarthy, M.A., Tingley, R. & Wintle, B.A. (2015) Is my species distribution model fit for purpose? Matching data and models to applications. *Global Ecology and Biogeography*, **24**, 276-292.

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

Guo, C., Lek, S., Ye, S., Li, W., Liu, J. & Li, Z. (2015) Uncertainty in ensemble modelling of large-scale species distribution: Effects from species characteristics and model techniques. *Ecological Modelling*, **306**, 67-75.

Hallgren, W.S., and Mackey, B. (2014) Analysis of parameter values for the configuration options of statistical Species Distribution Models. Technical Report of the Griffith Climate Change Response Program, Griffith University, Australia, 10pp.

Hastie, T. J., & Tibshirani, R. J. (1990). Generalized additive models (Vol. 43). CRC Press.

Hastie T, Tibshirani R, Buja A (1994) Flexible Discriminant Analysis by Optimal Scoring, Journal of the American Statistical Association, 89, 1255-1270.

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

Hijmans, R.J. & Elith, J. (2013) *Species distribution modelling with R*. http://cran.r-project.org/web/packages/dismo/index.html.

Hijmans, R.J., Phillips, S., Leathwick J. & Elith, J. (2011) *dismo: Species distribution modeling*. http://CRAN.R-project.org/package=dismo.

Houlder, D., Hutchinson, M., Nix, H. & McMahon, J. (2000) ANUCLIM user guide, version 5.1. *Centre for Resource and Environmental Studies, Australian National University, Canberra*,

Huntley, B. (1990) European post-glacial forests: compositional changes in response to climate change. *Journal of Vegetation Science*, **1**, 507-518.

Ihlow, F., Dambach, J., Engler, J.O., Flecks, M., Hartmann, T., Nekum, S., Rajaei, H. & Rodder, D. (2012) On the brink of extinction? How climate change may affect global chelonian species richness and distribution. *Global Change Biology*, **18**, 1520-1530.

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

Krause, C.M., Cobb, N.S. & Pennington, D.D. (2015) Range Shifts Under Future Scenarios of Climate Change: Dispersal Ability Matters for Colorado Plateau Endemic Plants. *Natural Areas Journal*, **35**, 428-438.

Lawler, J.J., Shafer, S.L., White, D., Kareiva, P., Maurer, E.P., Blaustein, A.R. & Bartlein, P.J. (2009) Projected climate-induced faunal change in the Western Hemisphere. *Ecology*, **90**, 588-597.

Lek, S., & Guégan, J. F. (1999). Artificial neural networks as a tool in ecological modelling, an introduction. *Ecological Modelling*, **120**, 65-73.

Li, X. & Wang, Y. (2013) Applying various algorithms for species distribution modelling. *Integrative Zoology*, **8**, 124-135.

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

Maraun, D. (2012) Nonstationarities of regional climate model biases in European seasonal mean temperature and precipitation sums. *Geophysical Research Letters*, **39**, L06706.

Marmion, M., Parviainen, M., Luoto, M., Heikkinen, R.K. & Thuiller, W. (2009) Evaluation of consensus methods in predictive species distribution modelling. *Diversity and distributions*, **15**, 59-69.

McCullagh P. and Nelder, J. A. (1989) *Generalized Linear Models.* London: Chapman and Hall.

Melles, S., Fortin, M.J., Lindsay, K. & Badzinski, D. (2010) Expanding northward: influence of climate change, forest connectivity, and population processes on a threatened species' range shift. *Global Change Biology*, **17**, 17-31.

Merow, C., Smith, M.J., Edwards, T.C., Guisan, A., McMahon, S.M., Normand, S., Thuiller, W., Wüest, R.O., Zimmermann, N.E. & Elith, J. (2014) What do we gain from simplicity versus complexity in species distribution models? *Ecography*, **37**, 1267-1281.

Moreno-Amat, E., Mateo, R.G., Nieto-Lugilde, D., Morueta-Holme, N., Svenning, J.-C. & García-Amorena, I. (2015) Impact of model complexity on cross-temporal transferability in Maxent species distribution models: An assessment using paleobotanical data. *Ecological Modelling*, **312**, 308-317.

Moss, R.H., Edmonds, J.A., Hibbard, K.A., Manning, M.R., Rose, S.K., van Vuuren, D.P., Carter, T.R., Emori, S., Kainuma, M., Kram, T., Meehl, G.A., Mitchell, J.F.B., Nakicenovic, N., Riahi, K., Smith, S.J., Stouffer, R.J., Thomson, A.M., Weyant, J.P. & Wilbanks, T.J. (2010) The next generation of scenarios for climate change research and assessment. *Nature*, **463**, 747-756.

Nix, H. (1986) A biogeographic analysis of Australian Elapid snakes. *Snakes: atlas of elapid snakes of Australia* (ed. by R. Longmore), pp. (4-10). Bureau of Flora and Fauna, Canberra.

Ochoa-Ochoa, L.M., Rodríguez, P., Mora, F., Flores-Villela, O. & Whittaker, R.J. (2012) Climate change and amphibian diversity patterns in Mexico. *Biological Conservation*, **150**, 94-102.

Parviainen, M., Marmion, M., Luoto, M., Thuiller, W. & Heikkinen, R.K. (2009) Using summed individual species models and state-of-the-art modelling techniques to identify threatened plant species hotspots. *Biological Conservation*, **142**, 2501-2509.

