

RESEARCH
PAPER



Reducing uncertainty in projections of extinction risk from climate change

Miguel B. Araújo^{1,2*}, Robert J. Whittaker¹, Richard J. Ladle¹ and Markus Erhard³

¹Biodiversity Research Group, School of Geography and the Environment, University of Oxford, Mansfield Road, Oxford, OX1 3TB, UK, ²Biogeography and Conservation Laboratory, Natural History Museum, Cromwell Road, London, SW7 5BD, UK, ³Institute for Meteorology and Climate Research, Forschungszentrum Karlsruhe, Kreuzackbahnstr, 19, 82467 Garmisch-Partenkirchen, Germany

ABSTRACT

Aim Concern over the implications of climate change for biodiversity has led to the use of species–climate ‘envelope’ models to forecast risks of species extinctions under climate change scenarios. Recent studies have demonstrated significant variability in model projections and there remains a need to test the accuracy of models and to reduce uncertainties. Testing of models has been limited by a lack of data against which projections of future ranges can be tested. Here we provide a first test of the predictive accuracy of such models using observed species’ range shifts and climate change in two periods of the recent past.

Location Britain.

Methods Observed range shifts for 116 breeding bird species in Britain between 1967 and 1972 (t_1) and 1987–91 (t_2) are used. We project range shifts between t_1 and t_2 for each species based on observed climate using 16 alternative models (4 methods \times 2 data parameterizations \times 2 rules to transform probabilities of occurrence into presence and absence records).

Results Modelling results were extremely variable, with projected range shifts varying both in magnitude and in direction from observed changes and from each other. However, using approaches that explore the central tendency (consensus) of model projections, we were able to improve agreement between projected and observed shifts significantly.

Conclusions Our results provide the first empirical evidence of the value of species–climate ‘envelope’ models under climate change and demonstrate reduction in uncertainty and improvement in accuracy through selection of the most consensual projections.

Keywords

Bioclimatic envelope modelling, British birds, climate change, consensus forecasting, model variability, probabilistic modelling, species distributions, uncertainty.

*Correspondence: Miguel B. Araújo, Biodiversity Research Group, School of Geography and Environment, University of Oxford, Mansfield Road, Oxford, OX1 3TB, UK. E-mail: maraujo@ouce.ox.ac.uk

INTRODUCTION

The species–climate ‘envelope’ modelling approach (also known as ecological niche modelling) uses present climate–range relationships to characterize species’ limits of tolerance to climate conditions, and apply climate change scenarios to enable projections of altered species distributions. Alongside the increasing application of this approach (e.g. Austin, 1992; Huntley *et al.*, 1995; Sykes *et al.*, 1996; Iverson & Prasad, 1998; Bakkenes *et al.*, 2002; Berry *et al.*, 2002; Erasmus *et al.*, 2002; Peterson *et al.*, 2002; Thuiller, 2003; Araújo *et al.*, 2004; Skov & Svenning, 2004; Thomas *et al.*, 2004), criticisms have been raised as to the useful-

ness of such models for guiding policy making and planning on the grounds that they make unrealistic assumptions of species distributions being at equilibrium with current climate conditions, interpret species–climate correlations as if indicating causal mechanisms, and ignore important parameters such as dispersal and biotic interactions (for discussion see Pearson & Dawson, 2003; Hampe, 2004). It has recently been shown that projections may be sensitive not only to the underlying assumptions of models and climate scenarios adopted but also to variations in methods used for calibrating models (e.g. Thuiller, 2004; Thuiller *et al.*, 2004; Pearson *et al.*, unpublished data). One approach to dealing with this variability is to fit a number of

alternative models and to explore the resulting range of projections. Unfortunately, as these models relate to future distributions, we lack an objective basis for selecting the 'best', i.e. most realistic modelling techniques (Araújo *et al.*, 2005). Herein, we make the starting assumptions that all projections are equally likely and that collectively they delimit the range of uncertainties associated with forecasting the future range of a given species. A 'majority-vote' criterion can then be used to assign higher probabilities to the most consensual projections (Clemen, 1989). This idea is based on the central limit theorem in statistics, where central limits of particular projection are expected to be more likely than the extremes because of the concatenation of a number of input factors. If all major sources of error have been taken into account, it is further assumed that the remaining error must be the result of a large number of small additive effects, hence producing a normal frequency.

Testing the ability of a model to predict climate-induced range shifts requires use of standardized distribution data for species in at least two different periods undergoing climate change. Such observations are rarely available, but existing data on the distribution of British breeding-bird species (Sharrock, 1976; Gibbons *et al.*, 1993) between 1968 and 1972 (t_1) and 1995–99 (t_2) meet the requirements for the test. Furthermore, they include species whose ranges have shifted northwards in apparent response to recent climate changes (Thomas & Lennon, 1999). We modelled the association between six climate variables and 116 species' observed distributions for t_1 and used these models to project potential distributions for t_2 , for comparison with observed data. In order to measure the spread of model outputs, an ensemble of 16 projections for every species was obtained by varying modelling techniques, data for calibration of models and rules for transforming model outputs into presence/absence records (i.e. four different models, each with two calibrations and two rules for transforming model outputs, as described in the Methods section). Distributional changes were measured for every species as the difference in total number of 10-km² grid cells occupied (or predicted to be occupied in the case of modelled distributions) in each time period. In order to identify patterns of central tendency (i.e. consensus) in projections, we explored two alternative strategies (herein termed consensus models and consensus forecasts). To test the relative performance of all models and of the consensus strategies, we examined their predictive performance directly by comparison with observed range shifts.

