ECOGRAPHY 29: 129–151, 2006

Novel methods improve prediction of species' distributions from occurrence data

Jane Elith*, Catherine H. Graham*, Robert P. Anderson, Miroslav Dudík, Simon Ferrier, Antoine Guisan, Robert J. Hijmans, Falk Huettmann, John R. Leathwick, Anthony Lehmann, Jin Li, Lucia G. Lohmann, Bette A. Loiselle, Glenn Manion, Craig Moritz, Miguel Nakamura, Yoshinori Nakazawa, Jacob McC. Overton, A. Townsend Peterson, Steven J. Phillips, Karen Richardson, Ricardo Scachetti-Pereira, Robert E. Schapire, Jorge Soberón, Stephen Williams, Mary S. Wisz and Niklaus E. Zimmermann

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

Prediction of species' distributions is central to diverse applications in ecology, evolution and conservation science. There is increasing electronic access to vast sets of occurrence records in museums and herbaria, yet little effective guidance on how best to use this information in the context of numerous approaches for modelling distributions. To meet this need, we compared 16 modelling methods over 226 species from 6 regions of the world, creating the most comprehensive set of model comparisons to date. We used presence-only data to fit models, and independent presence-absence data to evaluate the predictions. Along with well-established modelling methods such as generalised additive models and GARP and BIOCLIM, we explored methods that either have been developed recently or have rarely been applied to modelling species' distributions. These include machine-learning methods and community models, both of which have features that may make them particularly well suited to noisy or sparse information, as is typical of species' occurrence data. Presence-only data were effective for modelling species' distributions for many species and regions. The novel methods consistently outperformed more established methods. The results of our analysis are promising for the use of data from museums and herbaria, especially as methods suited to the noise inherent in such data improve.

J. Elith (j.elith@unimelb.edu.au), School of Botany, Univ. of Melbourne, Parkville, Victoria, 3010 Australia. – C. H.Graham, Dept of Ecology and Evolution, 650 Life Sciences Building, Stony Brook Univ., NY 11794, USA. – R. P. Anderson, City College of the City Univ. of New York, NY, USA. – M. Dudík, Princeton Univ., Princeton, NJ, USA. – S. Ferrier, Dept of Environment and Conservation Armidale, NSW, Australia. – A. Guisan, Univ. of Lausanne, Switzerland. – R. J. Hijmans, The Univ. of California, Berkeley, CA, USA. – F. Huettmann, Univ. of Alaska Fairbanks, AK, USA. – J. R. Leathwick, NIWA, Hamilton, NZ. – A. Lehmann, Swiss Centre for Faunal Cartography (CSCF), Neuchâtel, Switzerland. – J. Li, CSIRO Atherton, Queensland, Australia. – L. Lohmann, Univ.de São Paulo, Brasil. – B. A. Loiselle, Univ. of Missouri, St. Louis, USA. – G. Manion, Dept of Environment and Conservation, NSW, Australia. – C. Moritz, The Univ. of California, Berkeley, USA. – M. Nakamura, Centro de Invest. en Matemáticas (CIMAT), México. – Y. Nakazawa, The Univ. of Kansas, Lawrence, KS, USA. – J. McC. Overton, Landcare Research, Hamilton, NZ. – A. T. Peterson, The Univ. of Kansas, Lawrence, KS, USA. – S. J. Phillips, AT&T Labs-Research, Florham Park, NJ, USA. – K. S. Richardson, McGill Univ., QC, Canada. – R. Scachetti-Pereira, Centro de Referência em Informação Ambiental, Brazil. – R. E. Schapire, Princeton Univ., Princeton, NJ, USA. – J. Soberón, The Univ. of Kansas, Lawrence, KA, USA. – S. E.Williams, James Cook Univ., Queensland, Australia. – M. S. Wisz, National Environmental Research Inst., Denmark. – N. E. Zimmermann, Swiss Federal Research Inst. WSL, Birmensdorf, Switzerland.

*The first two authors have contributed equally to this paper.

Accepted 25 January 2006

Detailed knowledge of species' ecological and geographic distributions is fundamental for conservation planning and forecasting (Ferrier 2002, Funk and Richardson 2002, Rushton et al. 2004), and for understanding ecological and evolutionary determinants of spatial patterns of biodiversity (Rosenzweg 1995, Brown and Lomolino 1998, Ricklefs 2004, Graham et al. 2006). However, occurrence data for the vast majority of species are sparse, resulting in information about species distributions that is inadequate for many applications. Species distribution models attempt to provide detailed predictions of distributions by relating presence or abundance of species to environmental predictors. As such, distribution models have provided researchers with an innovative tool to explore diverse questions in ecology, evolution, and conservation. For example, they have been used to study relationships between environmental parameters and species richness (Mac Nally and Fleishman 2004), characteristics and spatial configuration of habitats that allow persistence of species in landscapes (Araújo and Williams 2000, Ferrier et al. 2002a, Scotts and Drielsma 2003), invasive potential of non-native species (Peterson 2003, Goolsby 2004), species' distributions in past (Hugall et al. 2002, Peterson et al. 2004) or future climates (Bakkenes et al. 2002, Skov and Svenning 2004, Araújo et al. 2004, Thomas et al. 2004, Thuiller et al. 2005), and ecological and geographic differentiation of the distributions of closely-related species (Cicero 2004, Graham et al. 2004b).

Most research on development of distribution modelling techniques has focused on creating models using presence/absence or abundance data, where regions of interest have been sampled systematically (Austin and Cunningham 1981, Hirzel and Guisan 2002, Cawsey et al. 2002). However, occurrence data for most species have been recorded without planned sampling schemes, and the great majority of these data consist of presence-only records from museum or herbarium collections that are increasingly accessible electronically (Graham et al. 2004a, Huettmann 2005, Soberón and Peterson 2005). The main problem with such occurrence data is that the intent and methods of collecting are rarely known, so that absences cannot be inferred with certainty. These data also have errors and biases associated with them, reflecting the frequently haphazard manner in which samples were accumulated (Hijmans et al. 2000, Reese et al. 2005). Thus, the considerable potential of occurrence data for analysis of biodiversity patterns will only be realised if we can use them critically. Simultaneous with increasing accessibility of species' occurrence data, environmental data layers of high spatial resolution, such as those derived from satellite images (Turner et al. 2003) and through sophisticated interpolation of climate data (Thornton et al. 1997, Hijmans et al. 2005), are now much more abundant and available. In parallel, development of methods for efficient exploration and summary of patterns in large databases has accelerated in other disciplines (Hastie et al. 2001), but only a few of these have been applied in ecological studies. Given the widespread use of distribution modelling, and the synergy of advances in data availability and modelling methods, a clear need exists for broad synthetic analyses of the predictive ability and accuracy of species' distribution modelling methods for presence-only data.

There is now a plethora of methods for modelling species' distributions that vary in how they model the distribution of the response, select relevant predictor variables, define fitted functions for each variable, weight variable contributions, allow for interactions, and predict geographic patterns of occurrence (Guisan and Zimmerman 2000, Burgman et al. 2005, Wintle and Bardos in press). Initial attempts to analyze presence-only data used methods developed specifically for that purpose, based either on calculations of envelopes or distance-based measures (Gómez Pompa and Nevling 1970, Rapoport 1982, Silverman 1986, Busby 1991, Walker and Cocks 1991, Carpenter et al. 1993). Attention then turned to adapting presence-absence methods (i.e. those that model a binomial response) to model presence-only data, using samples of the background environment (random points throughout the study area), or of areas designated as "non-use" or "pseudo-absence" (Stockwell and Peters 1999, Boyce et al. 2002, Ferrier et al. 2002a, Zaniewski et al. 2002, Keating and Cherry 2004, Pearce and Boyce in press).

More recently several novel modelling methods have been proposed that have foundations in ecological and/or statistical research, that may perform well for distribution modelling with noisy data, such as presence-only records. Some of these methods use information on the presences of other members of the community to supplement information for the species being modelled. Community methods are promising, especially for rare species, because the additional information carried by the wider community may help to inform the modelled relationships. Further, extensive research in the machine-learning and statistical disciplines has produced methods that are able to capture complex responses, even with noisy input data. These have received very little exposure in distribution modelling, though the work that has been done is promising (Phillips et al. 2006, Leathwick et al. in press.).

Regardless of the modelling method chosen, a major problem in evaluation is that species' distributions are not known exactly. In many instances evaluation focuses on predictive performance, some known occurrences are withheld from model development (either by splitting the data set, k-fold partitioning, or bootstrapping; Fielding and Bell 1997, Hastie et al. 2001, Araújo et al. 2005), and accuracy is assessed based on how well models predict the withheld data (Boyce et al. 2002,

Hirzel et al. unpubl.). In presence-only modelling, such withheld data are unlikely to provide a general test of model accuracy in predicting species' distributions, because the occurrence records often have biases in both geographic and environmental space (Bojórquez et al. 1995, Hijmans et al. 2000, Soberon et al. 2000, Kadmon et al. 2004) and such biases will persist in common resampling designs. More importantly, withheld data are presence-only, which limits the options for, and power of, statistical evaluations of predictive performance. One step towards improving evaluation of model performance in predicting distributions of species is to use independent, well structured presence-absence datasets for validation. Such datasets have rarely been used to evaluate predictions from presence-only models (but see Ferrier and Watson (1997) for an earlier example of this type of evaluation). Further possibilities include modelling artificial data and assessing whether responses are correctly predicted (Austin et al. 1995), or modelling with both presence-only and presence-absence data and comparing the fitted functions.

The primary aim of this research was to evaluate the capacity of presence-only occurrence data for predicting species' distributions. We chose to focus on evaluation at independent sites, using the performance of different modelling methods averaged over many species. Detailed evaluation of the ecological realism of all models was not practical. We tested the performance of a representative selection of modelling methods for presence-only data, using data sets typical of the types of species and environmental data that are commonly employed. This exercise also provides insights into whether and how these increasingly available data can be used for improving knowledge of species' ranges. We compiled data only for regions in which independent presence-absence data were available for evaluation and invited participation from researchers experienced in a range of distribution modelling methods, including several novel methods that have not been used widely in ecology. Our model comparison is very broad, applying 16 methods to modelling the distributions of 226 species from 6 regions around the globe.

Materials and methods

Data for modelling and evaluation

Our intention in selecting data was to collect representative examples of the types of data and species that are commonly used in species distribution modelling. Species locality data were assembled for six regions of the world (Tables 1 and 2): birds and plants of the Australian Wet Tropics (AWT); birds of Ontario, Canada (CAN); plants, birds, mammals and reptiles of north-east New South Wales, Australia (NSW), plants of New Zealand (NZ), plants from five countries (see footnote, Table 1) of South America (SA), and plants of Switzerland (SWI). The individual data sets had up to 54 species with a range of geographic extents and rarities. For each region, two sets of data were available: 1) presence-only (PO) data, from unplanned surveys or incidental records, including those from museums and herbaria; and 2) independent presence-absence (PA) data from planned surveys, with accurate location records. The independence of these two data sets for each region is an important underlying assumption of this study, and the totally different data collection methods for the two sets give us a high level of confidence that this assumption is well founded. Table 2 summarises the key characteristics of the data. One feature is that the accuracy and sample size of the modelling data vary between regions: mostly they were small-to-moderate (10^1–10^2 occurrences), but were larger (10^1–10^3) for SWI. Evaluation (PA) sets were larger for CAN, NZ and SWI than the other regions. Further, the PA data sets for AWT, NSW, NZ and SWI all provide better environmental and geographic coverage than those for CAN and SA.

The environmental data used for each region were selected for their relevance to the species being modelled (Austin 2002), as determined by the data provider (Tables 1 and 3). Eleven to 13 predictors were supplied per region, with grid cell sizes ranging from ca 100 by 100 m (AWT, NSW, NZ, SWI) to 1000 by 1000 m (CAN, SA). Several regions had data sets where previous research informed the development of ecologically relevant predictors (NSW, NZ, SWI). Other regions simply used variables typical of those used in distribution modelling, with emphasis on climatic data (CAN, SA, AWT). None of the exact data sets used in our study have previously been used for modelling, but some published studies (Ferrier and Watson 1997, Guisan et al. 1998, Venier et al. 2001, Leathwick 2002) use subsets of these or closely related data.

The data as supplied required considerable grooming and manipulation to generate datasets of consistent quality both within and between regions. For each region, environmental predictor variables were manipulated so that projections, grid cell size and alignment, and spatial extent were consistent across all layers. Some species data (particularly in the PO datasets) had many records per grid cell because of either repeat observations over years, or sites in close proximity to each other. We reduced both the PO and PA datasets to one record per grid cell. For the PA data, if any record in the grid cell was a presence, presence was assigned to the one record that was kept regardless of whether there was also an absence in the same grid cell. This was not a common feature of the data, but it occurred at least once in all data sets except SA. In a small number of cases both a PO and PA record occurred in the same grid cell; in these

Table 1. Summary of data available for modelling and evaluation.