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

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

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

R Development Core Team (2014) R: A language and environment for statistical computing. In. R Foundation for Statistical Computing, Vienna, Austria

Radosavljevic, A. & Anderson, R.P. (2014) Making better Maxent models of species distributions: complexity, overfitting and evaluation. *Journal of Biogeography*, **41**, 629-643.

Rapacciuolo, G., Roy, D.B., Gillings, S., Fox, R., Walker, K., & Purvis, A. (2012) Climatic associations of British species distributions show good transferability in time but low predictive accuracy for range change. *Plos One*, **7**, e40212.

Reside, A., VanDerWal, J., Phillips, B., Shoo, L., Rosauer, D., Anderson, B., Welbergen, J., Moritz, C., Ferrier, S., Harwood, T., Williams, K., Mackey, B., Hugh, S. & Williams, S. (2013) *Climate change refugia for terrestrial biodiversity: defining areas that promote species persistence and ecosystem resilience in the face of global climate change*. In. National Climate Change Adaptation Research Facility, Gold Coast, Qld, Australia.

Roeckner, E., Coauthors (2003) The atmospheric general circulation model ECHAM5. Part I: Model description. Max Planck Institute for Meteorology Rep. 349, 127 pp.

Rubidge, E.M., Monahan, W.B., Parra, J.L., Cameron, S.E. & Brashares, J.S. (2011) The role of climate, habitat, and species co-occurrence as drivers of change in small mammal distributions over the past century. *Global Change Biology*, **17**, 696-708.

Sauer, J., Domisch, S., Nowak, C. & Haase, P. (2011) Low mountain ranges: summit traps for montane freshwater species under climate change. *Biodiversity and Conservation*, **20**, 3133-3146.

Segurado, P. & Araújo, M.B. (2004) An evaluation of methods for modelling species distributions. *Journal of Biogeography*, **31**, 1555-1568.

Sheskin, D. (2004) *Handbook of parametric and nonparametric statistical procedures*, 3rd edn. Chapman and Hall, Boca Raton, Florida, USA.

Syphard, A.D. & Franklin, J. (2009) Differences in spatial predictions among species distribution modeling methods vary with species traits and environmental predictors. *Ecography*, **32**, 907-918.

Thibaud, E., Petitpierre, B., Broennimann, O., Davison, A.C. & Guisan, A. (2014) Measuring the relative effect of factors affecting species distribution model predictions. *Methods in Ecology and Evolution*, **5**, 947-955.

Thomas, C., Cameron, A., Green, R., Bakkenes, M., Beaumont, L., Collingham, Y., Erasmus, B., de Siqueira, M., Grainger, A., Hannah, L., Hughes, L., Huntley, B., Van Jaarsveld, A., Midgley, G., Miles, L., Ortega-Huerta, M., Peterson, A., Phillips, O. & Williams, S. (2004) Extinction risk from climate change. *Nature*, **427**, 145-148.

Thuiller, W., Georges, D. & Engler, R. (2012) *biomod2: Ensemble platform for species distribution modeling*. R package.

Tsoar, A., Allouche, O., Steinitz, O., Rotem, D. & Kadmon, R. (2007) A comparative evaluation of presence-only methods for modelling species distribution. *Diversity and Distributions*, **13**, 397-405.

Veloz, S.D. (2009) Spatially autocorrelated sampling falsely inflates measures of accuracy for presence-only niche models. *Journal of Biogeography*, **36**, 2290-2299.

Virkkala, R., Marmion, M., Heikkinen, R.K., Thuiller, W. & Luoto, M. (2010) Predicting range shifts of northern bird species: Influence of modelling technique and topography. *Acta Oecologica*, **36**, 269-281.

Warren, D.L. & Seifert, S.N. (2011) Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. *Ecological Applications*, **21**, 335-342.

Warren, R., VanDerWal, J., Price, J., Welbergen, J., Atkinson, I., Ramirez-Villegas, J., Osborn, T., Jarvis, A., Shoo, L. & Williams, S. (2013) Quantifying the benefit of early climate change mitigation in avoiding biodiversity loss. *Nature Climate Change*, **3**, 678-682.

Watling, J.I., Brandt, L.A., Bucklin, D.N., Fujisaki, I., Mazzotti, F.J., Romañach, S.S. & Speroterra, C. (2015) Performance metrics and variance partitioning reveal sources of uncertainty in species distribution models. *Ecological Modelling*, **309**, 48-59.

Xinhai, L.I. & Wang, Y. (2013) Applying various algorithms for species distribution modelling. *Integrative Zoology*, **8**, 124-135.

Yen, J.D.L., Thomson, J.R., Vesk, P.A. & Mac Nally, R. (2011) To what are woodland birds responding? Inference on relative importance of in-site habitat variables using several ensemble habitat modelling techniques. *Ecography*, **34**, 946-954.

Zhang, J., Nielsen, S.E., Stolar, J., Chen, Y. & Thuiller, W. (2015) Gains and losses of plant species and phylogenetic diversity for a northern high-latitude region. *Diversity and Distributions*, **21**, 1441-1454.