Table 1 Frequency distribution of the climate variables used in this study

	1967–72			1987–1991		
	Min	Mean	Max	Min	Mean	Max
Growing degree days	528.1	1462.4	2144.6	534.1	1549.1	2336.1
Annual precipitation	522.5	1096.9	3096.6	503.7	1153.1	3609.4
Summer precipitation	125.1	245.7	737.7	108.3	256.8	864.6
Annual temperature	3.9	8.5	10.8	4.2	8.8	11.4
Minimum temperature of the coldest month	–2.4	2.4	5.8	–1.6	2.7	6.0
Maximum temperature of the warmest month	10.5	14.9	17.7	10.8	15.4	18.2

METHODS

Climate data

Mean values of six predictor climate variables were derived from an updated version of the Climate Research Unit (CRU) monthly climate data (New *et al.*, 2000) for the two recording periods for birds plus an additional year before the start of the surveys (1967–72, 1987–91). The inclusion of this additional year was needed because distributions of birds in a particular breeding season are not independent of the climate conditions observed in the previous year. Climate variables included: annual temperature (°C), temperature of the coldest month (°C), temperature of the warmest month (°C), annual precipitation (mm), summer precipitation (mm) and growing season (°C), defined as the temperature sum of all consecutive days with mean temperature > 5 °C. Choice of variables was made to reflect primary qualities of the climate — energy and water — that, on the basis of prior knowledge, have known roles in imposing constraints upon species distributions as a result of widely shared physiological limitations (e.g. Lennon *et al.*, 2000; Crick, 2004). During the study period there was a relatively consistent trend of increasing temperatures in Great Britain (Table 1). Precipitation trends were more variable with increases recorded near the mean and the upper tail of the frequency distribution, and decreases near the lower tail of this distribution (Table 1).

Species data

British birds are one of the most intensively recorded faunas in the world (Sharrock, 1976; Gibbons *et al.*, 1993), with 98% to 100% sampling coverage of all 2861 of Britain's mainland 10-km² grid cells being available for two different periods of time: 1968–72 (t_1) and 1995–99 (t_2). Species with fewer than 20 records in the calibration set were not modelled as they were considered to have an insufficient number of observations for modelling. We modelled distributions of 116 British breeding-bird species after excluding marine, waterfowl and wader bird species. The list of species modelled is provided in Appendix 1.

Species–climate envelope modelling

Four familiar modelling techniques were used to model species distributions with the BIOMOD (Thuiller, 2003) application in

s-PLUS: (1) generalized linear models (GLM) with linear, quadratic and polynomial terms (second and third orders) with the possibility of interactions only for linear terms; (2) generalized additive models (GAM) with cubic-smooth splines, whereby the degree of smoothness was automatically selected by cross-validation and bounded to 4 for each variable; (3) classification tree analysis (CTA) using a 10-fold cross-validation to select the best trade-off between the number of leaves of the tree and the explained deviance; and (4) feed-forward artificial neural networks (ANN) with seven hidden units in a single layer with weight decay equal to 0.03; because of the heuristic nature of ANN, models were run 10 times and the mean projection was used (Ripley, 1996). In order to quantify the variability associated with using response data of different qualities, two calibrations were made for each model. In the first, models were calibrated using a random sample of 70% of the original t_1 data (e.g. GLM₇₀). The size of the calibration set was determined by application of a commonly used heuristic for identifying the ratio of training and cross-validation sets in presence and absence models: $[1 + (p - 1)^{1/2}]^{-1}$, where p is the number of predictors (Fielding & Bell, 1997). The second set of models was calibrated using 100% of t_1 data (e.g. GLM₁₀₀). To assess species expansions and contractions we transformed probabilities of occurrence from models into presence and absence using both the area under the Receiver Operator Characteristic (ROC) curve and maximum Cohen's Kappa (for more details see Fielding & Bell, 1997; Thuiller, 2003).

Consensus forecasting

Forecasts are said to be in consensus if they are insignificantly different from the sample mean (Gregory *et al.*, 2001). To find consensus in model projections, we explored two alternative strategies. In the first, we selected individual models best summarizing the overall pattern of variation in projections (herein termed *consensus models*). Consensus models were selected using principal components analysis (PCA), which identifies orthogonal groups of linearly covarying projections (e.g. Sengupta & Boyle, 1998; Thuiller, 2004). There are many variants of PCA that can be used to explore consensus among projections, but standard PCA has often been recommended (e.g. Westerhuis *et al.*, 1998). PCA was performed using projected range shifts for every species. The first principal component (PC1) is equivalent to a line that goes through the centroid of all sets of model projections and minimizes the square of the distance of each set of projections to that line. PC1 is as close to all of the data as possible and this is why it is called the consensus axis (see also Thuiller, 2004); it represents the greatest proportion of variance among projections. Subsequent axes also seek to represent as much variation as possible but they are orthogonal (unrelated) to the previous principal components. Component loadings in PCA (i.e. weights given to individual model projections within each component) represent the relative importance of each projection within components. Inspection of the component loadings allowed us to identify the group of projections that was more related to the first axis of PCA. In practice, this amounted to

selecting the highest loading model and then the closest models to it in both strength and direction. In the second strategy, we derived a single projection representing the central tendency across selected models (herein termed the *consensus forecast*). We obtained two consensus forecasts: the first represented the median projected range shift for each species across all 16 model projections; the second represented the median projected range shift across consensus models alone.