Region	Species records			Predictor variables	
	Groups (number species)	PO[1]: number (mean)	PA[1]: number sites; mean	Broad class	Cell size (m) and extent ($km^2 \times 10^6$)
Australian Wet Tropics (AWT)	birds (20) plants (20)	155 35	340; 97 102; 30	10 climate, 3 topography	80×80 m 0.024
Ontario, Canada (CAN)	birds (20)	255	14571; 1282	6 climate, 3 topography, 1 distance, 1 vegetation	1×1 km 1.088
New South Wales, Australia (NSW)	birds (10) plants (29) mammals (7) reptiles (8)	162 22 27 84	mean[3] 920; 74 mean[3] 1333; 214 570; 76 1008; 62	5 climate, 2 soil, 1 site moisture, 3 topography, 1 disturbance, 1 vegetation	100×100 m 0.089
New Zealand (NZ)	plants (52)	18–211	19120; 1801	8 climate, 2 substrate, 3 topography	100×100 m 0.265
South America[2] (SA)	plants (30)	17–216	152; 12	11 climate	1×1 km 14.654
Switzerland (SWI)	plants (30)	36–5822	10013; 810	7 climate, 2 substrate, 2 topography, 2 vegetation	100×100 m 0.041

[1] PO is the presence-only modelling data and PA, the presence-absence evaluation data.
[2] Five countries: continental Brazil, Ecuador, Colombia, Bolivia, and Peru.
[3] Means of subsets of sites appropriate to different groups of plants and birds (Table 2).

the PA record was usually deleted in order to ensure that there was no spatial overlap between the modelling and evaluation data sets. The exception was for the AWT data, where the PO record was deleted due to limited PA data.

As some of the modelling methods required data akin to absences, background samples (sometimes also referred to as "pseudo-absences"; Ferrier et al. 2002a) were generated by drawing a random sample of 10 000 sites for each region. These were intended as a sample of the whole region, and it is possible that a background sample coincided with a presence record. In another study we address alternative strategies for creating such data (and see Zaniewski et al. 2002, Graham et al. 2004a). It is important to note that the community methods (MARS-COMM and GDM) described below use the species data differently to single species methods (see Appendix, Text S1). In particular, they define a model using all sites available for all the species of the relevant biological group, and assume absences at sites if presence is not recorded, effectively treating all presence sites for a group of species as indicating an absence for any species not recorded at a particular site. To preserve as much consistency as possible between the single species and community modelling approaches, we used background samples in addition to the community data for fitting the community models. Nevertheless, the inclusion of the community data sets these apart from the single-species methods.

Modelling methods

Eleven distinct modelling methods were used, but a number of these were implemented in more than one way, resulting in the 16 approaches presented here (Table 4 and Appendix, Text S1). The methods form two broad groups based on the type of data they use: those that only use presence records (BIOCLIM, DOMAIN, LIVES), and those that characterise the background with a sample (all other methods). Among the techniques that characterise the background, a critical distinction exists between those that use only the presence records for the modelled species vs those that use data describing the presences of other species, i.e. community-based techniques. Details and key references for each method are in Appendix, Text S1 and Table 4, and the following briefly introduces the methods, with more detail for lesser-known techniques.

The first group of methods (those that only use presence records) includes one envelope-style method (BIOCLIM) that characterises sites that are located within the environmental hyper-space occupied by a species, and two distance-based methods (DOMAIN and LIVES) that assess new sites in terms of their environmental similarity to sites of known presence. The two distance-based methods differ both in their theoretical assumptions and the procedures used for calculating similarities.

The second group of methods includes several regression approaches (Table 4). Generalised linear models

Table 2. Species data.

Region	Species information	PO data	PA data
AWT	20 birds and 20 plants.	Birds from incidental surveys (accurate locations[1]); plants from herbarium records within last 40 yr (±3 km) #presence records: Birds mean 155, range 32–265; 78% occupied cells have >1 species. Plants: mean 35, range 9–74; 37% occupied cells have >1 species.	Birds from planned field surveys at 340 sites (accurate locations); plants from planned surveys over 20 yr at 102 sites (accurate locations) #presence records: birds mean 97, range 32–265; plants: mean 30, range 13–214.
CAN	20 birds.	Ontario Nest Records, Royal Ontario Museum (ROM). Temporal Span 1870–2002 (mostly 1960–2001). Coordinates derived from map by ROM; some locations with GPS. #presence records: mean: 255, range: 16–749; 26% occupied cells have >1 species.	Breeding Bird Atlas (BBA) for Ontario. 14571 sites #presence records: mean: 1282, range: 24–4512.
NSW	54 species: 7 bats (ba); 8 diurnal birds (db); 2 nocturnal birds (nb); 8 open-forest trees (ot); 8 open-forest understorey plants (ou); 7 rainforest trees (rt); 6 rainforest understorey plants (ru); 8 small reptiles (sr).	Fauna incidental records from Atlas of NSW Wildlife; flora from herbaria. Across all groups, #presence records: mean: 62, range: 2–426; 21% occupied cells have >1 species.	From designed surveys (Ferrier and Watson 1996, Pearce et al. 2001). Accurate locations. #of sites: ba 570, db 702, nb 1137, ot 2076, ou 1309, rt 1036, ru 909, sr 1008. Across all groups, #presence records: mean: 148, range: 4–693.
NZ	52 plants, mostly trees and shrubs.	Herbarium data. #presence records: mean: 59, range 18–211; 16% occupied cells have >1 species.	Designed surveys; 19 120 sites. Accurate locations. #presence records: mean: 1801, range 20–10 581.
SA	30 plants, family Bignoniaceae.	Herbarium data. #presence records: mean: 74, range 17–216; 36% occupied cells have >1 species.	Al Gentry's data (Missouri Botanic Gardens); 152 sites. #presence records: mean: 12, range 7–29.
SWI	30 trees.	Forest vegetation data from non-systematic surveys. 1913–1998; 75% post-1940. #presence records: mean: 1170, range 36–5822; 80% occupied cells have >1 species.	Forest inventory plots on regular lattice; 10 013 sites. Accurate locations. #presence records: mean: 810, range: 19–6953.

[1] Locations are considered accurate if location error is estimated to be ≤100 m. Where accuracy data are not presented, no information was provided.

Table 3. Environmental data.

Region	Grid cell size	Projection	Variables	Correlated pairs (pearson r > 0.85)
AWT	80 m	UTM	10 climate from BIOCLIM (1: annual mean temp, 2: temp seasonality, 3: max temp of warmest month, 4: min temp of coolest month, 5: annual precipitation, 6: precip seasonality, 7: precip driest quarter, 8: annual mean radiaton, 9: moisture index (MI) seasonality, 10: mean MI of lowest quarter MI), 11: slope, 12: topographic position, 13: terrain ruggedness index.	1 with 3,4;5 and 8 with 7;6 and 8 with 9;10 with 6,7,8,9.
CAN	1 km	unprojected	6 climate from WORLDCLIM (1: annual mean temp, 2: april temp, 3: temperature seasonality, 4: annual precipitation, 5: precip seasonality, 6: precip driest quarter), 7: altitude, 8: aspect (northness), 9: slope, 10: distance from Hudson Bay, 11: vegetation class (5 classes).	1 with 2,5,6,10;2 with 3,5,6,10;3 with 5,6;4 with 5,6; 5 with 6.
NSW	100 m	unprojected	5 climate (mean annual rainfall, mean rainfall of driest quarter, annual mean temperature, minimum temperature of the coldest month, annual mean solar radiation), moisture index, soil fertility, soil depth, ruggedness, topographic position, compound topographic index, disturbance, vegetation class (9 classes).	min temp with mean temp.
NZ	100 m	NZ map grid	8 climate (mean October vapour pressure deficit at 09:00 h, vapour pressure deficit, mean annual solar radiation, mean annual temperature, temperature seasonality, average monthly ratio of rainfall to potential evaporation(r2pet), annual precipitation, solar radiation seasonality), age of bedrock, toxic cations in soil, altitude, slope, hillshade.	r2pet with annual precip; mean annual temp with altitude.
SA	1 km	unprojected	11 climate variables from WORLDCLIM (1: annual mean temperature, 2: mean diurnal range, 3: temperature seasonality, 4: max temp of warmest month, 5: min temp of coldest month, 6: temp annual range, 7: mean temp of wettest quarter, 8: annual precipitation, 9: precip seasonality, 10: precip driest quarter, 11: precip warmest quarter).	1 with 4,5,7;2 with 6,4 and 5 with 7.
SWI	100 m	UTM	7 climate (1: growing days above freezing, 2: average temp coldest month, 3: days of summer frost, 4: annual precip, 5: days rain > 1 mm, 6: site water balance, 7: potential yearly global radiation), slope, topographic position, soil nutrient index, calcareous bedrock, broadleaf cover, conifer cover.	1 with 2.

(GLMs) and generalised additive models (GAMs) are used extensively in species' distribution modelling because of their strong statistical foundation and ability to realistically model ecological relationships (Austin 2002). GAMs use non-parametric, data-defined smoothers to fit non-linear functions, whereas GLMs fit parametric terms, usually some combination of linear, quadratic and/or cubic terms. Because of their greater flexibility, GAMs are more capable of modelling complex ecological response shapes than GLMs (Yee and Mitchell 1991). BRUTO provides a rapid method to identify both the variables to include and the degree of smoothing to be applied in a GAM model, but has only recently been used in ecological applications (Leathwick et al. unpubl.). Multivariate adaptive regression splines (MARS) provide an alternative regression-based method for fitting non-linear responses, using piecewise linear fits rather than smooth functions. They are much faster to implement than GAMs, and simpler to use in GIS applications when making maps of predictions. An added feature that we investigate here is their ability to analyse community data (Leathwick et al. 2005), i.e. to simultaneously relate variation in the occurrence of all species to the environmental predictors in one analysis, and then estimate individual model coefficients for each species (MARS-COMM). Most of our implementations of the regression methods did not attempt to model interactions, with the exception of the MARS models, where we allowed the fitting of first-order interactions in single species models (MARS-INT).

We implemented two versions of GARP: the desktop version that has been used widely for modelling data from natural history collections (DK-GARP), and a new open modeller implementation (OM-GARP), that has updated algorithms for developing rule sets. These both use a genetic algorithm to select a set of rules (e.g. adaptations of regression and range specifications) that best predicts the species distribution (Stockwell and Peters 1999).

Two methods have been developed within the machine learning community: maximum entropy models (MAXENT and MAXENT-T) and boosted regression trees (BRT, also called stochastic gradient boosting). MAXENT estimates species' distributions by finding the distribution of maximum entropy (i.e. closest to uniform) subject to the constraint that the expected value of each environmental variable (or its transform and/or interactions) under this estimated distribution matches its empirical average (Phillips et al. 2006). In the MAXENT application for modelling presence-only species' data, choices can be made about the complexity of the fitted functions. BRT combines two algorithms: the boosting algorithm iteratively calls the regression-tree algorithm to construct a combination or "ensemble" of trees. Regression trees are

Table 4. Modelling methods implemented.

Method	Class of model, and explanation	Data[1]	Software	Std errors?[2]	Contact person
BIOCLIM	envelope model	p	DIVA-GIS	no	CG, RH
BRT	boosted decision trees	pa	R, gbm package	no	JE
BRUTO	regression, a fast implementation of a gam	pa	R and Splus, mda package	yes	JE
DK-GARP	rule sets from genetic algorithms; desktop version	pa	DesktopGarp	no	ATP
DOMAIN	multivariate distance	p	DIVA-GIS	no	CG, RH
GAM	regression: generalised additive model	pa	S-Plus, GRASP add-on	yes	AG,AL,JE
GDM	generalised dissimilarity modelling; uses community data	pacomm	Specialized program not general released; uses Arcview and Splus	no	SF
GDM-SS	generalised dissimilarity modelling; implementation for single species	pa	as for GDM	no	SF
GLM	regression; generalised linear model	pa	S-Plus, GRASP add-on	yes	AG,AL,JE
LIVES	multivariate distance	p	Specialized program not general released	no	JLi
MARS	regression; multivariate adaptive regression splines	pa	R, mda package plus new code to handle binomial responses	yes	JE, FH
MARS-COMM	as for MARS, but implemented with community data	pacomm	as for MARS	yes	JE
MARS-INT	as or MARS; interactions allowed	pa	as for MARS	yes	JE
MAXENT	maximum entropy	pa	Maxent	no	SP
MAXENT-T	maximum entropy with threshold features	pa	Maxent	no	SP
OM-GARP	rule sets derived with genetic algorithms; open modeller version	pa	new version of GARP not yet available	no	ATP

[1] p = only presence data used; pa = presence and some form of absence required – e.g. a background sample; comm = community data contribute to model fitting.
[2] any method can have an uncertainty estimate derived from bootstrapping the modelling; these data refer to estimates that are available as a statistical part of the method.

used because they are good at selecting relevant variables and can model interactions; boosting is used to overcome the inaccuracies inherent in a single tree model. The regression trees are fitted sequentially on weighted versions of the data set, where the weights continuously adjust to take account of observations that are poorly fitted by the preceding models. Boosting can be seen as a method for developing a regression model in a forward stage-wise fashion, at each step adding small modifications in parts of the model space to fit the data better (Friedman et al. 2000). When using BRT, we avoided overfitting by using cross-validation to progressively grow models while testing predictive accuracy on withheld portions of the data.