Table 1. Number of models per SDM method. 220 mammal species were modelled using 14 SDM methods. Projections with TSS values < 0.4 were excluded from further analyses, leaving from 63 to 220 species modelled by each method. For a given method, the expected frequency of each type of extreme change in habitat area (see Table 2) is equal to the proportion of all models (2,694) that were fit using that method.

SDM	No. of models	Expected Frequency
ANN	178	0.066
BIOCLIM	209	0.078
BRT$_{bio}$	219	0.081
BRT$_{dis}$	214	0.079
CTA	63	0.023
Domain	208	0.077
FDA	186	0.069
GAM	215	0.080
BRT$_{bio}$	219	0.081
GLM	201	0.075
Mahal	152	0.056
MARS	205	0.076
Maxent	219	0.081
RF	220	0.082
SRE	205	0.076
Total	**2,694**	**1.000**

Table S3 is going here in final version.

Table 3. Chi-squared statistic for each of the 'extreme' range change categories assessed. The statistic is reported for two climate futures (ECHAM5 and CCSM3). *cH* = current habitat; *fH* = future habitat. In all instances, p < 0.0001. df = 13.

	χ^2 statistic	
Extreme range change	**ECHAM5**	**CCSM3**
No *fH*	230.9	282.8
Low stability of *cH*	154.8	297.6
No gain of new *fH*	72.3	49.3
All *fH* is gain	147.4	163.0
fH at least double size of *cH*	217.1	164.9

Figures

Figure 1. Changes in habitat suitability projected by species distribution models can be summarised along two axes: 'Proportion stable' (i.e. proportion of the area of current habitat [*cH*] projected to remain suitable in the future) and 'Proportion gained' (i.e. proportion of the area of future habitat [*fH*] projected to occur in currently unsuitable regions). Based on these two axes we calculated the number of predictions from 14 SDM methods that fell into the following categories: a) no *fH* (Prop. stable = 0, Prop. gained = 0); b) low stability of *cH* (Prop. $\text{stable} \leqslant 0.1$, $\text{Prop. gained} > 0$); c) no gain of new habitat (Prop. stable > 0; Prop. gained = 0); d) all *fH* is currently unsuitable (Prop. stable = 0, Prop. gained > 0); and e) substantial increase in area of *fH* compared to *cH* ($fH \geq 2cH$).

Figure 2. Box-and-whisker plot of AUC and size of current habitat projected for 220 Australian mammal species by 14 SDM methods. SDM methods can be broadly placed into one of four categories: distance-based, classification, regression, and machine learning. The dotted line in 2A indicates the TSS threshold (0.4) below which models were excluded due to poor predictive performance.

Figure 3. Proportional change in size of suitable habitat for Australian mammals under two climate futures. 14 SDM methods were calibrated for 220 species. Models with poor performance (TSS < 0.4) were excluded from these analyses, resulting in 63–220 species being modelled per SDM method (see Table 1). Data are presented on a cube-root scale.

Figure 4. Standardised residuals from Chi-squared goodness-of-fit tests. Chi-squared tests were used to assess whether 14 SDM methods vary from each other in the frequency with which they project five categories of extreme changes to the size of habitat for 220 Australian mammal species, under two climate futures. Standardised residuals can be used to identify which SDM method contributed the most to the significance of the Chi-squared test. Here, residuals more extreme than ±2 indicate that the observed frequency of the SDM methods differed significantly from expected ($p < 0.05$). This threshold is represented by the two dotted lines. Figure 1 describes the categories of extreme range changes.

Figure 1.
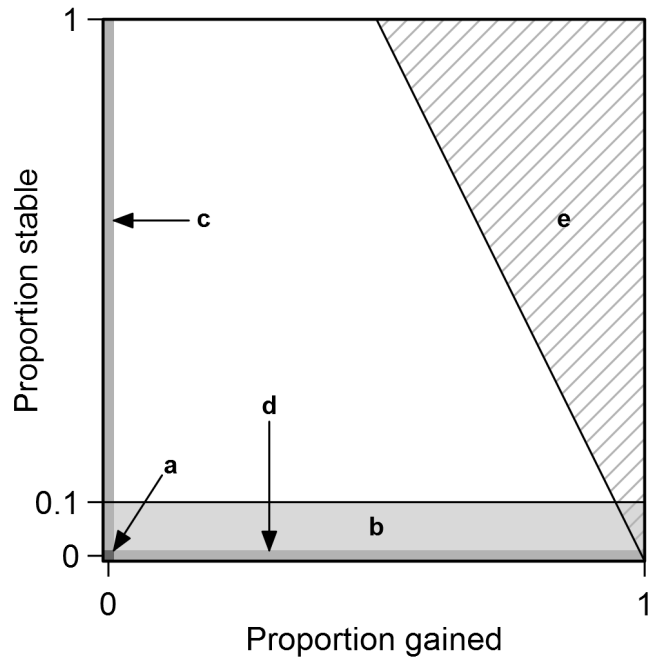
Figure 2.