RESULTS

Model-based uncertainties

Taking all models into account, the model projections were highly variable both in magnitude and direction (Fig. 1a), with 90% of the species projected both to expand and to contract depending on the modelling technique and calibration used. For example, the common crossbill, *Loxia curvirostra*, was projected to lose 325 suitable grid cells according to the GLM₁₀₀.ROC model and to gain 1522 new suitable grid cells with ANN₁₀₀.ROC. The actual shift was expansion by 356 grid cells. Amongst the 10% of species for which models always yielded consistent projections, 50% species had projections that were contrary to the direction of their observed shift. This was the case for the red-backed shrike (*Lanius collurio*), projected to expand its range in Britain by all 16 models but contracting from presence in 111 cells in t_1 to presence in 15 cells in t_2 . Similar findings pertain for the little owl (*Athene noctua*) and the jay (*Garrulus glandarius*), which each had contracting distributions in Britain but were projected to expand by all models. Among the 74 British-bird species that contracted during the reported period, species were correctly projected to contract in [median] 50% of the models (1st quartile = 31%; 3rd quartile = 69%). For the remaining 42 species that expanded during the reported period, species were correctly projected to expand in [median] 56% of the models (1st quartile = 44%; 3rd quartile = 88%) (Table 2). In short, assessment of all projections indicates a performance in terms of predicting the direction of range shifts no better than tossing a coin.

Reducing uncertainty by consensus

The first principal component (PC1 or consensus axis) explained 29.9% of model variation. Consensus models were discriminated by the first axis of the PCA and included ANN₁₀₀.Kap, GAM₁₀₀.Kap, GLM₁₀₀.Kap and CTA₁₀₀.Kap model projections (Appendix 2). Projections of expansion and contraction from these four consensus models coincided closely with observed range shifts (Fig. 2) and projected shifts were significantly different between groups of species that contracted and species that expanded ($\chi^2 = 39.93, 34.06, 34.63, 28.48$, respectively; Kruskal–Wallis test, $P < 0.0001$, Appendix 3). Outputs from the remaining (i.e. non-consensus) models diverged visibly from observed patterns (Fig. 2) and for these models range-shift projections were not significantly different between groups of species that contracted and expanded ($P > 0.05$ for most non-consensus

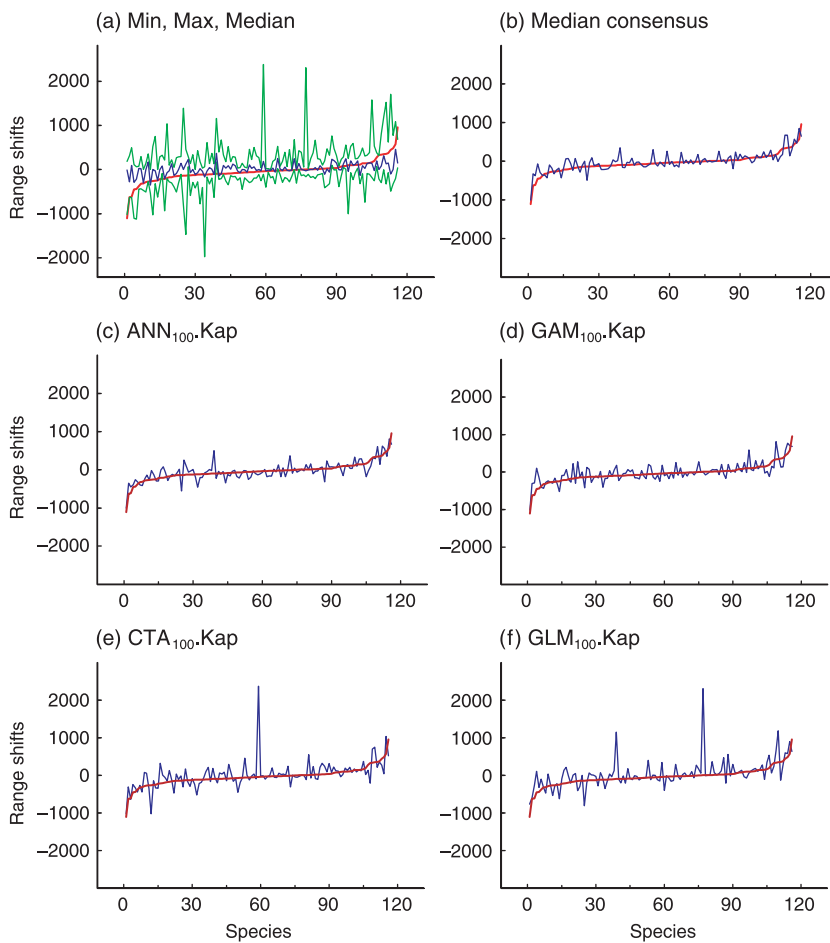


Figure 1 Projected and observed range shifts among the 116 British breeding-bird species. Shifts occurred in the *c.* 20-year time period between 1968 and 1972 and 1995–99: (a) observed (red line), median (blue) and projected maximum and minimum (green line) shifts from models; (b) observed and projected median shifts from the four consensus models; (c) observed and projected shifts using ANN₁₀₀.Kap (artificial neural networks, calibrated with all data and using the Kappa statistic to transform probabilities of occurrence into presence and absence); (d) GAM₁₀₀.Kap (generalized additive models with all data and Kappa); (e) CTA₁₀₀.Kap (classification tree analysis with all data and Kappa); and (f) GLM₁₀₀.Kap (generalized linear models with all data and Kappa). Species are ordered on the *x*-axis by their direction and magnitude of range shift and the units on the *y*-axis are the number of grid cells. Peaks of projections above the red line (observed shifts) indicate false positive errors whilst peaks below the red line indicate false negative errors (see Table 2).

models and $P > 0.01$ for ANN₁₀₀.ROC, GLM₇₀.Kap, GLM₁₀₀.ROC, Appendix 2). Consensus forecasts based on averaging projections from all models improved accuracy over each one of the 12 nonconsensus models (Appendix 3). However, they performed less well than each one of the four individual consensus models, especially for species showing the greatest expansions and contractions, where consensus forecasts consistently underestimated range shifts (Fig. 1, Appendix 3). By contrast, when the averaging of projections was performed based only on the four consensus models, the derived consensus forecast was more strongly related to observed range shifts than any of the individual consensus-model projections (chi-square = 45.99, Kruskal–Wallis test, $P < 0.0001$, Appendix 3). Indeed, the proportion of species that were correctly projected to contract and expand with this refined consensus forecast was [median] 100% (1st quartile = 100%; 3rd quartile = 100%). When using projections from consensus models, individually, the proportion of species that were correctly projected to contract was [median] 75% (1st quartile = 50%; 3rd quartile = 100%), whereas the correctly projected expansions were [median] 88% (1st quartile = 50%; 3rd quartile = 100%) (Table 1).