Finally, generalised dissimilarity models (GDM) model spatial turnover in community composition (or "compositional dissimilarity") between pairs of sites as a function of environmental differences between these sites. The approach combines elements of matrix regression and generalised linear modelling, thereby allowing it to model non-linear responses to the environment that capture ecologically realistic relationships between dissimilarity and ecological distance (Ferrier 2002, Ferrier et al. 2002b). For predicting species' distributions, an additional kernel regression algorithm (Lowe 1995) is applied within the transformed environmental space generated by GDM, to estimate likelihoods of occurrence of a given species at all sites. Two versions of this approach were applied in the current study: 1) "GDM" in which a single GDM was fitted to the combined data for all species in a given biological group, such that the output from this GDM was then used as a common basis for all of the subsequent kernel regression analyses; and 2) "GDM-SS" in which a separate GDM was fitted to the data for each species alone, such that kernel regression analysis for each species was based on the output from a GDM tailored specifically to that species.

As the manner in which a particular method is implemented can have substantial effects on model performance, we deliberately used experienced analysts to develop all models. We also implemented batch processing of methods in order to run the large number of models presented here, using settings judged by the modellers to provide a robust and reliable implementation of the method. Details are contained in Appendix, Text S1 and Table S1.

Experimental design

All analyses were carried out by modellers (Appendix, Table S2) blind to the evaluation data, which was not examined until after all modelling was complete. We provided modellers with the presence-only (PO) locations for each species and random background locations (one set for each of the 6 regions). For each region, environmental data were presented in two formats: either as a table that included the environmental variables for each PO and random locality, or as environmental grids.

The modellers could use the modelling data in any way they chose, to decide how to develop their models (e.g. they could run cross-validation testing on the PO data). Based on these decisions, modellers made predictions for each species using a set of evaluation data, one set of which was prepared for each region. This consisted of a table of environmental data for the evaluation sites, but it contained no species records, i.e., predictions were made from the PO data with no knowledge of the pattern of the species as described by the PA dataset. Modellers who applied more than one method to the data (Appendix, Table S2) approached each method as a new situation and aimed to implement the method in the best way possible for a multi-species analysis (Appendix, Text S1).

Evaluation

The evaluation focussed on predictive performance at sites. We used three statistics, the area under the Receiver Operating Characteristic curve (AUC), correlation (COR) and Kappa, to assess the agreement between the presence-absence records and the predictions. AUC has been used extensively in the species' distribution modelling literature, and measures the ability of a model to discriminate between sites where a species is present, versus those where it is absent (Hanley and McNeil 1982). This provides an indication of the usefulness of the models for prioritising areas in terms of their relative importance as habitat for the particular species. The AUC ranges from 0 to 1, where a score of 1 indicates perfect discrimination, a score of 0.5 implies predictive discrimination that is no better than a random guess, and values <0.5 indicate performance worse than random. "Worse than random" can occur because a model may fit the modelling data but predict badly, and we tested predictive performance on independent data rather than model fit. AUC values can be interpreted as indicating the probability that, when a presence site and an absence site are drawn at random from the population, the first will have a higher predicted value than the second. It is closely related to a Mann-Whitney U statistic, and it is in this context it is seen to be a rank-based statistic – the prediction at the presence site can be higher than the prediction at the absence site by a small or large amount, and the value of the statistic will be the same. Standards errors were calculated with the methods of Hanley and McNeil (1982).

The correlation, COR, between the observation in the PA dataset (a dichotomous variable) and the prediction, is known as the point biserial correlation, and can be calculated as a Pearson correlation coefficient (COR; (Zheng and Agresti 2000)). It is similar to AUC, but carries with it extra information: instead of being rank-based, it takes into account how far the prediction varies from the observation. This gives further insight into the distribution of the predictions, and in the technical evaluation framework of Murphy and Winkler (1992) further informs the user about the model's discrimination.

Kappa (Cohen 1960), which is a chance-corrected measure of agreement, is commonly used in ecological studies with presence-absence data. It requires a threshold to be applied to the predictions, to convert them to presence-absence predictions. Kappa provides an index that considers both omission and commission errors. We calculated a maximum kappa (KAPPA) for each model by calculating kappa at all possible thresholds for each species-specific set of predictions and identifying both the maximum kappa and the threshold at which this occurred. This method used information not available to the modeller – i.e. information from the evaluation data set. It returns the best possible KAPPA score for each method. Liu et al. (2005) have demonstrated that other methods are more reliable for selecting thresholds from the training data (i.e. that used in model development), but in this case we wanted to characterise the best possible KAPPA value that could be attained on these evaluation data, so that threshold selection did not confound the results.

Variation in AUC and COR values were analysed using Generalized Linear Mixed Models, with the statistic (e.g. AUC) as the response and modelling method fitted as a fixed effect. Both species and an interaction between modelling method and region were fitted as random effects, the interaction term allowing for differing performance of methods across regions. Analyses were performed using WinBUGS (Spiegelhalter et al. 2003a), which fits a Bayesian model. We assumed uninformative priors for all parameters, resulting in a GLMM that is equivalent to one fitted using maximum likelihood. Comparisons of the sixteen methods were summarized from 50 000 Monte Carlo iterations after a burn-in period of 10 000. The performance of modelling method was summarized as the mean and standard deviation of the posterior distributions. The percentage of runs where the response (e.g. AUC) for method X is greater than that for method Y estimates the probability that the true difference between the methods is greater than zero. This is a 2-tailed test, and values close to 1 mean that method X's response is greater than that of method Y, and vice-versa for values close to zero. The importance of each term in the GLMM was assessed by change in the Deviance Information Criterion (DIC, Spiegelhalter et al. 2003b) for the full model compared with subsets where each term was excluded from the model. The DIC is the Bayesian equivalent of Akaike's Information Criterion, and rules of thumb suggest that changes in DIC of >10 units indicate that the excluded term had an important effect (Burnham and Anderson 2002, McCarthy and Masters 2005).

To further explore the results, we calculated a series of metrics that define the distances between sites, and the area occupied, in both environmental and geographic space. These are useful for assessing to what extent a species "fills" the regional space. The nearest neighbour (NN, also called p-median) summarises the distances between points in multidimensional environmental space, using a Manhattan distance (Sneath and Sokal 1973). We used the 10 000 random points and the presences in the evaluation data set and calculated the median of the minimum distances between any one random point and all the presence points. In each dimension, the distance was scaled by the range of the random points. NN values are larger for species that occupy only part of the environmental space of the region. The range overlap method used the same sets of points but summarised the mean overlap in ranges over all environmental dimensions; high values indicate species that span most of the environmental space in the regions. In contrast to those two methods, maximum distance and area of convex polygons are measures in geographic space, calculated on presence only data but presented in relation to the maximum measures in the presence absence data for each biological group in each region. Small distances or areas indicate species that are restricted geographically.

Finally, we recorded time taken for modelling and made comments on potential improvements, and these are presented in Appendix, Table S2.

Results

To display mapped model results, we show distributions predicted with several modelling methods for species in NSW (Fig. 1 and Appendix, Fig. S1). These maps illustrate variation in model predictions among techniques. There is considerable agreement between some methods. The most obvious differences are in the proportion of the region that appears to be predicted most suitable for the species, and although these might stem from predictions that are scaled in different ways, the variation in AUC suggests actual differences between methods.

Evaluating the results at independent sites across all species and methods, we found clear indications that presence-only data can provide the basis for accurate predictions, but also marked variation in modelling success (Fig. 2 and Appendix, Table S3). For example, although AUC varied from 0.07 to 0.97, 40% of models had an AUC >0.75 (a useful amount of discrimination; see methods), and 90% of methods performed better than random (AUC >0.5). The following analyses explore trends and sources of variation among methods, regions and species.

A) Broad trends across regions and species

Assessments of modelling success using AUC and COR indicate that methods can be analysed in three groups (Fig. 3 and Appendix, Fig. S2). The first and highest-performing group (above-right of the solid black line in Fig. 3) included MARS community (MARS-COMM), boosted regression trees (BRT), generalised dissimilarity (GDM and GDM-SS) and maximum entropy (MAXENT and MAXENT-T) models, all of which performed relatively well according to each of the evaluation measures (Figs 3 and 4). A second group of methods (between the solid and dashed diagonal black lines, Fig. 3) showed intermediate performance for AUC and COR. It included most of the standard regression methods (generalised additive models – GAM and BRUTO; generalised linear models – GLM; individual multivariate adaptive regression splines – MARS), and OM-GARP. Models in the third group all performed relatively poorly (lower left of the dashed black line, Fig. 3) and included the 3 methods that use only presence data (BIOCLIM, LIVES, DOMAIN) with no inferred absences, along with DK-GARP, and the MARS individual models fitted with interactions (MARS-INT). This group also deviated from the generally linear relationship between AUC and COR results, i.e. assessment of their predictive success depends on which measure is used. DOMAIN was close to the middle group in relation to AUC but not COR, and DK-GARP was close for COR but less successful with AUC. DK-GARP results are not completely comparable to the others because this method could not be applied to for one region (NZ) because of computational constraints.

The GLMMs confirmed our graphical interpretation and indicated that differences in the performance of modelling methods were statistically important, because the removal of the modelling method term from the full model resulted in a change in the Deviance Information Criterion of 472 units (Table 5). These results are for AUC, and those for COR are comparable. Results from the pair-wise comparison of methods (Table 6) indicate that those methods occurring to the top right in Fig. 3 provide significantly better performance that those to the lower left, and also to several methods located in the central part of this figure. The error bars in Fig. 3 are those estimated from the model, and bars that do not overlap coincide with high probabilities that the methods are different (Fig. 3, Table 6).

The same general trends are evident also for KAPPA (Fig. 4), although methods are not so clearly separated because KAPPA is a less sensitive measure, estimating performance at a single prediction threshold. Nevertheless, the highest-performing methods as assessed by AUC and COR also had the highest KAPPA scores (to the right in Fig. 4) and the presence-only methods also ranked among the lowest.

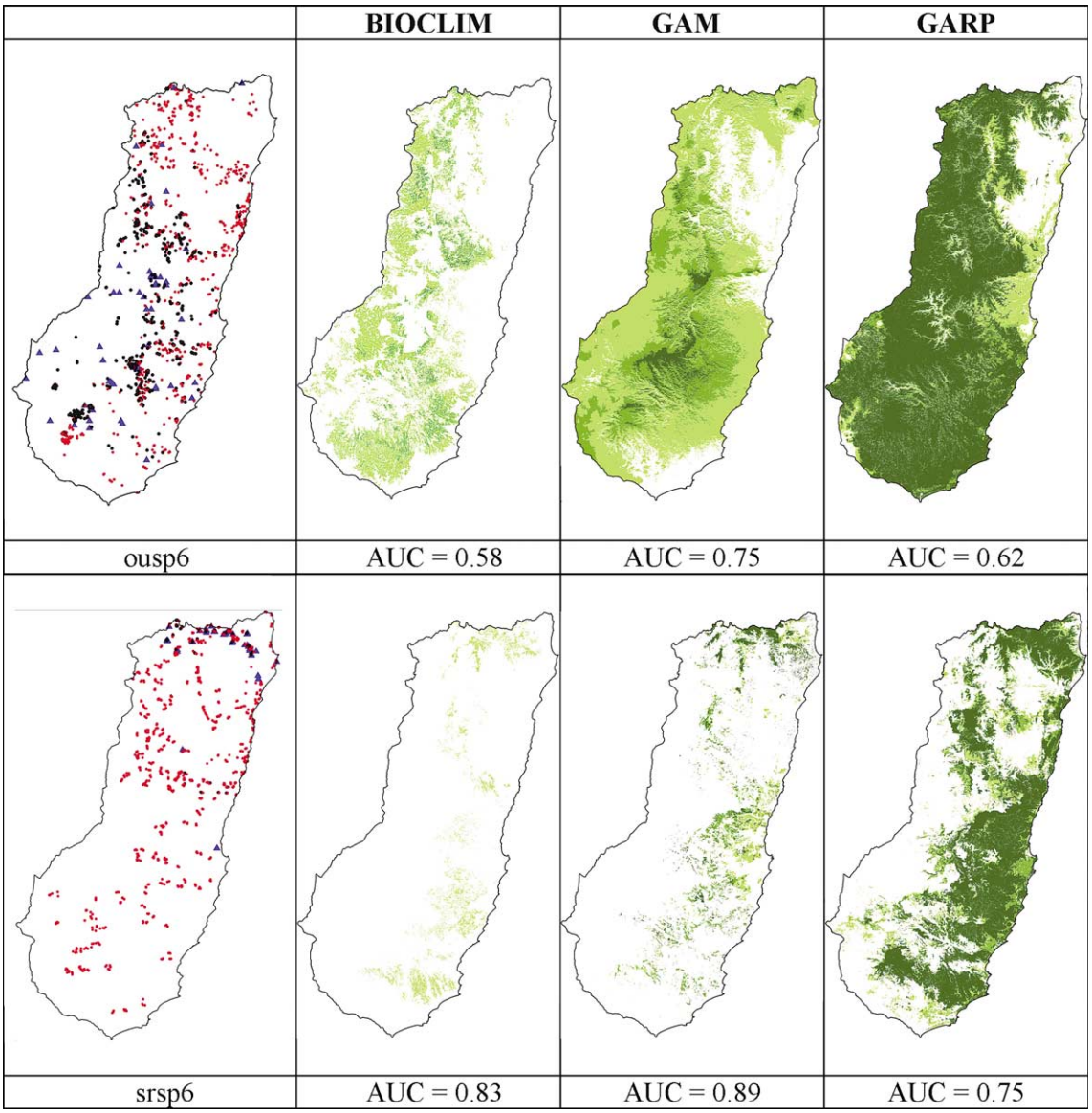

Fig. 1. Maps for two species from NSW for each of three selected techniques. Details: ousp6, *Poa sieberiana* (53 records for modelling and 512 presence/797 absence for evaluation); srsp6 *Ophioscincus truncatus* (79 model, 74/932 eval). The first column shows modelling sites (grey triangles) and evaluation sites: presence = black circle, absence = black cross.