Figure 3.
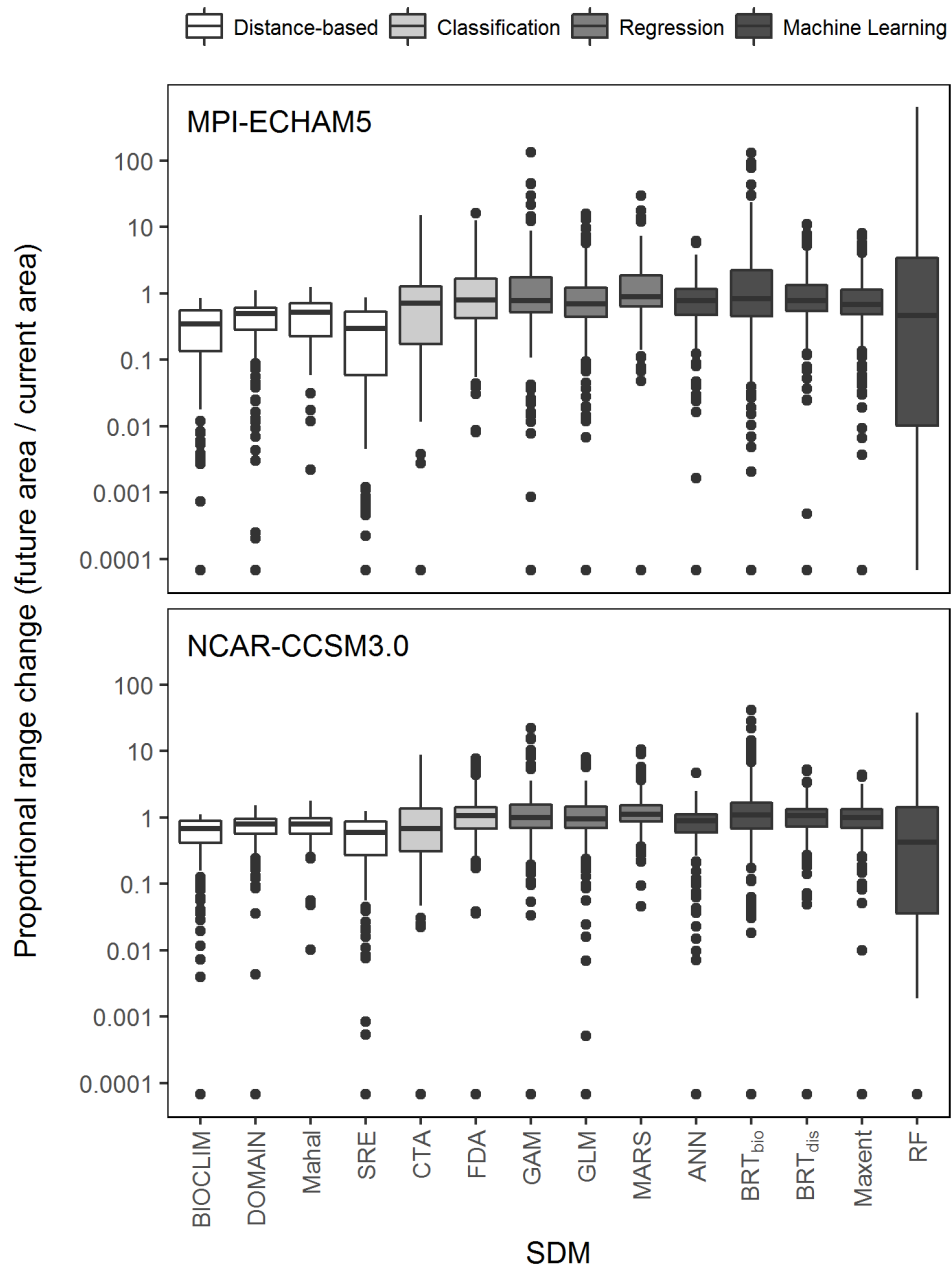
Figure 4.
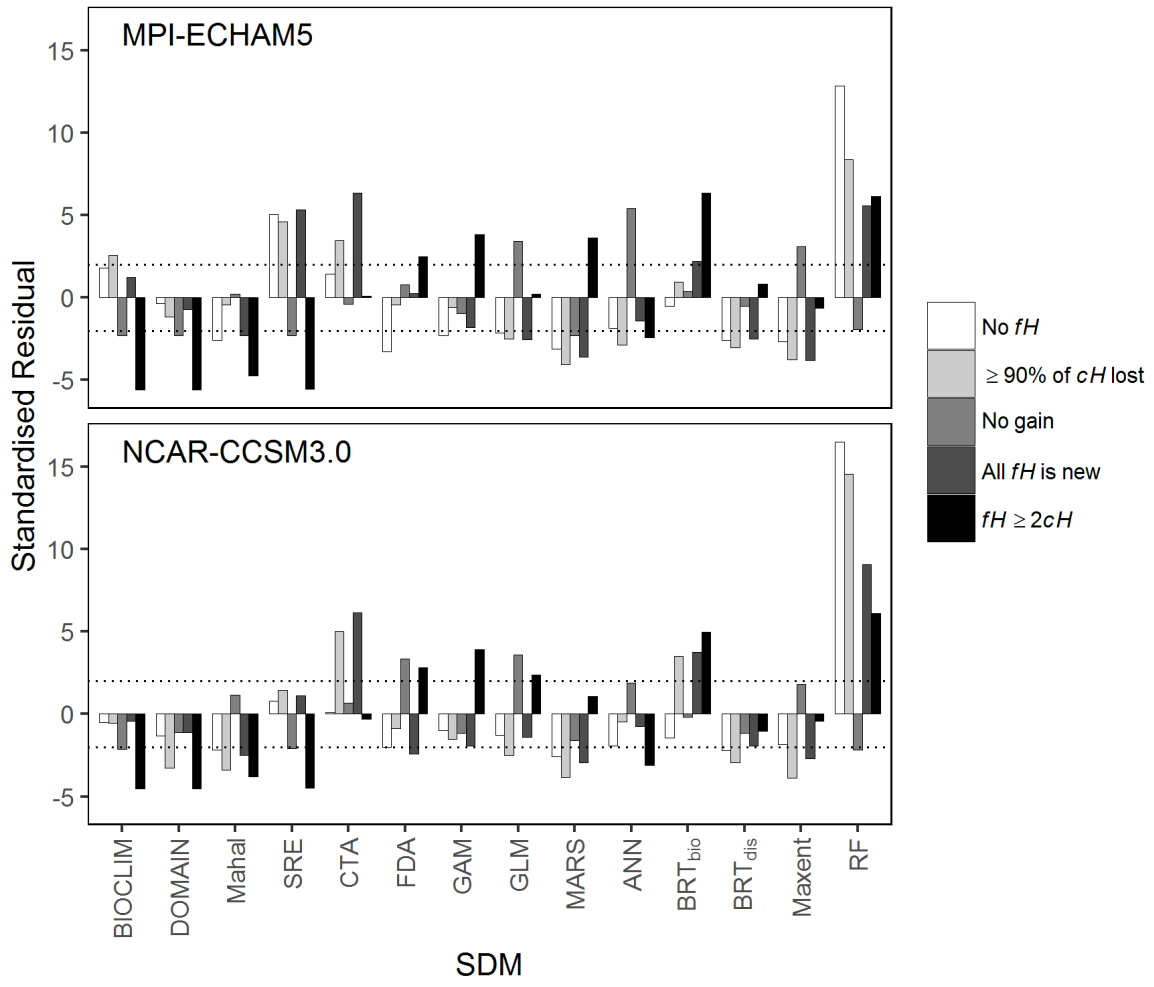
Which species distribution models are more (or less) likely to project broad-scale, climate-induced shifts in species ranges?