To investigate how non-consensus models contributed to uncertainty in assessments of range shift, we grouped non-consensus model projections into three groups with four model projections for each species in each of the groups (i.e. ₁₀₀ROC,

₇₀Kappa and ₇₀ROC). The groups were defined so as to reflect variations in the size of the calibration (i.e. 70% or 100% of data) set and the rules used to transform probabilities of occurrence into species presence and absence (i.e. Kappa and ROC). Results show that models calibrated with only 70% of the data for t_1 markedly underperform, and that when using 100% of the t_1 data, the four models using the Kappa statistic to transform probabilities of occurrence into presence and absence collectively outperform the ROC procedure (Table 2).

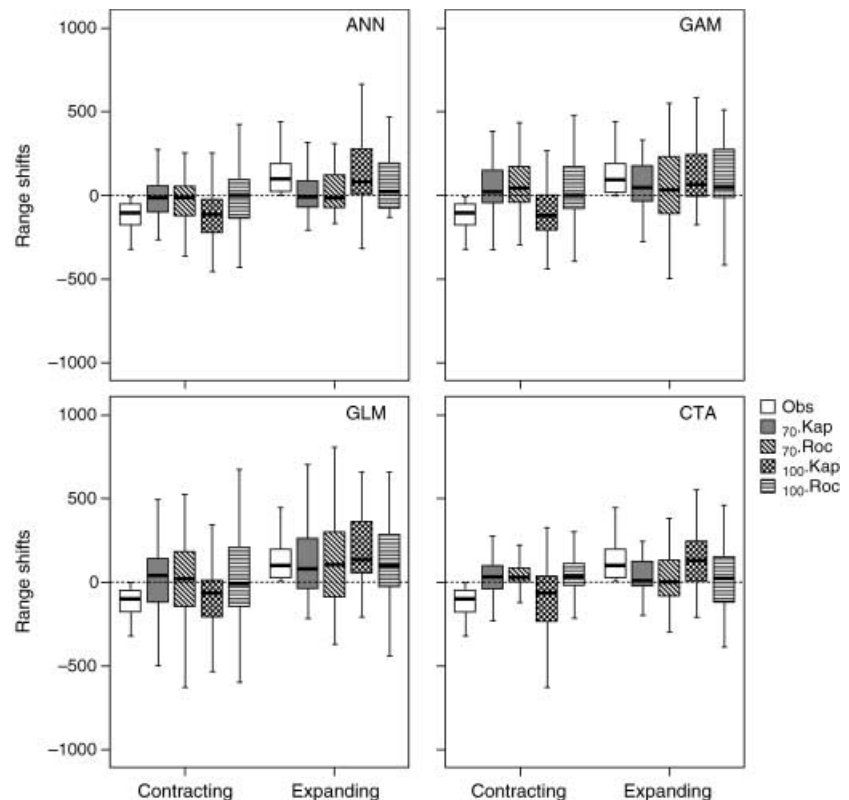
DISCUSSION

Our finding that model outputs are so variable corroborates claims that variability from species–climate envelope models can be so large as to compromise their usefulness in forecasting species extinctions risk under climate change (e.g. Thuiller *et al.*, 2004; Pearson *et al.*, unpublished data). However, by selecting models best representing a consensus of model outcomes, or by deriving a single projection reflecting the central tendency of selected forecasts, we were able not only to reduce the spread of projected species' responses to climate change but, more importantly, increase the level of agreement between projected and observed range shifts. The usefulness of consensus (or 'ensemble') forecasting has been illustrated in other fields of research including economics (e.g. Gregory *et al.*, 2001), management (e.g. Makridakis

Table 2 Model range-shift uncertainties among 116 British bird species. Percentage values were calculated in two steps. First, we calculated the percentage of times model outputs correctly predicted the direction of observed range shifts for each species. Then the frequency distribution values of percentage values across all species were calculated: median, lower-quartile and upper-quartile values of this frequency distribution values are presented in the table. Results were aggregated into seven sets: (1) 'all' includes results from all 16 model projections per species; (2) consensus models, include results from ANN₁₀₀.Kap, GAM₁₀₀.Kap, GLM₁₀₀.Kap and CTA₁₀₀.Kap; (3) Consensus forecast 1, includes averaging results from all 16 model projections; (4) Consensus forecast 2, includes averaging results from the four consensus models; (5) ₁₀₀ROC includes results from all four modelling techniques that were calibrated with all data and used the ROC procedure to transform probabilities of occurrence into presence and absence; (6) ₇₀Kap includes results from all four modelling techniques that were calibrated with 70% random sample of the data and used the Kappa statistic to transform probabilities of occurrence into presence and absence and; (7) ₇₀ROC includes results from all four modelling techniques that were calibrated with 70% random sample of the data and used the ROC procedure to transform probabilities of occurrence into presence and absence