B) Comparison of regional results

Predictive success varied markedly across regions (Fig. 5, Appendix, Figs S3 and S4; Table 7 and Appendix, Table S3). For example, AUC scores were generally high for SWI and SA (mean values of 0.77 and 0.76, Table 7, and details in Appendix, Table S3), intermediate for NSW, NZ and AWT (mean AUC 0.69, 0.71, 0.67 respectively) and mostly poor for CAN (mean AUC 0.58). Regional differences as assessed using COR (Table 7, Appendix, Table S3) and KAPPA (Appendix, Tables S3 and S4) indicated a slightly different ordering of regions: SA and AWT were better modelled than the next group (SWI, NSW, NZ), but again predictions for CAN species were generally poor. The relative success of different methods did not differ greatly across the three test statistics, therefore we focus on AUC statistics for the remainder of this section.

In some regions there were relatively large differences in mean AUC among methods, whereas in others differences were more muted, as shown by the vertical spread of the lines in Fig. 5 (and see Table 7). When data are analysed within regions there is less power to detect differences, so error bars were relatively larger (compare Fig. 3 and Appendix, Fig. S5) and fewer pairwise differences between methods were significant. Our

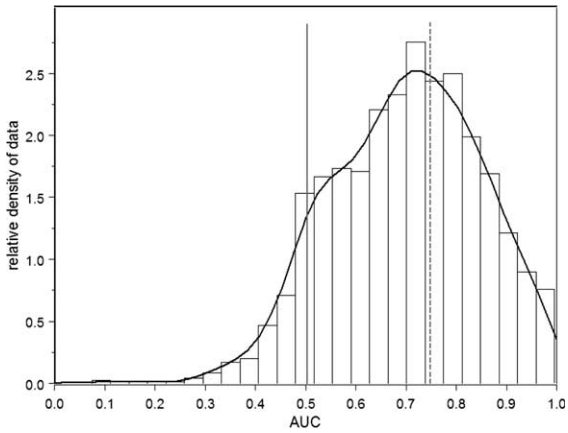

Fig. 2. Distribution of AUC for all species and from all methods. The solid curve is a density plot, and the y axis is scaled to show the relative density of points; for the histogram bars the sum of the area below all bars is 1. The grey vertical solid line shows random predictions, and grey dashed indicates reasonable predictive performance.

ability to clearly distinguish between methods was partly related to the amount of evaluation data; SWI, NZ, and CAN had more data (i.e. number of records per species is to the right of Appendix, Fig. S6, leading to lower standard errors, Appendix, Fig. S7), and more records provide more opportunity to find differences in these regions. For example, Table 8 presents pairwise differences for SA and SWI, and demonstrates more distinctions between methods in SWI. Nevertheless, relative rankings of methods were broadly consistent across regions (Fig. 5, Table 7, and see Appendix, Table S3, Fig. S5) and the group of highest-performing methods identified in Fig. 3 was generally reliable across all regions, but with some variation depending on the evaluation statistic (Appendix, Fig. S5).

Some interesting patterns and exceptions to model performance by region were apparent. Performance of methods in NSW, NZ, SA, and SWI was generally similar, with BRT, MAXENT and MAXENT-T, MARS-COMM, and GDM and GDM-SS usually performing well. In NZ, presence-only methods (BIOCLIM, DOMAIN and LIVES) performed particularly poorly. In most regions DOMAIN and LIVES had lower COR values in relation to AUC than the average method (i.e. they sit below the line fitted to the means in Appendix, Fig. S5).

The importance of the interaction term (method × region) in the GLMM (Table 5) indicated anomalies in the relative performance of methods in particular regions. These tended to occur in regions with lower overall performance – i.e. to the right in Fig. 5 (see also Table 7 and Appendix, S3), and with highest uncertainty in estimates (see standard errors, Appendix, Fig. S5). For example, in AWT, OM-GARP ranked with the better methods (GDM-SS and MAXENT-T), whereas it generally had only intermediate overall performance in all other regions (Fig. 5, Table 7). However, AUC only varied from 0.64 to 0.70 in AWT and many differences were not statistically important (Appendix, Fig. S5). Canada had the lowest AUC, COR and KAPPA scores of any region, and many methods performed poorly, with evaluation statistics only marginally better than random (Table 7 and Appendix, Table S3; Figs S3, S4 and S5). Of the two methods with the highest AUC scores in CAN, one (MARS-COMM) ranked consis-

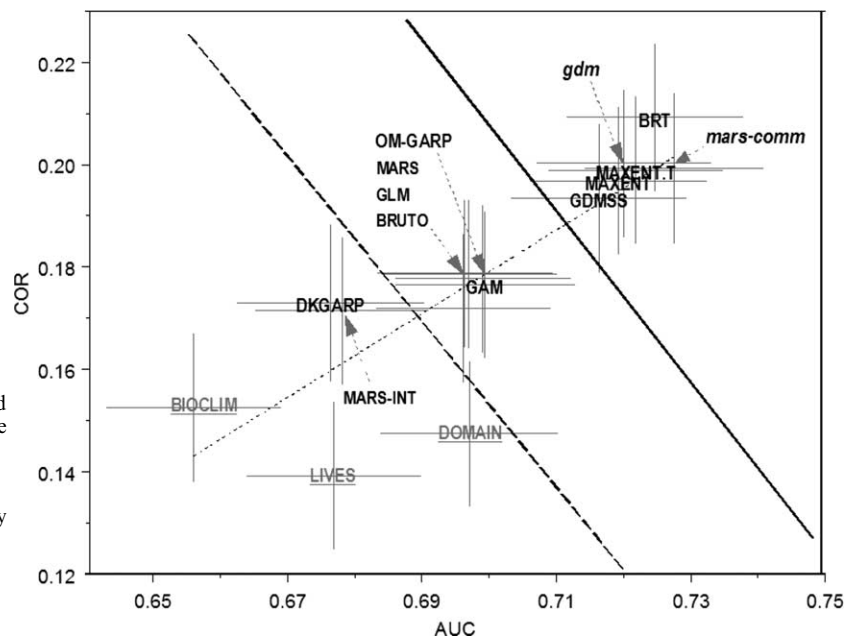

Fig. 3. Mean AUC vs mean correlation (COR) for modelling methods, summarised across all species. The grey bars are standard errors estimated in the GLMM (see Appendix), reflecting variation for an average species in an average region. The labels are broad classifications of the methods: grey underlined = only use presence data, black capitals = use presence and background samples, black lower case italics = community methods.

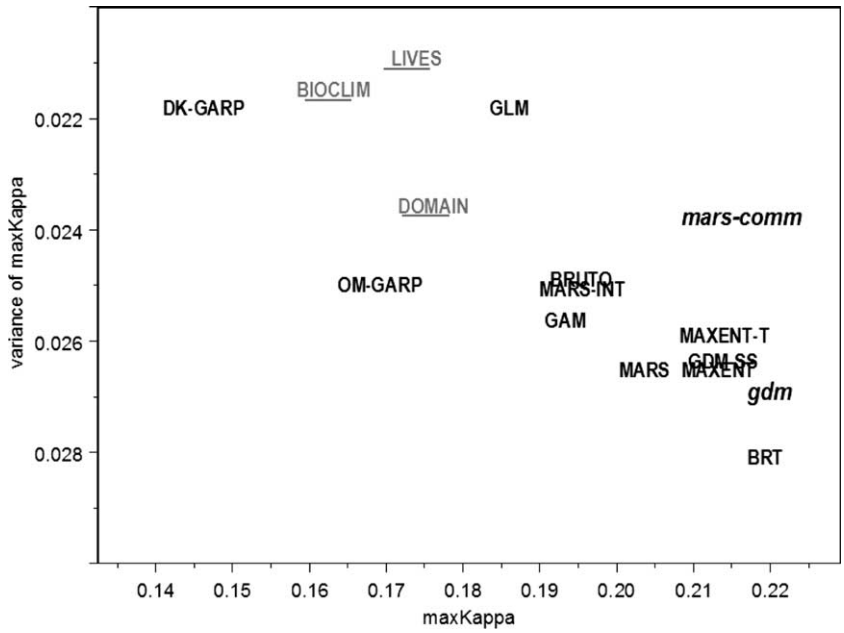

Fig. 4. Performance measured by maximum kappa and its variance across all species. Variance axis reversed so low is higher on plot; it is desirable to have high kappa and low variance (i.e. upper right in plot), for consistent and good performance. Labels as for Fig. 3.

tently among the best across regions, whereas the other (BIOCLIM) tended to be among the lowest.

C) Results at the species level

The greatest variation in the performance of different methods was apparent at the species level, but with similar trends: methods shown above to perform well when averaged across species and regions (i.e. MARS-COMM, BRT, MAXENT/MAXENT-T and GDM/GDM-SS) also tended to perform well when ranked against other methods on an individual species' basis (Table 9 and Appendix, Fig. S2). Figure 6a shows results from SA that are typical for most regions, in that there is marked variation in which methods perform best for different species (i.e. lines cross in the graphs). Most methods occasionally failed badly, although the better methods tended to have more stable performance. The exception to this general pattern was SWI, in which we observed clear and reasonably consistent separation between methods (Fig. 6b), probably reflecting the larger amounts of accurately located data for both modelling and evaluation. Results for SWI using both AUC and COR indicate that the highest-performing methods for most species were MARS-COMM, BRT and MAXENT/MAXENT-T, while the lowest were LIVES, BIOCLIM, and DK-GARP (Appendix, Fig. S5). Across all species in all regions, and considering the best method only, 78% of species had AUC scores of 0.70 or more, and 64% had scores of 0.75 or more (Appendix, Table S3). In some cases a species was modelled well (or poorly) by most methods, whereas in others there was considerable variation in predictive accuracy depending on the method used. Mean AUC scores per species varied from 0.36 to 0.97, with coefficients of variation (cv) ranging from 2 to 47% (Appendix, Table S3). The correlation between mean AUC and cv was weakly negative (Pearson $r = -0.39$), indicating a slight trend for more variation across methods for species with low mean AUC scores.

Predictive performance did not vary consistently with number of presence records available for modelling (Table 10, Fig. 7, Appendix, Table S5). Species that are rarer because they are environmentally or geographically restricted appeared to be modelled with greater accuracy than more common and generalist species (Table 10). We note here and discuss later that this result depends on the spatial extent of analysis and the type of evaluation.

Table 5. Analysis of importance of factors affecting predictive performance, from Generalized Linear Mixed Model.

	DIC[1]
Full model:AUC ~ Method + Method × Region + Species	−8996
Without Method	−8524
Without Interaction (Method × Region)	−8803
Without Species	−4071

[1]Deviance Information Criterion. Changes in DIC >10 are important.

Discussion

The model comparison developed herein is unique for its broad geographic scope, application of numerous modelling methods (including several new techniques) to

Table 6. Probability that the method in the column gives a higher AUC than the method in a row. Low values indicate that the method in the row tends to give higher AUCs than the method in the column. Values outside the arbitrary limits p =(0.025, 0.975) are highlighted in bold; for this two-tailed test.