Linda J. Beaumont[1], Erin Graham[2], Daisy Englert Duursma[1], Peter D. Wilson[1], Abigail Cabrelli[1], John B. Baumgartner[1], Willow Hallgren[3], Manuel Esperón-Rodríguez[1], David A. Nipperess[1], Dan L. Warren[1], Shawn W. Laffan[4], Jeremy VanDerWal[2]

[1]Department of Biological Sciences, Macquarie University, NSW, 2109, Australia

[2]Centre for Tropical Biodiversity and Climate Change, School of Marine and Tropical Biology, James Cook University, Townsville, QLD, 4811, Australia

[3]Griffith University, Gold Coast Campus, Parklands Drive, Southport, QLD, 4215, Australia

[4]Centre for Ecosystem Science, School of Biological, Earth and Environmental Sciences, UNSW, Sydney, NSW, 2052, Australia

Supplementary Information

Table S1. 220 native Australian mammal species, their taxonomic order, and number of occurrence records used to model habitat suitability.

Order	Species	No. Records
Chiroptera		
	Chaerephon jobensis	211
	Chalinolobus dwyeri	308
	Chalinolobus gouldii	3,532

Chalinolobus morio	2,326
Chalinolobus nigrogriseus	373
Chalinolobus picatus	303
Falsistrellus tasmaniensis	642
Hipposideros ater	135
Hipposideros diadema	27
Hipposideros stenotis	35
Kerivoula papuensis	264
Macroderma gigas	167
Macroglossus minimus	68
Miniopterus australis	739
Miniopterus schreibersii	1,411
Mormopterus beccarii	125
Murina florium	11
Myotis macropus	596
Nyctimene robinsoni	102
Nyctophilus arnhemensis	130
Nyctophilus bifax	186
Nyctophilus geoffroyi	3,301
Nyctophilus gouldi	1,724
Nyctophilus timoriensis	281
Nyctophilus walkeri	85
Pipistrellus adamsi	86
Pipistrellus westralis	39
Pteropus alecto	431
Pteropus conspicillatus	89
Pteropus poliocephalus	999
Pteropus scapulatus	643
Rhinolophus megaphyllus	906
Rhinonicteris aurantia	105
Saccolaimus flaviventris	724
Scoteanax rueppellii	611
Scotorepens balstoni	849
Scotorepens greyii	829
Scotorepens orion	703
Scotorepens sanborni	43
Syconycteris australis	125
Tadarida australis	2,401
Taphozous australis	24
Taphozous georgianus	375
Taphozous hilli	98
Taphozous kapalgensis	18
Vespadelus baverstocki	302
Vespadelus caurinus	180
Vespadelus darlingtoni	1,221