Observed range change	Model prediction	
	Contracted	Did not contract
Species contracted	<p>True positives</p> <p>All models: 50% (LQ = 31%; UQ = 69%)</p> <p>Cons. models: 75% (LQ = 50%; UQ = 100%)</p> <p>Cons. forecast 1: 100% (LQ = 0%; UQ = 100%)</p> <p>Cons. forecast 2: 100% (LQ = 100%; UQ = 100%)</p> <p>₁₀₀ROC models: 50% (LQ = 25%; UQ = 75%)</p> <p>₇₀Kap models: 25% (LQ = 0%; UQ = 75%)</p> <p>₇₀ROC models: 25% (LQ = 25%; UQ = 50%)</p>	<p>False negatives</p> <p>All models: 50% (LQ = 31%; UQ = 69%)</p> <p>Cons. models: 25% (LQ = 0%; UQ = 50%)</p> <p>Cons. forecast 1: 0% (LQ = 0%; UQ = 100%)</p> <p>Cons. forecast 2: 0% (LQ = 0%; UQ = 0%)</p> <p>₁₀₀ROC models: 50% (LQ = 25%; UQ = 75%)</p> <p>₇₀Kap models: 75% (LQ = 25%; UQ = 100%)</p> <p>₇₀ROC models: 75% (LQ = 50%; UQ = 75%)</p>
Species did not contract	<p>False positives</p> <p>All models: 44% (LQ = 12%; UQ = 56%)</p> <p>Cons. models: 12% (LQ = 0%; UQ = 50%)</p> <p>Cons. forecast 1: 0% (LQ = 0%; UQ = 100%)</p> <p>Cons. forecast 2: 0% (LQ = 0%; UQ = 0%)</p> <p>₁₀₀ROC models: 25% (LQ = 0%; UQ = 50%)</p> <p>₇₀Kap models: 50% (LQ = 6%; UQ = 75%)</p> <p>₇₀ROC models: 50% (LQ = 0%; UQ = 75%)</p>	<p>True negatives</p> <p>All models: 56% (LQ = 44%; UQ = 88%)</p> <p>Cons. models: 88% (LQ = 50%; UQ = 100%)</p> <p>Cons. forecast 1: 100% (LQ = 0%; UQ = 100%)</p> <p>Cons. forecast 2: 100% (LQ = 100%; UQ = 100%)</p> <p>₁₀₀ROC models: 75% (LQ = 50%; UQ = 100%)</p> <p>₇₀Kap models: 50% (LQ = 25%; UQ = 94%)</p> <p>₇₀ROC models: 50% (LQ = 25%; UQ = 100%)</p>

Figure 2 Frequency distribution of projected and observed range shifts among the 116 British breeding-bird species modelled in relation to observed (1) contractions and (2) expansions. There are four graphs representing results for each set of modelling techniques. Within these graphs, each box represents observed and projected range shifts varying the latter in relation to the amount of data used for calibration (70% vs. 100%) and the method used to transform probabilities of occurrence into presence and absence records (ROC or Kappa). Within each panel the fourth and ninth boxes represent the consensus set of models. The shaded bars represent the interquartile range and the median is marked within this; the line extensions from each box are the largest and smallest values, excluding outliers (points more than 1.5 times the interquartile range distant from the box edges). The asymmetrical position of the median within interquartile ranges indicates that distributions are skewed. A test of difference between frequency distributions of range shifts for contracting and expanding species is given in Appendix 3.



& Winkler, 1983), biomedicine (e.g. Nilsson *et al.*, 2000), meteorology (e.g. Sanders, 1963) and climatology (e.g. Benestad, 2004), and is embedded in heuristic modelling that uses multiple projections to derive a single combination (e.g. GARP, Stockwell & Peters, 1999). The rationale for combining projections stems from the demonstration that, when estimated variability spans the full range of predictive uncertainties, no more than half of individual projections that define the median projection of multiple runs can ever be more accurate than the median projection (McNees, 1992). A problem with translating intermodel spread into objective assessments of uncertainty is that there is no guarantee that estimated variability spans the full range of modelling uncertainties (Allen *et al.*, 2000). Indeed, any combination of projections is a synthetic construct with no existence independent of the individual projections of which it is composed. Hence, although considerable effort has been expended on finding the best way to combine individual projections (Palm & Zellner, 1992), improved accuracy will still depend on traditional tasks of trying to build better models with improved data. As put forward by Stephen K. McNees (1987), 'there would be no advantage to combining my forecast with yours if we agreed precisely on the forecasting techniques, if I simply mimicked your forecast, or if my forecast were always inferior' (p.15). These ideas are supported by our study. For example, the average of 16 models was better than most individual model projections. While the result from averaging 16 models was not as good as any of the four consensus models, the average of the four consensus models gave the best predictive performance. This provides evidence in support of the intuitive expectation that the increased accuracy expected from averaging forecasts should improve mainly when better and not more models are taken into account.

In this study, two major sources of error were identified. First, incompleteness of data resulting from calibrating models with a random sample of 70% of the original data was found to dramatically reduce the performance of models. This may seem a trivial finding. However, it has great relevance to real-world planning as first, in the absence of future data, reserving a portion of the data from t_1 for model evaluation is a standard protocol (Fielding & Bell, 1997), and second, most species occur in regions with poor data and our ability to predict their fate is therefore more limited than previously anticipated (e.g. Erasmus *et al.*, 2002; Midgley *et al.*, 2002; Peterson *et al.*, 2002; Thomas *et al.*, 2004). Secondly, we also found that rules to transform probabilities of occurrence into species presence/absence have an important effect as a source of model variability. This suggests that more robust forecasts might be obtained in the future by using estimates of change in the probabilities of occurrence rather than changes in the presences and absences of species (e.g. Araújo *et al.*, 2002).