	BIOCLIM	BRT	BRUTO	DOMAIN	GAM	GLM	DK-GARP	OM-GARP	GDM	GDM-SS	LIVES	MAXENT	MAXENT-T	MARS	MARS-INT	MARS-COMM
BIOCLIM		**0.000**														
BRT	**0.000**															
BRUTO	**0.019**	**0.988**														
DOMAIN	**0.013**	**0.979**	0.422													
GAM	**0.011**	**0.976**	0.413	0.495												
GLM	**0.012**	**0.982**	0.447	0.527	0.534											
DK-GARP	0.200	**0.999**	0.882	0.914	0.916	0.909										
OM-GARP	**0.010**	**0.976**	0.400	0.475	0.487	0.449	0.078									
GDM	**0.000**	0.757	0.061	0.089	0.093	0.076	**0.004**	0.095								
GDM-SS	**0.000**	0.791	0.076	0.107	0.110	0.095	**0.006**	0.119	0.549							
LIVES	0.124	**0.999**	0.832	0.877	0.877	0.864	0.394	0.888	**0.993**	**0.992**						
MAXENT	**0.000**	0.733	0.052	0.075	0.081	0.065	**0.004**	0.085	0.469	0.424	**0.007**					
MAXENT-T	**0.000**	0.658	0.032	0.049	0.050	0.043	**0.002**	0.054	0.381	0.337	**0.003**	0.411				
MARS	**0.017**	**0.986**	0.479	0.552	0.560	0.530	0.111	0.580	0.933	0.918	0.151	0.942	0.964			
MARS-INT	0.169	**0.999**	0.877	0.914	0.916	0.900	0.465	0.924	**0.996**	**0.995**	0.576	**0.997**	**0.998**	0.888		
MARS-COMM	**0.000**	0.425	**0.009**	**0.015**	**0.016**	**0.012**	**0.000**	**0.016**	0.190	0.157	**0.000**	0.213	0.279	**0.010**	**0.000**	

the analysis of presence-only data, and incorporation of extensive presence/absence data to enable well-informed evaluations of predictive performance. We interpret results in the context of the feasibility of using such techniques to predict accurately species' distributions in situations in which only presence data are available. Our evaluation also explores the utility of the vast storehouse of occurrence information in resources such as world natural history museums.

We can draw two major conclusions from our results. First, presence-only data are useful for modelling species' distributions. This result bolsters the recent movement to capitalize on the growing availability of both species' occurrence records (Soberon et al. 2000, Graham et al. 2004a) and high-resolution spatial environmental data. Not all species were predicted well according to our evaluation data, but we found that 64% of the best models for each species had AUCs >0.75 and an additional 14% had AUCs between 0.7 and 0.75. These AUC scores indicate that predictions based on presence-only data can be sufficiently accurate to be used in conservation planning (Pearce and Ferrier 2000a) and in numerous other applications in which estimates of species' distribution are relevant. Second, new modelling methods that have only recently been applied to the challenge of modelling species' distributions generally outperformed established methods. Some of these new methods originated in other disciplines and have had little exposure in ecological analyses. These methods appear to offer considerable promise across a much broader range of ecological applications, providing an exciting avenue for future research. The other strong performers were community methods, and these also deserve further scrutiny, particularly where data for the species in question are sparse.

Broad comparison of methods

We demonstrated differences in predictive performance among modelling methods, despite substantial variation at both regional and species levels. Within the suite of relatively commonly used methods, those that characterise the background environment and that can differentially weight variables outperform those that use presence data alone (BIOCLIM, LIVES, and – for some measures – DOMAIN). These results give no support to using methods that do not attempt to characterise the distribution of a species relative to the background environment in which it occurs. The various regression-based methods are largely indistinguishable from one another in terms of predictive performance. The new version of GARP (OM-GARP), first implemented for this study, is comparable to, but slower than, the regression methods and outperforms the widely used desktop version.

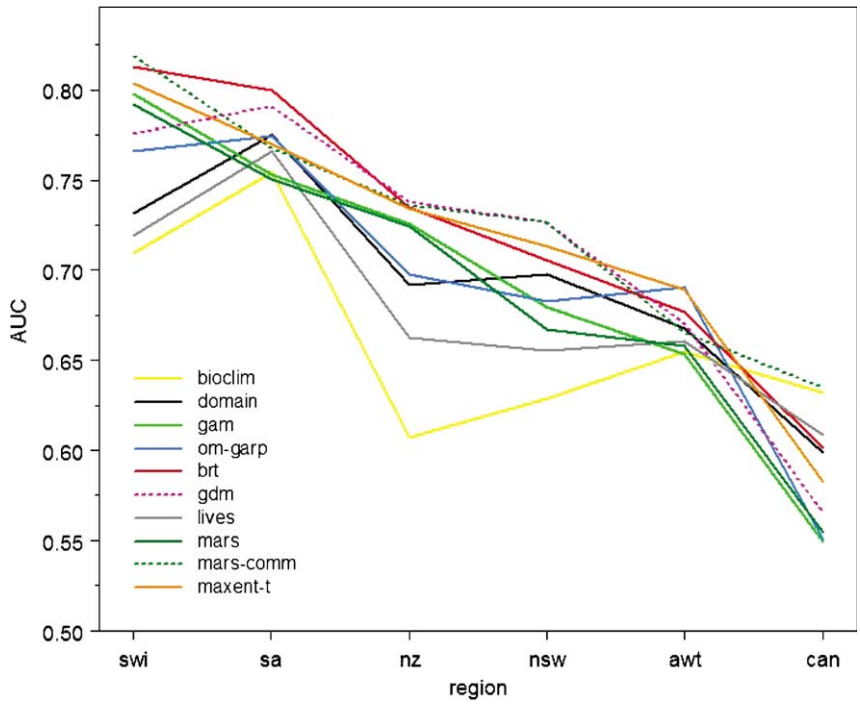

Fig. 5. Predictive success across regions, for 10 methods. Regions are sorted by the mean AUC across all 16 methods and all species per region.

Results for the more common approaches are consistent with previous studies of presence-absence modelling methods, and with the relatively few comparisons of methods used to model presence-only data. Studies of presence-absence modelling methods suggest that several non-linear techniques (e.g. GAMs, artificial neural networks, and MARS) are comparable in terms of predictive ability, and are often superior to methods such as traditional single decision trees (Ferrier and Watson 1997, Elith and Burgman 2002, Moisen and Frescino 2002, Muñoz and Fellicísimo 2004, Segurado and Araujo 2004). Comparisons of methods using presence-only records are less common, but tell a similar story: GLMs and GAMs generally outperform simpler methods (Ferrier and Watson 1997, Brotons et al. 2004), MAXENT outperforms GARP (Phillips et al. 2006) and some presence-only methods (e.g. DOMAIN, ENFA, (Hirzel et al. 2002)) have advantages over BIOCLIM (Loiselle et al. 2003). Most of these studies, however, have focused on single geographic regions and/or smaller numbers of species. Several have been evaluated on the same data sets as were used for model development, and this makes it difficult to generalise results, and to discern whether models have good predictive performance or whether they are simply overfit (Leathwick et al. unpubl.). Our use of independent presence/absence test

Table 7. Regional data: mean AUC and mean COR per method, per region.

Method	Mean AUC						Mean COR					
	AWT	CAN	NSW	NZ	SA	SWI	AWT	CAN	NSW	NZ	SA	SWI
BIOCLIM	0.65	0.63	0.63	0.61	0.75	0.71	0.22	0.08	0.12	0.08	0.29	0.15
BRT	0.68	0.60	0.71	0.73	0.80	0.81	0.24	0.08	0.19	0.18	0.32	0.24
BRUTO	0.64	0.55	0.68	0.72	0.75	0.79	0.20	0.04	0.15	0.17	0.22	0.20
DK-GARP	0.68	0.56	0.66	na[1]	0.75	0.70	0.28	0.05	0.15	na[1]	0.21	0.13
DOMAIN	0.67	0.60	0.70	0.69	0.77	0.73	0.22	0.05	0.15	0.10	0.21	0.14
GAM	0.65	0.55	0.68	0.73	0.75	0.80	0.22	0.04	0.15	0.17	0.23	0.21
GDM	0.67	0.57	0.73	0.74	0.79	0.78	0.24	0.06	0.21	0.16	0.30	0.19
GDM-SS	0.70	0.56	0.70	0.73	0.79	0.79	0.29	0.04	0.17	0.15	0.28	0.20
GLM	0.66	0.57	0.68	0.71	0.74	0.78	0.24	0.06	0.16	0.16	0.22	0.19
LIVES	0.66	0.61	0.66	0.66	0.77	0.72	0.23	0.06	0.12	0.08	0.21	0.13
MARS	0.66	0.55	0.67	0.72	0.75	0.79	0.23	0.04	0.15	0.17	0.24	0.21
MARS-COMM	0.67	0.64	0.73	0.74	0.77	0.82	0.20	0.11	0.19	0.18	0.26	0.26
MARS-INT	0.65	0.54	0.64	0.70	0.73	0.78	0.22	0.05	0.14	0.16	0.24	0.21
MAXENT	0.68	0.58	0.71	0.74	0.78	0.80	0.23	0.05	0.18	0.18	0.27	0.24
MAXENT-T	0.69	0.58	0.71	0.73	0.77	0.80	0.24	0.06	0.18	0.18	0.26	0.25
OM-GARP	0.69	0.55	0.68	0.70	0.77	0.77	0.29	0.04	0.15	0.13	0.25	0.19
mean	0.67	0.58	0.69	0.71	0.76	0.77	0.24	0.06	0.16	0.15	0.25	0.20

[1] DK-GARP could not be run for NZ; the large number of grid cells could not be accomodated by the available computers.

Table 8. Probability that the method in the column gives a higher AUC than the method in a row for (a) SA and (b) SWI.Format follows Table 6.

	BIOCLIM	BRT	BRUTO	DOMAIN	GAM	GLM	DK-GARP	OM-GARP	GDM	GDM-SS	LIVES	MAXENT	MAXENT-T	MARS	MARS-INT	MARS-COMM
(a)																
BIOCLIM																
BRT	**0.002**															
RUTO	0.653	**0.999**														
DOMAIN	0.098	0.940	0.045													
GAM	0.521	**0.998**	0.367	0.912												
GLM	0.740	**1.000**	0.600	0.973	0.723											
DK-GARP	0.649	**0.999**	0.495	0.953	0.628	0.395										
OM-GARP	0.102	0.944	0.049	0.513	0.093	0.028	0.049									
GDM	**0.011**	0.706	**0.003**	0.155	**0.009**	**0.002**	**0.003**	0.147								
GDM-SS	**0.024**	0.814	**0.009**	0.254	**0.022**	**0.004**	**0.010**	0.243	0.638							
LIVES	0.223	**0.982**	0.124	0.706	0.206	0.081	0.126	0.694	0.939	0.886						
MAXENT	0.083	0.931	0.038	0.467	0.074	**0.022**	0.039	0.456	0.826	0.721	0.268					
MAXENT-T	0.160	0.969	0.083	0.618	0.145	0.051	0.085	0.609	0.907	0.834	0.408	0.650				
MARS	0.598	**0.999**	0.444	0.939	0.577	0.347	0.448	0.936	**0.995**	**0.987**	0.844	0.949	0.893			
MARS-INT	0.906	**1.000**	0.824	**0.996**	0.897	0.748	0.825	**0.995**	**1.000**	**1.000**	**0.980**	**0.997**	**0.989**	0.857		
MARS-COMM	0.193	**0.976**	0.104	0.667	0.178	0.065	0.105	0.657	0.927	0.863	0.459	0.696	0.552	0.133	**0.014**	
(b)																
BIOCLIM																
BRT	**0.000**															
BRUTO	**0.000**	**0.995**														
DOMAIN	**0.006**	**1.000**	**1.000**													
GAM	**0.000**	0.954	0.193	**0.000**												
GLM	**0.000**	**0.999**	0.719	**0.000**	0.925											
DK-GARP	0.809	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**										
OM-GARP	**0.000**	**1.000**	**0.996**	**0.000**	**1.000**	**0.981**	**0.000**									
GDM	**0.000**	**1.000**	0.943	**0.000**	**0.993**	0.841	**0.000**	0.140								
GDM-SS	**0.000**	**0.992**	0.439	**0.000**	0.762	0.231	**0.000**	**0.002**	0.041							
LIVES	0.142	**1.000**	**1.000**	0.919	**1.000**	**1.000**	0.026	**1.000**	**1.000**	**1.000**						
MAXENT	**0.000**	0.900	0.102	**0.000**	0.342	0.033	**0.000**	**0.000**	**0.002**	0.132	**0.000**					
MAXENT-T	**0.000**	0.840	0.060	**0.000**	0.243	**0.016**	**0.000**	**0.000**	**0.001**	0.081	**0.000**	0.387				
MARS	**0.000**	**0.988**	0.391	**0.000**	0.719	0.194	**0.000**	**0.002**	0.032	0.448	**0.000**	0.839	0.899			
MARS-INT	**0.000**	**1.000**	0.853	**0.000**	0.971	0.676	**0.000**	0.053	0.295	0.886	**0.000**	**0.989**	**0.995**	0.907		
MARS-COMM	**0.000**	0.251	**0.000**	**0.000**	**0.009**	**0.000**	**0.000**	**0.000**	**0.000**	**0.001**	**0.000**	**0.025**	0.048	**0.002**	0.000	

Table 9. Comparison of performance at a species level and over all species. Rank of 1 = highest to 16 = lowest.