Vespadelus douglasorum	34
Vespadelus finlaysoni	327
Vespadelus pumilus	827
Vespadelus regulus	1,386
Vespadelus troughtoni	129
Vespadelus vulturnus	2,422
Dasyuromorphia	
Antechinomys laniger	285
Antechinus adustus	21
Antechinus agilis	893
Antechinus bellus	114
Antechinus flavipes	1,301
Antechinus godmani	25
Antechinus leo	15
Antechinus minimus	175
Antechinus stuartii	1,296
Antechinus subtropicus	13
Antechinus swainsonii	736
Dasycercus cristicauda	62
Dasyuroides byrnei	20
Dasyurus geoffroii	140
Dasyurus hallucatus	446
Dasyurus maculatus	2,724
Dasyurus viverrinus	629
Ningaui ridei	373
Ningaui yvonneae	264
Phascogale calura	64
Phascogale pirata	34
Phascogale tapoatafa	650
Planigale gilesi	197
Planigale ingrami	221
Planigale maculata	366
Planigale tenuirostris	210
Pseudantechinus bilarni	73
Pseudantechinus macdonnellensis	201
Pseudantechinus ningbing	40
Sarcophilus harrisii	1,336
Sminthopsis aitkeni	10
Sminthopsis bindi	37
Sminthopsis butleri	21
Sminthopsis crassicaudata	1,628
Sminthopsis dolichura	350
Sminthopsis gilberti	85
Sminthopsis griseoventer	127
Sminthopsis hirtipes	125

Sminthopsis leucopus	174
Sminthopsis longicaudata	28
Sminthopsis macroura	1,234
Sminthopsis murina	823
Sminthopsis ooldea	364
Sminthopsis psammophila	46
Sminthopsis virginiae	168
Sminthopsis youngsoni	303
Diprotodontia	
Acrobates pygmaeus	1,041
Aepyprymnus rufescens	418
Bettongia gaimardi	421
Bettongia penicillata	65
Bettongia tropica	26
Burramys parvus	21
Cercartetus caudatus	21
Cercartetus lepidus	157
Cercartetus nanus	398
Dactylopsila trivirgata	88
Dendrolagus bennettianus	25
Dendrolagus lumholtzi	60
Gymnobelideus leadbeateri	28
Hemibelideus lemuroides	53
Hypsiprymnodon moschatus	71
Lasiorhinus latifrons	467
Macropus agilis	729
Macropus antilopinus	239
Macropus bernardus	56
Macropus dorsalis	84
Macropus eugenii	112
Macropus fuliginosus	3,412
Macropus giganteus	3,139
Macropus parma	218
Macropus parryi	182
Macropus robustus	3,117
Macropus rufogriseus	2,503
Macropus rufus	5,081
Onychogalea fraenata	18
Onychogalea lunata	27
Onychogalea unguifera	441
Petauroides volans	1,713
Petaurus australis	1,356
Petaurus breviceps	3,031
Petaurus gracilis	63
Petaurus norfolcensis	786

Petrogale assimilis	59
Petrogale brachyotis	230
Petrogale concinna	38
Petrogale godmani	28
Petrogale herberti	24
Petrogale inornata	38
Petrogale lateralis	308
Petrogale mareeba	33
Petrogale penicillata	367
Petrogale persephone	11
Petrogale purpureicollis	21
Petrogale rothschildi	34
Petrogale sharmani	11
Petrogale xanthopus	207
Petropseudes dahli	93
Phascolarctos cinereus	3,258
Potorous longipes	19
Potorous tridactylus	562
Pseudocheirus peregrinus	2,377
Pseudochirops archeri	79
Pseudochirulus cinereus	15
Setonix brachyurus	68
Spilocuscus maculatus	12
Thylogale billardierii	821
Thylogale stigmatica	242
Thylogale thetis	330
Trichosurus caninus	601
Trichosurus vulpecula	5,153
Vombatus ursinus	3,994
Wallabia bicolor	3,025
Monotremata	
Ornithorhynchus anatinus	2,389
Tachyglossus aculeatus	7,135
Notoryctemorphia	
Notoryctes typhlops	105
Peramelemorphia	
Isoodon macrourus	1,085
Isoodon obesulus	1,106
Macrotis lagotis	632
Perameles gunnii	498
Perameles nasuta	1,410
Rodentia	
Conilurus penicillatus	107
Hydromys chrysogaster	960
Leggadina forresti	392

Leggadina lakedownensis	157
Mastacomys fuscus	126
Melomys burtoni	592
Melomys capensis	21
Melomys cervinipes	431
Mesembriomys gouldii	215
Mesembriomys macrurus	12
Notomys alexis	1,020
Notomys amplus	14
Notomys aquilo	16
Notomys cervinus	64
Notomys fuscus	134
Notomys mitchellii	348
Pseudomys apodemoides	169
Pseudomys australis	91
Pseudomys bolami	262
Pseudomys calabyi	33
Pseudomys delicatulus	453
Pseudomys desertor	505
Pseudomys fumeus	54
Pseudomys gracilicaudatus	106
Pseudomys hermannsburgensis	1,858
Pseudomys higginsi	104
Pseudomys johnsoni	43
Pseudomys nanus	479
Pseudomys novaehollandiae	208
Pseudomys occidentalis	28
Pseudomys oralis	84
Pseudomys pilligaensis	46
Pseudomys shortridgei	43
Rattus colletti	152
Rattus fuscipes	2,657
Rattus leucopus	71
Rattus lutreolus	1,186
Rattus sordidus	57
Rattus tunneyi	518
Rattus villosissimus	308
Uromys caudimaculatus	131
Xeromys myoides	14
Zyzomys argurus	582
Zyzomys maini	35
Zyzomys pedunculatus	26
Zyzomys woodwardi	44

Table S2. Settings used in *biomod2*, *dismo* and Maxent.