Although our particular implementation of consensus modelling provides the first evidence of agreement between observed and projected responses of species to climate change, one has to be aware of the limitations of species–climate 'envelope' models before supporting their unsupervised use in policy making. First, it is important to acknowledge that species–climate models for future climate scenarios can never be truly validated (e.g. Rastetter, 1996; Araújo *et al.*, 2005); this is because natural systems are

not closed, and truly predictive models would require input parameters that are incompletely known (Oreskes *et al.*, 1994). Hence, it is difficult to confirm whether agreement between observed and predicted patterns reflects that the models have captured the underlying mechanisms responsible for range shifts, or whether this is contingent on the data within essentially black-box correlative models. Only by providing repeated empirical evidence for the value of models can we reinforce confidence that their outputs are useful for prediction. This kind of repeated evidence is unlikely to be produced in the near future due to lack of suitable data for model validation. Secondly, it is important to emphasize that species–climate 'envelope' models do not project species-distributional shifts but changes in the potential distribution and availability of suitable climate for species. Assessments of species extinction risk that make assumptions on the relationship between shifts in climate space for species and species range losses assume a linear relationship between species' range size and availability of suitable climates, which may not always apply (Svenning & Skov, 2004; Araújo & Pearson, 2005). Finally, our analyses benefited from having observed climate for two periods in time, while models of future responses are heavily dependent on assumptions inherent in the climate-change scenarios adopted: a further crucial source of uncertainty.

By acknowledging limitations of models, we should be in a better position to make the best use of their results (Whittaker *et al.*, 2005). This should include: using models to corroborate hypotheses by offering evidence to strengthen what may be already established through other means; using models to elucidate discrepancies in other models; and using models for exploring 'what if' questions and thereby illuminating which aspects of a system are most in need of further investigation and where more empirical data are needed. As model outputs will always be open to question and the direction and magnitude of error be uncertain, their primary value should be heuristic more than predictive. Nonetheless, our findings suggest that while there is good reason to remain highly sceptical of particular range shift projections, the simple form of meta-analysis of climate-driven range shift modelling described herein can provide a means of constraining range shift projections to within reasonable bounds.

ACKNOWLEDGEMENTS

We thank the many volunteer fieldworkers who contributed to the atlas records; Wilfried Thuiller for discussion and for providing BROMOD, Richard Pearson, Townsend Peterson, David Stockwell and anonymous referees for comments on the manuscript. This research was funded by the EC Integrated FP6 ALARM (GOCE-CT-2003-506675) project. MBA is a EC FP6 Marie Curie Research Fellow.

REFERENCES

- Allen, M.R., Stott, P.A., Mitchell, J.F.B., Schnur, R. & Delworth, T.L. (2000) Quantifying the uncertainty in forecasts of anthropogenic climate change. *Nature*, **407**, 617–620.

- Araújo, M.B., Cabeza, M., Thuiller, W., Hannah, L. & Williams, P.H. (2004) Would climate change drive species out of reserves? An assessment of existing reserve selection methods. *Global Change Biology*, **10**, 1618–1626.
- Araújo, M.B. & Pearson, R.G. (2005) Equilibrium of species' distributions with climate. *Ecography*, in press.
- Araújo, M.B., Pearson, R.G., Thuiller, W. & Erhard, M. (2005) Validation of species–climate impact models under climate change. *Global Change Biology*, in press.
- Araújo, M.B., Williams, P.H. & Fuller, R.J. (2002) Dynamics of extinction and the selection of nature reserves. *Proceedings of the Royal Society of London, Series B, Biological Sciences*, **269**, 1971–1980.
- Austin, M.P. (1992) Modelling the environmental niche of plants — implications for plant community response to elevated CO₂ levels. *Australian Journal of Botany*, **40**, 615–630.
- Bakkenes, M., Alkemade, R.M., Ihle, F., Leemans, R. & Latour, J.B. (2002) Assessing effects of forecasted climate change on the diversity and distribution of European higher plants for 2050. *Global Change Biology*, **8**, 390–407.
- Benestad, R.E. (2004) Tentative probabilistic temperature scenarios for northern Europe. *Tellus*, **56A**, 89–101.
- Berry, P.M., Dawson, T.E., Harrison, P.A. & Pearson, R.G. (2002) Modelling potential impacts of climate change on the bioclimatic envelope of species in Britain and Ireland. *Global Ecology and Biogeography*, **11**, 453–462.
- Clemen, R.T. (1989) Combining forecasts: a review and annotated bibliography. *International Journal of Forecasting*, **5**, 559–583.
- Crick, H.Q.P. (2004) The impact of climate change on birds. *ISIS*, **146**, 48–56.
- Erasmus, B.F.N., Van Jaarsveld, A.S., Chown, S.L., Kshatriya, M. & Wessels, K.J. (2002) Vulnerability of South African animal taxa to climate change. *Global Change Biology*, **8**, 679–693.
- Fielding, A.H. & Bell, J.F. (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, **24**, 38–49.
- Gibbons, D.W., Reid, J.B. & Chapman, R.A. (1993) *The new atlas of breeding birds in Britain and Ireland: 1988–1991*. Poyser, London.
- Gregory, A.W., Smith, G.W. & Yetman, J. (2001) Testing for forecast consensus. *Journal of Business and Economic Statistics*, **19**, 34–43.
- Hampe, A. (2004) Bioclimate envelope models: what they detect and what they hide. *Global Ecology and Biogeography*, **13**, 469–471.
- Huntley, B., Berry, P.M., Cramer, W. & McDonald, A.P. (1995) Modelling present and potential future ranges of some European higher plants using climate response surfaces. *Journal of Biogeography*, **22**, 967–1001.
- Iverson, L.R. & Prasad, A. (1998) Predicting abundance for 80 tree species following climate change in the Eastern United States. *Ecological Monographs*, **68**, 465–485.
- Lennon, J.J., Greenwood, J.J.D. & Turner, J.R.G. (2000) Beta-diversity and environmental gradients in Britain: a test of the species–energy hypothesis. *Journal of Animal Ecology*, **69**, 581–598.
- Makridakis, S. & Winkler, R.L. (1983) Averages of forecasts: some empirical results. *Management Science*, **29**, 987–996.
- McNees, S.K. (1987) Consensus forecasts: tyranny of the majority? *New England Economic Review*, Nov/Dec, 15–21.
- McNees, S.K. (1992) The uses and abuses of 'consensus' forecasts. *Journal of Forecasting*, **11**, 703–710.
- Midgley, G.F., Hannah, L., Millar, D., Rutherford, M.C. & Powrie, L.W. (2002) Assessing the vulnerability of species richness to anthropogenic climate change in a biodiversity hotspot. *Global Ecology and Biogeography*, **11**, 445–451.
- New, M., Hulme, M. & Jones, P.D. (2000) Representing twentieth century space–time climate variability. Part 2: Development of 1901–96 monthly grids of terrestrial surface climate. *Journal of Climate*, **13**, 2217–2238.
- Nilsson, J., Persson, B. & von Heijne, G. (2000) Consensus predictions of membrane protein topology. *FEBS Letters*, **486**, 267–269.
- Oreskes, N., Shrader-Frechette, K.S. & Belitz, K. (1994) Verification, validation, and confirmation of numerical models in the earth sciences. *Science*, **263**, 641–646.
- Palm, F.C. & Zellner, A. (1992) To combine or not combine? Issues of combining forecasts. *Journal of Forecasting*, **11**, 687–701.
- Pearson, R.G. & Dawson, T.E. (2003) Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Global Ecology and Biogeography*, **12**, 361–371.
- Peterson, A.T., Ball, L.G. & Cohoon, K.P. (2002) Predicting distributions of Mexican birds using ecological niche modelling methods. *ISIS*, **144**, E27–E32.
- Rastetter, E.B. (1996) Validating models of ecosystem response to global change. *Bioscience*, **46**, 190–198.
- Ripley, B.D. (1996) *Pattern recognition and neural networks*. Cambridge University Press, Cambridge.
- Sanders, F. (1963) On subjective probability forecasting. *Journal of Applied Meteorology*, **2**, 191–201.
- Sengupta, S. & Boyle, J.S. (1998) Using common principal components for comparing GCN simulations. *Journal of Climate*, **11**, 816–830.
- Sharrock, J.T.R. (1976) *The atlas of breeding birds of Britain and Ireland*. Poyser, Berkhamsted.
- Skov, F. & Svenning, J.-C. (2004) Potential impact of climatic change on the distribution of forest herbs in Europe. *Ecography*, **27**, 366–380.
- Stockwell, D.R.B. & Peters, D.P. (1999) The GARP modelling system: problems and solutions to automated spatial prediction. *International Journal of Geographical Information Science*, **13**, 143–158.
- Svenning, J.-C. & Skov, F. (2004) Limited filling of the potential