Method (sorted by column 2)	Mean AUC rank per species	Rank of mean AUC over all species
MARS-COMM	6.15	1
BRT	6.20	2
MAXENT-T	6.42	3
MAXENT	6.69	5
GDM-SS	7.38	6
GDM	7.53	4
GAM	8.26	7
GLM	8.64	10
DOMAIN	8.70	9
BRUTO	8.79	12
MARS	8.92	8
OM-GARP	8.92	11
MARS-INT	9.72	13
LIVES	10.22	14
DK-GARP	10.47	15
BIOCLIM	10.85	16

data across multiple geographic regions provides a broader basis for comparisons.

A novel aspect of our work is the inclusion of newer modelling methods that have had little exposure in previous comparative studies and few applications in ecology in general. These novel methods outperform the established methods, and this observation should provoke attention and scrutiny. Several of the novel methods have been developed and tested in fields other than species' distribution modelling, and have been shown to handle noisy data and complex analytical challenges successfully. For example, boosted regression trees have been a focus of attention in the machine-learning and statistical fields for a number of years (Ridgeway 1999, Hastie et al. 2001), but the present paper and a companion application to New Zealand fish (Leathwick et al. in press) are among the first in ecology. Similarly, maximum entropy methods are well known in other fields (Jaynes 1982) but only recently developed for questions of species' distributions (Phillips et al. 2006).

One question of interest is whether our "best" methods share certain characteristics that set them apart from the others? One feature that they all share in common is a high level of flexibility in fitting complex responses. As a consequence they all have what we term here "expressiveness" – a well-developed ability to express or demonstrate the complex relationships that exist in the data. In several methods this includes effective mechanisms for modelling interactions among variables. However, expressiveness needs to be controlled so that models are not overfit, and to that end several methods use "regularization" techniques (Hastie et al. 2001) to achieve balance between complexity and parsimony.

These methods achieve those goals in different ways. BRT achieves expressiveness by combining the strengths of regression trees, namely omission of irrelevant variables and ability to model interactions, with those of boosting, that is, the building of an ensemble of models that approximate the true response surface more accurately than a single model by overcoming the misclassification problems inherent in single tree models. Both the model building procedure (a penalized forward stepwise search) and our cross-validation methods for finding optimal numbers of trees help to control overfitting. The application of maximum entropy methods to distribution modelling was developed specifically for use with presence-only occurrence data (Phillips et al. 2006). In MAXENT, strong focus has been placed on the role of penalty functions (i.e. regularization) in parameter estimation. Regularization has most impact when sample sizes are small, so the MAXENT modellers tuned their regularization in relation to sample size (Phillips et al. 2004). MAXENT can also fit complex functions between response and predictor variables, and can include interaction terms but to a more limited extent than BRT. GDM-SS models are single-species versions of the community-based GDM models. They are parameterized on data for individual species, not including community data. These models are developed in a 2-step fashion, which together achieves controlled expressiveness. The first step operates effectively like a GAM, fitting additive smoothing functions (albeit to dissimilarities rather than raw observations). The second step, a kernel regression, incorporates interactions by modelling distances and densities within a truly multivariate predictor space, with no assumption of additivity. The success of the kernel regression step depends on the first step accurately transforming the predictor space, thereby addressing the "curse of dimensionality" normally associated with kernel regression type techniques (Lowe 1995).

The success of these new methods suggests that predictive performance of some more common methods such as GAMs might be improved substantially if better tradeoffs between expressiveness and complexity could be incorporated into the model fitting process. In the case of GAMs the issue is not whether it can fit complex responses – it can – but whether the model building methods commonly used are optimal for species distribution modelling. One aspect is modelling interactions, which is possible in regression but rarely implemented when many species are being modelled. Another is model selection: alternative techniques already exist, but are rarely implemented in ecology. For example, the "lasso," used for regularization in MAXENT, can be applied to variable selection and coefficient estimation in regression (Hastie et al. 2001), and has been shown to perform better than stepwise selection for determining model complexity (Tibshirani 1996).

Ferrier et al. (2002b) developed and used GDM for reserve selection and survey design. The impetus for using GDM and other community modelling methods (e.g. MARS-COMM) in distribution modelling is based

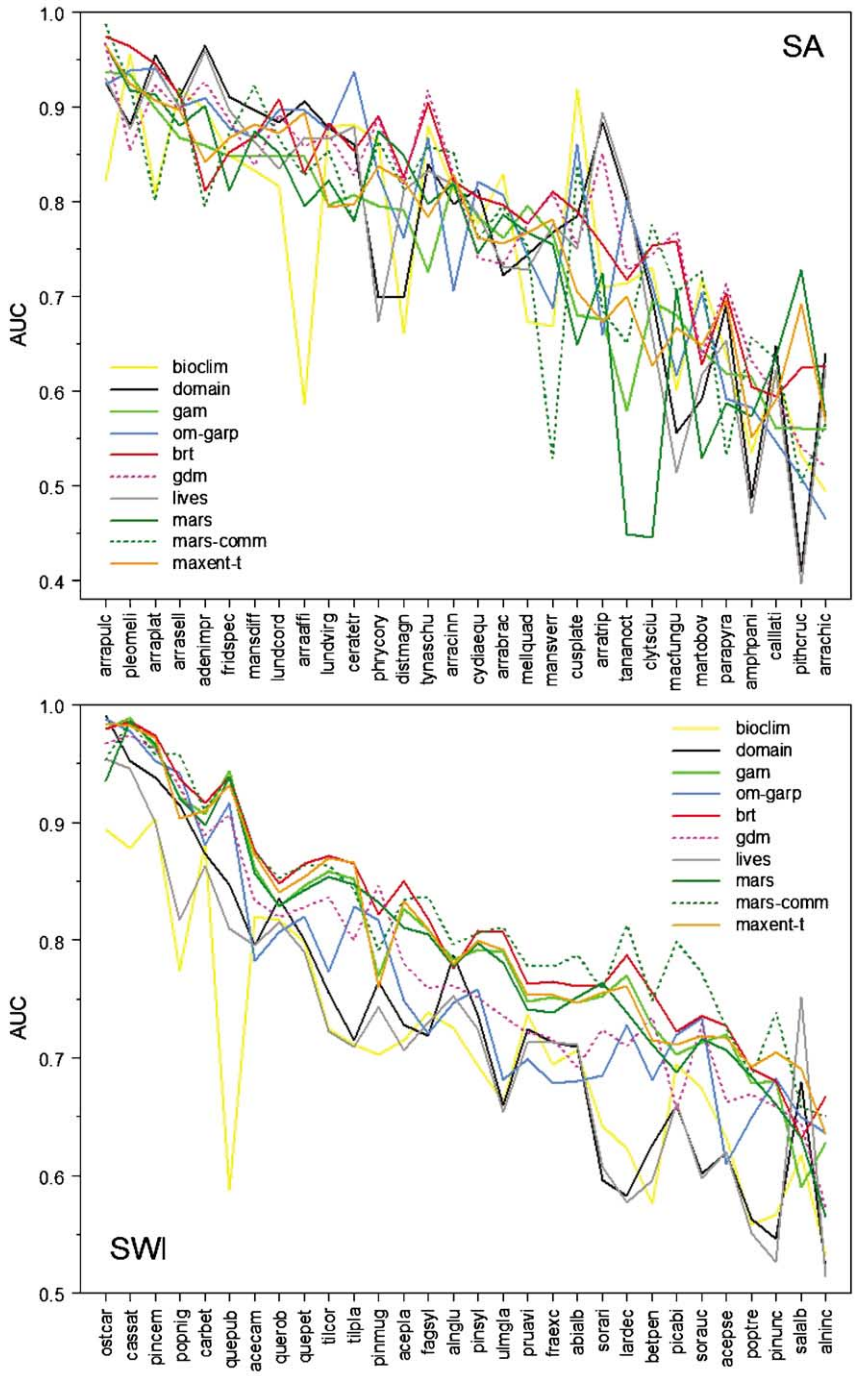

Fig. 6. Predictive performance for the 30 South American species (6a) and the 30 Swiss species (6b), for 10 methods. Species are sorted by the mean AUC across all 16 methods.

on the understanding that important but subtle environmental trends may only be apparent in the response of multiple species, and, more pragmatically, that rare species are difficult to model with standard statistical methods and, hence, community signals may improve model performance. Community models use information from suites of species to inform variable selection and modelling. Relevant predictors are included because of their strong community signal, whereas that signal might be insufficient to trigger inclusion in single species models (Leathwick et al. unpubl.). GDM is the more ecologically sophisticated approach to community modelling, but the simpler MARS-COMM models performed at least as well in this study. The comparison of MARS single species models with the MARS community models shows that the community models

Table 10. Pearson correlation coefficients between data characteristics and the maximum AUC achieved per species, summarised at a regional level.

	Sample sizes		Environmental space		Geographic space	
	P in PO	Prevalence	Nearest neighbour	Range overlap	Relative maximum distance	Relative area of convex polygon
AWT	−0.340*	−0.276	0.589*	−0.302	−0.456*	−0.421*
CAN	0.048*	−0.318	0.621*	−0.589*	−0.313	−0.304
NSW	−0.045	−0.654*	0.675*	−0.770*	−0.230	−0.359*
NZ	0.028	−0.298*	0.713*	−0.669*	−0.491*	−0.300*
SA	−0.601*	−0.714*	0.541*	−0.922*	−0.667*	−0.708*
SWI	−0.186	−0.165	0.619*	−0.520*	−0.534*	−0.620*

Asterisks indicate significant values ($p < 0.05$). Prevalence is the frequency of occurrence records in the PA data. The measures are detailed in the methods.

perform strongly in this trial. In contrast, a recent application of MARS and MARS-COMM to modelling a large presence-absence freshwater fish data set from New Zealand indicated no consistent advantage for community models (Leathwick et al. unpubl.). This difference probably reflects the advantage in using community data to infer absences for presence-only data (see Materials and methods), the importance of the community signal in the more biased and noisier data analyzed here, and the small amounts of data available for fitting some species. Further comparisons of alternative community modelling approaches would be useful to identify the factors influencing success of the various techniques. It is likely that the inference of absences is important, and could be adopted as an alternative to random background samples for single species models.

In summary, we suggest that the good performance of the novel methods result from their ability to fit complex responses (often including interactions) and select a relevant set of variables. We do not expect that the fitted models are too complex – for example, MAXENT as used here is similar in expressiveness to a GLM without interactions (for species with <80 presences) or a GAM with pairwise interactions (Appendix, Text S1). However, for any method an ability to fit complex responses is not useful in itself, and has to be balanced against the requirement for model features to be ecologically realistic (Austin 2002). Systematic assessment of whether individual fitted responses were realistic was beyond the scope of the current research as our aim was to test models across numerous species and regions. Nonetheless, this would be a fruitful avenue for research and model assessment.

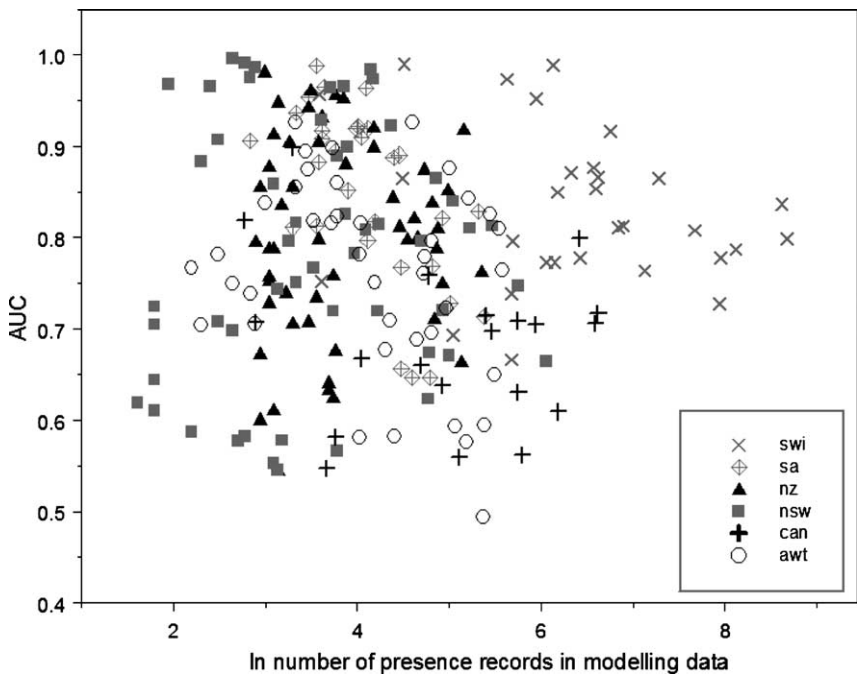

Fig. 7. The maximum AUC (over all methods) vs sample size in modelling data. Each point represents a species.