Algorithm	Description of modelling options adopted in BIOMOD	Function arguments

Artificial neural networks (ANN)	5-fold cross-validation to find best size and decay parameters. Initial random weights = 0.1, maximum number of iterations = 200.	`NbCV = 5`
		`size = NULL`
		`decay = NULL`
		`rang = 0.1`
		`maxit = 200`

BIOCLIM	—	—

Boosted Regression Tree (*biomod2* implementation; BRT$_{bio}$)	Bernoulli error distribution, interaction depth of 7, learning rate of 0.001, bag.fraction of 0.5, 5-fold cross-validations to select optimal number of trees with a maximum of 2500 trees fitted to the data.	`distribution = 'bernoulli'`
		`n.trees = 2500`
		`interaction.depth = 7`
		`n.minobsinnode = 5`
		`shrinkage = 0.001`
		`bag.fraction = 0.5`
		`train.fraction = 1`
		`cv.folds = 3`
		`keep.data = FALSE`
		`verbose = FALSE`
		`perf.method = 'cv'`

Boosted Regression Tree (*dismo* implementation; BRT$_{dis}$)	Bernoulli error distribution, interaction depth of 1, learning rate of 0.01, bag fraction of 0.75, 10-fold cross-validations to select optimal number of trees with a maximum of 10,000 trees fitted to the data.	`offset = NULL`
		`fold.vector = NULL`
		`tree.complexity = 1`
		`learning.rate = 0.01`
		`bag.fraction = 0.75`
		`site.weights = rep(1, nrow(data))`
		`var.monotone = rep(0, length(BRT.x))`
		`n.folds = 10`
		`prev.stratify = TRUE`
		`family = "bernoulli"`
		`n.trees = 50`
		`step.size = n.trees`
		`max.trees = 10000`

| | | ```
tolerance.method = "auto"
tolerance = 0.00
keep.data = FALSE
plot.main = TRUE
plot.folds = FALSE
verbose = TRUE
silent = FALSE
keep.fold.models = FALSE
keep.fold.vector = FALSE
keep.fold.fit = FALSE
``` |
| Classification tree analysis (CTA) | 5-fold cross-validation. | ```
method = 'class'
parms = 'default'
cost = NULL
control = list(xval = 5,
minbucket = 5, minsplit = 5, cp =
0.001, maxdepth = 25)
``` |
| DOMAIN | — | — |
| Flexible discriminant analysis (FDA) | Linear regression method used for optimal scaling, via function 'polyreg'. | ```
method = 'mars'
``` |
| Generalised Additive Model (GAM) | Algorithm = 'GAM_mgvc' with cubic spline smoother, with 4 degrees of freedom. Logistic link function and a binomial error distribution of the response. Note that GVC (Generalised Cross Validation) will automatically choose the number of knots to balance simplicity against explanatory power. | ```
algo = 'GAM_mgcv'
type = 's_smoother'
k = -1
interaction.level = 0
myFormula = NULL
family = binomial(link = 'logit')
method = 'GCV.Cp'
optimizer = c('outer','newton')
select = FALSE
knots = NULL
paraPen = NULL
control = list(nthreads = 1,
irls.reg = 0, epsilon = 1e-07,
maxit = 200, trace = FALSE,
mgcv.tol = 1e-07, mgcv.half = 15,
rank.tol = 1.49011611938477e-08,
nlm = list(ndigit=7, gradtol=1e-
06, stepmax=2, steptol=1e-04,
iterlim=200,
check.analyticals=0), optim =
list(factr=1e+07), newton =
list(conv.tol=1e-06, maxNstep=5,
maxSstep=2, maxHalf=30,
``` |

		`use.svd=0), outerPIsteps = 0,`
		`idLinksBases = TRUE, scalePenalty`
		`= TRUE, keepData = FALSE,`
		`scale.est = fletcher)`

Generalised Linear Model (GLM)	Logistic link function and a binomial error distribution of the response. Linear and quadratic terms with no interaction level. Stepwise procedure using the AIC criterion to select the most significant terms.	`type = 'quadratic'`
		`interaction.level = 0`
		`myFormula = NULL`
		`test = 'AIC'`
		`family = binomial(link = 'logit')`
		`mustart = 0.5`
		`control = glm.control(epsilon =`
		`1e-08, maxit = 50, trace = FALSE)`

Mahalanobis (Mahal)	—	—

Multivariate Adaptive Regression Spline (MARS)	Two-way interactions, cost per degree of freedom change = 2, forward stepwise stopping threshold = 0.001.	`degree = 2`
		`nk = NULL`
		`penalty = 2`
		`thresh = 0.001`
		`prune = TRUE`

Maxent	Linear, Product, Quadratic and Hinge features used; 10-fold cross-validations.	`cat('java -mx2048m -jar`
		`',maxent.jar, ' -e ',bkgd.data, '`
		`-s ',occur.data, ' -o`
		`',maxent.dir, ' nothreshold`
		`nowarnings novisible`
		`nowriteclampgrid nowritemess`
		`writeplotdata -P -J -r -a \n',`
		`sep="", file=shell.file))`

Random Forest (RF)	Classification RF, number of predictors to be chosen randomly at each node = $\sqrt{p}$ where p is number of variables. Number of trees grown = 500.	`do.classif = TRUE`
		`ntree = 500`
		`mtry = 'default'`
		`nodesize = 5`
		`maxnodes = NULL`