- range in European tree species. *Ecology Letters*, **7**, 565–573.
- Sykes, M.T., Prentice, I.C. & Cramer, W. (1996) A bioclimatic model for the potential distributions of north European tree species under present and future climate. *Journal of Biogeography*, **23**, 203–233.
- Thomas, C.D., Cameron, A., Green, R.E., Bakkenes, M., Beaumont, L.J., Collingham, Y., Erasmus, B.F.N., de Siqueira, M.F., Grainger, A., Hannah, L., Hughes, L., Huntley, B., van Jaarsveld, A.S., Midgley, G.F., Miles, L.J., Ortega-Huerta, M.A., Townsend Peterson, A., Phillips, O. & Williams, S.E. (2004) Extinction risk from climate change. *Nature*, **427**, 145–148.
- Thomas, C.D. & Lennon, J.J. (1999) Birds extend their ranges northwards. *Nature*, **399**, 213.
- Thuiller, W. (2003) BIOMOD: Optimising predictions of species distributions and projecting potential future shifts under global change. *Global Change Biology*, **9**, 1353–1362.
- Thuiller, W. (2004) Patterns and uncertainties of species' range shifts under climate change. *Global Change Biology*, **10**, 2220–2227.
- Thuiller, W., Araújo, M.B., Pearson, R.G., Whittaker, R.J., Brotons, L. & Lavorel, S. (2004) Uncertainty in predictions of extinction risk. *Nature*, **430**, doi: 10.1038/nature02716.
- Westerhuis, J.A., Kourti, T. & Macgregor, J.F. (1998) Analysis of multiblock and hierarchical PCA and PLS models. *Journal of Chemometrics*, **12**, 301–321.
- Whittaker, R.J., Araújo, M.B., Jepson, P., Ladle, R., Watson, J.E. & Willis, K.J. (2005) Conservation biogeography: assessment and prospect. *Diversity and Distributions*, **11**, 3–23.

BIOSKETCHES

Miguel B. Araújo has a PhD from the University of London. He is a Marie Curie Research Fellow in the University of Oxford and the Natural History Museum in London. His main interests include an investigation of processes affecting occurrence and persistence of species at different spatial and temporal scales. He is also interested in the development of effective conservation-planning tools to prevent local extinctions of species.

Robert J. Whittaker is a Professor of Biogeography in the University of Oxford. His research interests include island biogeography, community assembly, plant dispersal, conservation biogeography, scale and diversity theory.

Richard J. Ladle has a D.Phil. from the University of Oxford. He is the Director of the MSc Biodiversity, Conservation and Management in the School of Geography and the Environment, University of Oxford. He has broad academic interests, which include the development of biogeographical principles as a conservation tool, diversity metrics, evolutionary ecology, science communication and environmental interpretation, and the public understanding of conservation and biodiversity science.

Markus Erhard has a PhD from Potsdam University. He is a postdoctoral research fellow at the Forschungszentrum Karlsruhe. His main research interests include climate change impacts, vulnerability and adaptation from local to European scales.