Regional patterns

The trend for good performance of the novel methods in our experiment (BRT, MAXENT, GDM and MARS-COMM) was also apparent at the regional level. In regions such as SWI, where overall performance was high, the "best" methods performed well, but most methods were able to produce reasonable quality models for many of the species. Exceptions to the usual ranking of methods sometimes appeared in regions in which overall predictive performance was generally poor (AWT and CAN), with no clearly consistent cause. In these regions, the good methods usually still showed moderate success in prediction, and often were not impacted as severely as other methods. The fact that some regions had relatively poor average performance and even the best methods could not do well suggests that in such regions, modelling might most benefit from attempts to improve the training data set. This is particularly apparent in the CAN bird data set, in which the data display many of the biases possible in presence-only data (Anderson 2003). Sampling is severely biased towards the southern extreme of the study area, species' ranges are only partially represented, and environmental data may not capture the important conditions affecting bird distribution, especially those with large ranges. The effects of the partial sampling of ranges may be similar to those demonstrated by Randin et al. (in press) in their study of transferability of models. One would expect the sampling bias to be particularly problematic in cases, like CAN, where the gradient in sampling efforts corresponds to strong environmental gradients. As a consequence, only 4 of 20 species had a mean ROC >0.7, and the best methods (MARS-COMM and BIOCLIM) had mean ROC scores of only 0.63 across species. With poor modelling success, the data and ecology of the species need to be further investigated to identify how to supplement or otherwise improve both the species' occurrence records and the predictor variable set.

Species-level patterns

Though not a primary focus of our analysis, we can highlight two results to guide those needing to understand what species' attributes might affect model performance. Firstly, sample size (i.e. number of occurrence records) was not consistently related to modelling success. While other studies indicate that small sample size negatively influences modelling success (Pearce and Ferrier 2000a, Harrell 2001, Stockwell and Peterson 2002, Kadmon et al. 2003), this was not evident here, perhaps because in general we had adequate localities for modelling (mean number of modelling records $=233$; range $=2-5822$)

The other pattern that emerged is that species judged to be specialists by the distribution of records in environmental or geographic space tended to have higher AUC scores than generalists, an effect also observed elsewhere (Guisan and Hofer 2003, Segurado and Araújo 2004, Thuiller et al. 2004, Luoto et al. 2005). However, this result requires further scrutiny because it may well be largely a function of spatial extent of the analysis. If the extent is fixed (say, to a region) and the evaluation data are a constant set of sites, specialists by definition are those that exist in only a subset of that space. Therefore, the evaluation data set will have many zero records for such species, and any model that can restrict its non-zero predictions to the zone that the species occupies will have a good AUC score, due to the comparison with the many absence sites. The question then becomes one of whether a constant extent of analysis is appropriate for all species in relation to the purpose of the predictions. This consideration needs to be defined by the end-use of the predictions, and merits attention in any evaluation of modelling success.

Limitations

Whereas the present study points to both the potential informativeness of presence-only data and to improved methods for predicting species distributions from them, we also need to emphasize what this doesn't tell us. Our evaluation strategy is specific to the question of how best to predict species distribution under current environmental conditions, which has broad relevance to conservation and ecological or macro-ecological studies (Ferrier 2002, Funk and Richardson 2002, Rushton et al. 2004). Our approach, however, does not inform selection of methods for predicting potential ranges or extrapolation from the current to alternative climates. Modelling of potential ranges (Soberón and Peterson 2005) has applications to predicting expansions of invasive species and investigating speciation processes (Hugall et al. 2002, Peterson 2003). However, as the true potential range may differ from the realized range because of dispersal limitation, competition or other factors (Van Horne 1983, Hanski 1994, Tyre et al. 2001, Anderson et al. 2002), evaluating model performance is a complex task and use of observed absences may be misleading.

Projection of modelled relationships to alternative climates (past or future) is an increasingly popular application of distributional modelling, e.g. in relation to potential effects of global warming. However, strong performance of a particular method in the present climate does not guarantee similar performance under different climates (Thuiller 2004), particularly where this requires prediction outside the range of environments on which the original model was based (Araújo et al. 2005).

Community-based models could fail if patterns of species co-occurrence change due to idiosyncratic responses by individual species to climate change. Similar caution should also be exercised in making predictions for sites distant from the geographic domain from which the modelling data were drawn, given the potential for accurate prediction to be confounded either by unrecognised environmental factors or by large-scale geographic variation in disturbance regimes (Randin et al. in press). Clearly, different models may be necessary, and different approaches needed to evaluate relative performance of modelling methods under these scenarios. Araújo et al. (2005) argue that whilst evaluation of models to new climates and new regions is particularly difficult, evaluations of models even within one region are prone to overoptimism. This occurs when modelling and evaluation data sets are not sufficiently independent. Because the data in this study were sourced from collections with substantially different survey protocols, the evaluation is unlikely to suffer from such problems. Nevertheless, further explorations of the ability of the novel methods to predict reliably to unsampled areas – for example, by spatially stratifying the evaluation data – would be useful.

A further qualification needs to be made. Whilst these models have been shown to produce predictions that are useful in their ability to rank sites for relative suitability for many species, we do not suggest that models built with presence-only data will be well calibrated. That is, they do not accurately predict probability of presence because they do not have access to reliable information on the prevalence (frequency of occurrence) of the species in the region. Rather, they provide relative indices of suitability. Our implementations of most of the methods that take into account background were simple applications of the methods, not attempting to make sophisticated adjustments at might improve this aspect. Alternative approaches such as case control methods for regression (Keating and Cherry 2004, Pearce and Boyce in press), and Bayesian methods (Gelfand et al. 2006) continue to be developed and have potentially important contributions because they deal more appropriately with the presence-only paradigm.

This is by no means a complete analysis, and important questions remain. To advance our understanding of the strengths and weaknesses of methods and the differences between them requires theoretical investigations of the techniques (Austin 2002), testing on simulated data (Austin et al. 1995), analysis of the modelled response shapes (e.g. Austin et al. 1994, Bio 2000, Leathwick 2002) and evaluation of the spatial trends in errors (Fielding and Bell 1997, Barry and Elith in press). Our method of evaluation relies on the evaluation site data and summarises over these, but this might mask effects that can only be clarified with these more detailed evaluations. Operator expertise will affect model performance, and tests of the sensitivity of the methods to how they are applied would be informative. These were beyond the scope of this paper but are necessary for a deeper understanding of the models and their predictive capacities and limitations.

Finally, we stress that modelling can never provide a complete substitute for detailed, ongoing collection of field data, including data on species' distribution, demography, abundance, and interactions (Guisan and Thuiller 2005). Modelling approaches which attempt to integrate such information include Bayesian approaches (Gelfand et al. 2006), investigation of competitors (Leathwick and Austin 2001, Anderson et al. 2002), and studies of connectivity (Ferrier et al. 2002a, Moilanen et al. 2005). To date, few models have been validated via collection of new data (but see Ferrier and Watson 1997, Elith and Burgman 2002, Raxworthy et al. 2003, Peterson 2005). Likewise, collaborative efforts between modellers and users such as conservation managers are rare (Pielke Jr 2003). Ideally, models should be developed and tested in iterative cycles that take account of the desired uses of model, investigate the ecological rationality of the modelled responses and explore errors in predictions (Burgman et al. 2005, Barry and Elith in press). We hope that our model comparisons will stimulate more research into both modelling methods, further development of efficient user interfaces for the more successful methods, and greater integration among modellers and end-users.

Acknowledgements – Data were kindly provided by organizations for whom a number of authors worked. We also thank Mark and George Peck, Royal Ontario Museum, for access to Ontario nest record data (<http://www.birdsontario.org/onrs/onrsmain.html>); Mike Cadman, Bird Studies Canada, Canadian Wildlife Service of Environment Canada, for access to BBS data; Missouri Botanical Garden, especially Robert Magill and Trisha Consiglio, for access to TROPICOS and Gentry transect databases; T. Wohlgemuth and U. Braendi from WSL Switzerland for access to the NFI and forest plots data, and Andrew Ford, CSIRO Atherton, for AWT PA plant records. Michael McCarthy helped with the GLMM analysis and WinBUGs programming. The comments of Mike Austin and Miguel Araújo improved the manuscript substantially. J. E. was funded by ARC Grant DP0209303. This research was initiated in a working group at the National Center for Ecological Analysis and Synthesis (NCEAS), Santa Barbara, USA: "Testing Alternative Methodologies for Modelling Species' Ecological Niches and Predicting Geographic Distributions", conceived of and led by Peterson and Moritz.

References

Anderson, R. P. 2003. Real vs artefactual absences in species distributions: tests for *Oryzomys albigularis* (Rodentia: Muridae) in Venezuala. – J. Biogeogr. 30: 591–605.

Anderson, R. P., Peterson, A. T. and Gómez-Laverde, M. 2002. Using niche-based GIS modeling to test geographic predictions of competitive exclusion and competitive release in South American pocket mice. – Oikos 98: 3–16.

Araújo, M. B. and Williams, P. H. 2000. Selecting areas for species persistence using occurrence data. – Biol. Conserv. 96: 331–345.
Araújo, M. B. et al. 2004. Would climate change drive species out of reserves? An assessment of existing reserve-selection methods. – Global Change Biol. 10: 1618–1626.
Araújo, M. B. et al. 2005. Validation of species-climate impact models under climate change. – Global Change Biol. 11: 1504–1513.
Austin, M. P. 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. – Ecol. Modell. 157: 101–118.
Austin, M. P. and Cunningham, R. B. 1981. Observational analysis of environmental gradients. – Proc. Ecol. Soc. Aust. 11: 109–119.
Austin, M. P. et al. 1994. Determining species response functions to an environmental gradient by means of a beta-function. – J. Veg. Sci. 5: 215–228.
Austin, M. P. et al. 1995. Modelling of landscape patterns and processes using biological data. Subproject 5: simulated data case study. – In: Division of Wildlife and Ecology, CSIRO.
Bakkenes, M. et al. 2002. Assessing effects of forecasted climate change on the diversity and distribution of European higher plants for 2050. – Global Change Biol. 8: 390–407.
Barry, S. C. and Elith, J. in press. Error and uncertainty in habitat models. – J. Appl. Ecol.
Bio, A. M. F. 2000. Does vegetation suit our models? Data and model assumptions and the assessment of species distribution in space. – Fac. of Geographical Sciences, Utrecht Univ., Netherlands, Ph.D. thesis.
Bojórquez, L. I. et al. 1995. Identifying conservation priorities in México through GIS and modeling. – Ecol. Appl. 5: 215–231.
Boyce, M. S. et al. 2002. Evaluating resource selection functions. – Ecol. Modell. 157: 281–300.
Brotons, L. et al. 2004. Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. – Ecography 27: 437–448.
Brown, J. and Lomolino, M. 1998. Biogeography. – Sinauer.
Burgman, M., Lindenmayer, D. B. and Elith, J. 2005. Managing landscapes for conservation under uncertainty. – Ecology 86: 2007–2017.
Burnham, K. P. and Anderson, D. R. 2002. Model selection and inference: a practical information – theoretic approach, 2nd ed. – Springer.
Busby, J. R. 1991. BIOCLIM – a bioclimate analysis and prediction system. – In: Margules, C. R. and Austin, M. P. (eds), Nature conservation: cost effective biological surveys and data analysis. CSIRO, pp. 64–68.
Carpenter, G., Gillison, A. N. and Winter, J. 1993. DOMAIN: a flexible modelling procedure for mapping potential distributions of plants and animals. – Biodiv. Conserv. 2: 667–680.
Cawsey, E. M., Austin, M. P. and Baker, B. L. 2002. Regional vegetation mapping in Australia: a case study in the practical use of statistical modelling. – Biodiv. Conserv. 11: 2239–2274.
Cicero, C. 2004. Barriers to sympatry between avian sibling species (Paridae: Beolophus) in tenuous secondary contact. – Evolution 58: 1573–1587.
Cohen, J. 1960. A coefficient of agreement for nominal scales. – Educ. Psyschol. Meas. 20: 37–46.
Elith, J. and Burgman, M. A. 2002. Predictions and their validation: rare plants in the Central Highlands, Victoria, Australia. – In: Scott, J. M. et al. (eds), Predicting species occurrences: issues of accuracy and scale. Island Press, pp. 303–314.
Ferrier, S. 2002. Mapping spatial pattern in biodiversity for regional conservation planning: where to from here? – Syst. Biol. 51: 331–363.
Ferrier, S. and Watson, G. 1997. An evaluation of the effectiveness of environmental surrogates and modelling techniques in predicting the distribution of biological diversity. – Environment Australia, Canberra, <http://www.deh.gov.au/biodiversity/publications/technical/surrogates/>.
Ferrier, S. et al. 2002a. Extended statistical approaches to modelling spatial pattern in biodiversity: the north-east New South Wales experience. I. Species-level modelling. – Biodiv. Conserv. 11: 2275–2307.
Ferrier, S. et al. 2002b. Extended statistical approaches to modelling spatial pattern in biodiversity: the north-east New South Wales experience. II. Community-level modelling. – Biodiv. Conserv. 11: 2309–2338.
Fielding, A. H. and Bell, J. F. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. – Environ. Conserv. 24: 38–49.
Friedman, J. H., Hastie, T. and Tibshirani, R. 2000. Additive logistic regression: a statistical view of boosting. – Ann. Stat. 28: 337–407.
Funk, V. and Richardson, K. 2002. Systematic data in biodiversity studies: use it or lose it. – Syst. Biol. 51: 303–316.
Gelfand, A. E. et al. 2006. Explaining species distribution patterns through hierarchical modeling. – Bayesian Analysis 1: 41–92.
Gómez Pompa, A. and Nevling, L. I. 1970. La Flora de Veracruz. – Anales del Inst. de Biología de la UNAM, Bot. 31: 137–171.
Goolsby, J. A. 2004. Potential distribution of the invasive old world climbing, fern, *Lygodium microphyllum* in north and south America. – Nat. Areas J. 24: 351–353.
Graham, C. H. et al. 2004a. New developments in museum-based informatics and applications in biodiversity analysis. – Trends Ecol. Evol. 19: 497–503.
Graham, C. H. et al. 2004b. Integrating phylogenetics and environmental niche models to explore speciation mechanisms in dendrobatid frogs. – Evolution 58: 1781–1793.
Graham, C. H., Moritz, C. and Williams, S. E. 2006. Habitat history improves prediction of biodiversity in a rainforest fauna. – Proc. Natl. Acad. Sci. USA 103: 632–636.
Guisan, A. and Zimmerman, N. E. 2000. Predictive habitat distribution models in ecology. – Ecol. Modell. 135: 147–186.
Guisan, A. and Hofer, U. 2003. Predicting reptile distributions at the mesoscale: relation to climate and topography. – J. Biogeogr. 30: 1233–1243.
Guisan, A. and Thuiller, W. 2005. Predicting species distribution: offering more than simple habitat models. – Ecol. Lett. 8: 993–1009.
Guisan, A., Theurillat, J. P. and Kienast, F. 1998. Predicting the potential distribution of plant species in an alpine environment. – J. Veg. Sci. 9: 65–74.
Hanley, J. A. and McNeil, B. J. 1982. The meaning and use of the area under a Receiver Operating Characteristic (ROC) curve. – Radiology 143: 29–36.
Hanski, I. 1994. Patch occupancy dynamics in fragmented landscapes. – Trends Ecol. Evol. 9: 131–134.
Harrell, F. E. 2001. Regression modeling strategies with applications to linear models, logistic regression and survival analysis. – Springer.
Hastie, T., Tibshirani, R. and Friedman, J. H. 2001. The elements of statistical learning: data mining, inference, and prediction. – Springer.
Hijmans, R. J. et al. 2000. Assessing the geographic representativeness of genebank collections: the case of Bolivian wild potatoes. – Conserv. Biol. 14: 1755–1765.
Hijmans, R. J. et al. 2005. Very high resolution interpolated climate surfaces for global land areas. – Int. J. Climatol. 25: 1965–1978.
Hirzel, A. H. and Guisan, A. 2002. Which is the optimal sampling strategy for habitat suitability modelling? – Ecol. Modell. 157: 331–341.
Hirzel, A. H. et al. 2002. Ecological-niche factor analysis: how to compute habitat- suitability maps without absence data? – Ecology 83: 2027–2036.