Surface Range Envelope (SRE)	Quantile of extreme environmental variable to be removed from envelope.	`quant = 0.025`

Table S3. Habitat suitability under current and future climates (ECHAM5 and CCSM3) was modelled for 220 native Australian mammal species, using 14 SDM methods. Number of species for which SDM predictive power was equal to or greater than TSS 0.4 is given. Expected frequency is the proportion of the total number of models (2,694) that a given SDM method accounts for. Also given is the number of predictions in each of the five categories of extreme range changes that we assessed, i.e. a) no *fH*, b) low stability of *cH*, c) no gain of new *fH*, d) all *fH* is gain and e) *fH* is at least double the size of *cH*. (*cH* = current habitat; *fH* = future habitat).

SDM	No. of species	Expected frequency	MPI-ECHAM5					NCAR-CCSM3.0				
			a	b	c	d	e	a	b	c	d	e
ANN	178	0.066	4	21	15	10	13	1	19	7	5	4
BIOCLIM	209	0.078	17	61	0	23	0	5	22	0	7	0
BRT$_{bio}$	219	0.081	10	53	6	28	64	3	43	4	19	41
BRT$_{dis}$	214	0.079	3	26	4	8	34	1	11	2	3	15
CTA	63	0.023	6	26	1	20	9	2	21	2	12	5
Domain	208	0.077	10	37	0	15	0	3	9	2	5	0
FDA	186	0.069	0	37	6	17	38	1	18	10	1	28
GAM	215	0.080	4	42	3	11	50	4	18	2	3	36
GLM	201	0.075	4	27	12	7	29	3	12	11	4	28
Mahal	152	0.056	1	30	4	5	0	0	4	5	0	0
MARS	205	0.076	1	18	0	3	47	0	6	1	0	23
Maxent	219	0.081	3	22	12	3	27	2	7	8	1	18
RF	220	0.082	54	102	1	42	63	47	97	0	34	46
SRE	205	0.076	27	73	0	39	0	8	31	0	11	0
Sum	**2,694**		144	575	64	231	374	80	318	54	105	244

	No. predictions in each category of extreme range change	
	MPI-ECHAM5	NCAR-CCSM3.0

SDM	No. of species	Expected frequency	a	b	c	d	e	a	b	c
ANN	178	0.066	4 (9.5)	21 (38)	15 (4.2)	10 (15.3)	13 (24.7)	1 (5.3)	19 (21)	7 (3.6)
BIOCLIM	209	0.078	17 (11.2)	61 (44.6)	0 (5)	23 (17.9)	0 (29)	5 (6.2)	22 (24.7)	0 (4.2)
BRT$_{bio}$	219	0.081	10 (11.7)	53 (46.7)	6 (5.2)	28 (18.8)	64 (30.4)	3 (6.5)	43 (25.9)	4 (4.4)
BRT$_{dis}$	214	0.079	3 (11.4)	26 (45.7)	4 (5.1)	8 (18.3)	34 (29.7)	1 (6.4)	11 (25.3)	2 (4.3)
CTA	63	0.023	6 (3.4)	26 (13.4)	1 (1.5)	20 (5.4)	9 (8.7)	2 (1.9)	21 (7.4)	2 (1.3)
Domain	208	0.077	10 (11.1)	37 (44.4)	0 (4.9)	15 (17.8)	0 (28.9)	3 (6.2)	9 (24.6)	2 (4.2)
FDA	186	0.069	0 (9.9)	37 (39.7)	6 (4.4)	17 (15.9)	38 (25.8)	1 (5.5)	18 (22)	10 (3.7)
GAM	215	0.080	4 (11.5)	42 (45.9)	3 (5.1)	11 (18.4)	50 (29.8)	4 (6.4)	18 (25.4)	2 (4.3)
GLM	201	0.075	4 (10.7)	27 (42.9)	12 (4.8)	7 (17.2)	29 (27.9)	3 (6)	12 (23.7)	11 (4)
Mahal	152	0.056	1 (8.1)	30 (32.4)	4 (3.6)	5 (13)	0 (21.1)	0 (4.5)	4 (17.9)	5 (3)
MARS	205	0.076	1 (11)	18 (43.8)	0 (4.9)	3 (17.6)	47 (28.5)	0 (6.1)	6 (24.2)	1 (4.1)
Maxent	219	0.081	3 (11.7)	22 (46.7)	12 (5.2)	3 (18.8)	27 (30.4)	2 (6.5)	7 (25.9)	8 (4.4)
RF	220	0.082	54 (11.8)	102 (47)	1 (5.2)	42 (18.9)	63 (30.5)	47 (6.5)	97 (26)	0 (4.4)
SRE	205	0.076	27 (11)	73 (43.8)	0 (4.9)	39 (17.6)	0 (28.5)	8 (6.1)	31 (24.2)	0 (4.1)
Sum	2,694		144	575	64	231	374	80	318	54

Figure S1. Box-and-whisker plot of AUC and size of current habitat projected for 220 Australian mammal species belonging to seven taxonomic orders. Projections were made using 14 SDM methods. The dotted line indicates the TSS threshold (0.4) below which models were excluded due to poor predictive performance.

Figure S2.