Appendix 1 List of the British bird species modelled in the present study ($n = 116$). For scientific names see, for example, the British Ornithologists' Union (<http://www.bou.org.uk>)

Red grouse	Magpie	Willow warbler
Ptarmigan	Chough	Chiffchaff
Black grouse	Jackdaw	Wood warbler
Capercaillie	Rook	Sedge warbler
Red-legged partridge	Carrion crow	Reed warbler
Grey partridge	Raven	Woodlark
Quail	Starling	Skylark
Pheasant	Ring ouzel	Duncock
Wryneck	Blackbird	House sparrow
Green woodpecker	Fieldfare	Tree sparrow
Great spotted woodpecker	Song thrush	Yellow wagtail
Lesser spotted woodpecker	Redwing	Grey wagtail
Kingfisher	Mistle thrush	Pied wagtail
Cuckoo	Spotted flycatcher	Tree pipit
Swift	Pied flycatcher	Meadow pipit
Nightjar	Stonechat	Rock pipit
Barn owl	Whinchat	Yellowhammer
Little owl	Wheatear	Cirl bunting
Tawny owl	Redstart	Reed bunting
Long-eared owl	Black redstart	Corn bunting
Short-eared owl	Robin	Chaffinch
Collared dove	Nightingale	Greenfinch
Turtle dove	Nuthatch	Goldfinch
Feral pigeon	Treecreeper	Siskin
Stock dove	Dipper	Bullfinch
Woodpigeon	Wren	Hawfinch
Woodcock	Marsh tit	Linnet
Kestrel	Willow tit	Twite
Merlin	Crested tit	Lesser redpoll
Hobby	Coal tit	Common crossbill
Peregrine	Blue tit	Scottish crossbill
Osprey	Great tit	
Red kite	Long-tailed tit	
Sparrowhawk	Sand martin	
Goshawk	Swallow	
Buzzard	House martin	
Golden eagle	Lesser whitethroat	
Marsh harrier	Whitethroat	
Hen harrier	Garden warbler	
Montagu's harrier	Blackcap	
Red-backed shrike	Bearded tit	
Jay	Grasshopper warbler	
	Goldcrest	

Appendix 2 First principal components analysis (PCA) of the 16 range shift predictions from models among the 116 British breeding bird species between 1968 and 1972 and 1995–99. Highest loading models identified as consensus models are highlighted in bold

Principal Components Scores	PC1	PC2	PC3	PC4	PC5
Cumulative explained variance	29.9	45.1	55.0	63.8	71.4
ANN ₇₀ .Kap	-2.25	-2.70	-3.47	2.06	-4.67
ANN ₇₀ .ROC	-2.98	-3.23	-4.60	1.52	-2.72
ANN ₁₀₀ .Kap	-9.27	2.02	0.20	0.15	-1.13
ANN ₁₀₀ .ROC	-0.17	-4.37	-8.70	-0.09	5.33
GAM ₇₀ .Kap	3.94	-0.31	2.11	2.14	-1.64
GAM ₇₀ .ROC	5.91	0.32	2.36	2.76	-0.71
GAM ₁₀₀ .Kap	-9.16	3.29	0.82	1.30	-2.38
GAM ₁₀₀ .ROC	5.59	-0.16	2.83	7.23	4.92
GLM ₇₀ .Kap	5.33	4.47	-0.78	-0.55	-2.14
GLM ₇₀ .ROC	6.86	5.20	-1.87	-0.76	-0.99
GLM ₁₀₀ .Kap	-8.67	4.99	2.55	-1.28	2.29
GLM ₁₀₀ .ROC	7.44	5.83	-1.91	-6.24	0.95
CTA ₇₀ .Kap	1.43	-2.88	1.95	0.14	-1.14
CTA ₇₀ .ROC	1.10	-4.33	2.68	-0.82	-1.09
CTA ₁₀₀ .Kap	-6.62	0.68	1.61	-1.19	5.00
CTA ₁₀₀ .ROC	1.52	-8.84	4.23	-6.09	0.13

Appendix 3 Chi-square and Kruskal–Wallis tests (one-way analysis of variance by ranks) of the null hypothesis that range-shift projections for contracting and expanding species are no different from each other

	Chi-Square	d.f.	Asymp. sig.	Monte Carlo Sig.		
				Sig.	99% CI	
					Lower bound	Upper bound
ANN ₇₀ .Kap	1.034	1	0.309	0.314*	0.302	0.326
ANN ₇₀ .ROC	2.292	1	0.130	0.131*	0.122	0.139
ANN ₁₀₀ .Kap	39.932	1	0.000	0.000*	0.000	0.000
ANN ₁₀₀ .ROC	4.994	1	0.025	0.027*	0.023	0.031
GAM ₇₀ .Kap	0.098	1	0.754	0.750*	0.738	0.761
GAM ₁₀₀ .Kap	34.067	1	0.000	0.000*	0.000	0.000
GAM ₁₀₀ .ROC	1.823	1	0.177	0.182*	0.172	0.191
GLM ₇₀ .Kap	4.218	1	0.040	0.036*	0.031	0.041
GLM ₇₀ .ROC	2.063	1	0.151	0.155*	0.146	0.165
GLM ₁₀₀ .Kap	34.637	1	0.000	0.000*	0.000	0.000
GLM ₁₀₀ .ROC	4.147	1	0.042	0.042*	0.037	0.047
CTA ₇₀ .Kap	0.148	1	0.700	0.707*	0.696	0.719
CTA ₇₀ .ROC	0.768	1	0.381	0.378*	0.365	0.390
CTA ₁₀₀ .Kap	28.481	1	0.000	0.000*	0.000	0.000
CTA ₁₀₀ .ROC	0.093	1	0.761	0.763*	0.752	0.774
Median _{All}	10.499	1	0.001	0.001*	0.000	0.002
Median _{Consensus}	45.990	1	0.000	0.000*	0.000	0.000

*Based on 10,000 sampled tables with starting seed 2,000,000.