Huettmann, F. 2005. Databases and science-based management in the context of wildlife and habitat: towards a certified ISO standard for objective decision-making for the global community by using the internet. – J. Wildl. Manage. 69: 466–472.
Hugall, A. et al. 2002. Reconciling paleodistribution models and comparative phylogeography in the Wet Tropics rainforest land snail *Gnarosophia bellendenkerensis* (Brazier 1875). – Proc. Natl. Acad. Sci. USA 99: 6112–6117.
Jaynes, E. T. 1982. On the rationale of maximum entropy methods. – Proc. the IEEE 70: 939–952.
Kadmon, R., Farber, O. and Danin, A. 2003. A systematic analysis of factors affecting the performance of climatic envelope models. – Ecol. Appl. 13: 853–867.
Kadmon, R., Farber, O. and Danin, A. 2004. Effect of roadside bias on the accuracy of predictive maps produced by bioclimatic models. – Ecol. Appl. 14: 401–413.
Keating, K. A. and Cherry, S. 2004. Use and interpretation of logistic regression in habitat selection studies. – J. Wildl. Manage. 68: 774–789.
Leathwick, J. R. 2002. Intra-generic competition among *Nothofagus* in New Zealand's primary indigenous forests. – Biodiv. Conserv. 11: 2177–2187.
Leathwick, J. R. and Austin, M. P. 2001. Competitive interactions between tree species in New Zealand's old-growth indigenous forests. – Ecology 82: 2560–2573.
Leathwick, J. R. et al. 2005. Using multivariate adaptive regression splines to predict the distributions of New Zealand's freshwater diadromous fish. – Freshwater Biol. 50: 2034–2052.
Leathwick, J. R. et al. in press. Variation in demersal fish species richness in the oceans surrounding New Zealand: an analysis using boosted regression trees. – Mar. Ecol. Progr. Ser.
Liu, C. et al. 2005. Selecting thresholds of occurrence in the prediction of species distributions. – Ecography 28: 385–393.
Loiselle, B. A. et al. 2003. Avoiding pitfalls of using species distribution models in conservation planning. – Conserv. Biol. 17: 1591–1600.
Lowe, D. G. 1995. Similarity metric learning for a variable-kernel classifier. – Neural Comput 7: 72–85.
Luoto, M. et al. 2005. Uncertainty of bioclimate envelope models based on geographical distribution of species. – Global Ecol. Biogeogr. 14: 575–584.
Mac Nally, R. and Fleishman, E. 2004. A successful predictive model of species richness based on indicator species. – Conserv. Biol. 18: 646–654.
McCarthy, M. A. and Masters, P. 2005. Profiting from prior information in Bayesian analyses of ecological data. – J. Appl. Ecol. 42: 1012–1019.
Moilanen, A. et al. 2005. Prioritizing multiple-use landscapes for conservation: methods for large multi-species planning problems. – Proc. R. Soc. B 272: 1885–1891.
Moisen, G. G. and Frescino, T. S. 2002. Comparing five modeling techniques for predicting forest characteristics. – Ecol. Modell. 157: 209–225.
Muñoz, J. and Fellicísimo, Á. M. 2004. Comparison of statistical methods commonly used in predictive modeling. – J. Veg. Sci. 15: 285–292.
Murphy, A. H. and Winkler, R. L. 1992. Diagnostic verification of probability forecasts. – Int. J. Forecasting 7: 435–455.
Pearce, J. and Ferrier, S. 2000a. Evaluating the predictive performance of habitat models developed using logistic regression. – Ecol. Modell. 133: 225–245.
Pearce, J. and Ferrier, S. 2000b. An evaluation of alternative algorithms for fitting species distribution models using logistic regression. – Ecol. Modell. 128: 127–147.
Pearce, J. L. and Boyce, M. S. in press. Modelling distribution and abundance with presence-only data. – J. Appl. Ecol.
Pearce, J., Ferrier, S. and Scotts, D. 2001. An evaluation of the predictive performance of distributional models for flora and fauna in north-east New South Wales. – J. Environ. Mange. 62: 171–184.
Peterson, A. T. 2003. Predicting the geography of species' invasions via ecological niche modeling. – Quart. Rev. Biol. 78: 419–433.
Peterson, A. T. 2005. Kansas gap analysis: the importance of validating distributional models before using them. – Southwest. Nat, pp. 230–236
Peterson, A. T., Martinez-Meyer, E. and Gonzalez-Salazar, C. 2004. Reconstructing the pleistocene geography of the Aphelocoma jays (Corvidae). – Biodiv. Res. 10: 237–246.
Phillips, S. J., Dudik, M. and Schapire, R. E. 2004. A maximum entropy approach to species distribution modeling. – In: Proc. of the 21st International Conference on Machine Learning, Banff, Canada, 2004.
Phillips, S. J., Anderson, R. P. and Schapire, R. E. 2006. Maximum entropy modeling of species geographic distributions. – Ecol. Modell. 190: 231–259.
Pielke Jr, R. A. 2003. The role of models in prediction for decision. – In: Canham, C., Cole, J. and Lauenroth, W. K. (eds), Models in ecosystem science. Princeton Univ. Press, pp. 111–135.
Randin, C. F. et al. in press. Are species distribution models transferable in space? – J. Biogeogr.
Rapoport, E. H. 1982. Aerography. – Permagon Press.
Raxworthy, C. J. et al. 2003. Predicting distributions of known and unknown reptile species in Madagascar. – Nature 426: 837–841.
Reese, G. C. et al. 2005. Factors affecting species distribution predictions: a simulation modeling experiment. – Ecol. Appl. 15: 554–564.
Ricklefs, R. 2004. A comprehensive framework for global patterns in biodiversity. – Ecol. Lett. 7: 1–15.
Ridgeway, G. 1999. The state of boosting. – Comput. Sci. Stat. 31: 172–181.
Rosenzweig, M. 1995. Species diversity in space and time. – Cambridge Univ. Press.
Rushton, S. P., Ormerod, S. J. and Kerby, G. 2004. New paradigms for modelling species distributions? – J. Appl. Ecol. 41: 193–200.
Scotts, D. and Drielsma, M. 2003. Developing landscape frameworks for regional conservation planning: an approach integrating fauna spatial distributions and ecological principles. – Pac. Conserv. Biol. 8: 235–254.
Segurado, P. and Araújo, M. B. 2004. An evaluation of methods for modelling species distributions. – J. Biogeogr. 31: 1555–1568.
Silverman, B. W. 1986. Density estimation for statistics and data analysis. – Chapman and Hall.
Skov, F. and Svenning, J. C. 2004. Potential impact of climatic change on the distribution of forest herbs in Europe. – Ecography 27: 366–380.
Sneath, P. H. A. and Sokal, R. R. 1973. Numerical taxonomy – the principles and practice of numerical classification. – W. H. Freeman.
Soberon, J., Llorente, J. B. and Onate, L. 2000. The use of specimen – label databases for conservation purposes: an example using Mexican papilionid and pierid butterflies. – Biodiv. Conserv. 9: 1441–1466.
Soberón, J. M. and Peterson, A. T. 2005. Interpretation of models of fundamental ecological niches and species' distributional areas. – Biodiv. Inform. 2: 1–10.
Spiegelhalter, D. et al. 2003a. WinBUGS user manual, version 1.4. – MRC Biostatistics Unit, Cambridge, UK.
Spiegelhalter, D. J. et al. 2003b. Bayesian measures of model complexity and fit. – J. R. Stat. Soc. Ser. B 64: 583–639.
Stockwell, D. and Peters, D. 1999. The GARP modelling system: problems and solutions to automated spatial prediction. – Int. J. Geogr. Inform. Sci. 13: 143–158.
Stockwell, D. R. B. and Peterson, A. T. 2002. Effects of sample size on accuracy of species distribution models. – Ecol. Modell. 148: 1–13.

Thomas, C. D. et al. 2004. Extinction risk from climate change. – Nature 427: 145–148.
Thornton, P. E., Running, S. W. and White, M. A. 1997. Generating surfaces of daily meteorological variables over large regions of complex terrain. – J. Hydrol. 190: 214–251.
Thuiller, W. 2004. Patterns and uncertainties of species' range shifts under climate change. – Global Change Biol. 10: 2020–2027.
Thuiller, W. et al. 2004. Relating plant traits and species distributions along bioclimatic gradients for 88 *Leucadendron* species in the Cape Floristic Region. – Ecology 85: 1688–1699.
Thuiller, W. et al. 2005. Climate change threats to plant diversity in Europe. – Proc. Natl. Acad. Sci. USA 102: 8245–8250.
Tibshirani, R. 1996. Regression shrinkage and selection via the lasso. – J. R. Stat. Soc. Ser. B 58: 267–288.
Turner, W. et al. 2003. Remote sensing for biodiversity science and conservation. – Trends Ecol. Evol. 18: 306–314.
Tyre, A. J., Possingham, H. P. and Lindenmayer, D. B. 2001. Matching observed pattern with ecological process: can territory occupancy provide information about life history parameters? – Ecol. Appl. 11: 1722–1738.
Van Horne, B. 1983. Density as a misleading indicator of habitat quality. – J. Wildl. Manage. 47: 893–901.
Venier, L. A. et al. 2001. Models of large-scale breeding-bird distribution as a function of macro-climate in Ontario, Canada. – J. Biogeogr. 26: 315–328.
Walker, P. A. and Cocks, K. D. 1991. HABITAT: a procedure for modelling a disjoint environmental envelope for a plant or animal species. – Global Ecol. Biogeogr. Lett. 1: 108–118.
Wintle, B. A. and Bardos, D. C. in press. Modelling species habitat relationships with spatially autocorrelated observation data. – Ecol. Appl.
Yee, T. W. and Mitchell, N. D. 1991. Generalized additive models in plant ecology. – J. Veg. Sci. 2: 587–602.
Zaniewski, A. E., Lehmann, A. and Overton, J. M. 2002. Predicting species distribution using presence-only data: a case study of native New Zealand ferns. – Ecol. Modell. 157: 261–280.
Zheng, B. and Agresti, A. 2000. Summarizing the predictive power of a generalized linear model. – Stat. Med. 19: 1771–1781.

Download the appendix as file E4596 from
<www.oikos.ekol.lu.se/appendix>

Subject Editor: Miguel Araujo.