Selecting thresholds of occurrence in the prediction of species distributions

Canran Liu, Pam M. Berry, Terence P. Dawson and Richard G. Pearson

Transforming the results of species distribution modelling from probabilities of or suitabilities for species occurrence to presences/absences needs a specific threshold. Even though there are many approaches to determining thresholds, there is no comparative study. In this paper, twelve approaches were compared using two species in Europe and artificial neural networks, and the modelling results were assessed using four indices: sensitivity, specificity, overall prediction success and Cohen's kappa statistic. The results show that prevalence approach, average predicted probability/suitability approach, and three sensitivity-specificity-combined approaches, including sensitivity-specificity sum maximization approach, sensitivity-specificity equality approach and the approach based on the shortest distance to the top-left corner (0,1) in ROC plot, are the good ones. The commonly used kappa maximization approach is not as good as the afore-mentioned ones, and the fixed threshold approach is the worst one. We also recommend using datasets with prevalence of 50% to build models if possible since most optimization criteria might be satisfied or nearly satisfied at the same time, and therefore it's easier to find optimal thresholds in this situation.

C. Liu (cliu@une.edu.au), P. M. Berry, T. P. Dawson and R. G. Pearson, Environmental Change Inst., Centre for the Environment, Univ. of Oxford, Dyson Perrins Building, South Parks Road, Oxford, UK OX1 3QY. (present address of C. L.: Dept of Ecosystem Management, School of Environmental Sciences & Natural Resources Management, Univ. of New England, Armidale, NSW 2351, Australia.)

Predicting species distributions is becoming increasingly important since it is relevant to resource assessment, environmental conservation and biodiversity management (Fielding and Bell 1997, Manel et al. 1999, Austin 2002, D'heygere et al. 2003). Many modeling techniques have been used for this purpose, e.g. generalized linear models (GLM), generalized additive models (GAM), classification and regression trees (CARTs), principal components analysis (PCA), artificial neural networks (ANNs) (Guisan and Zimmermann 2000, Moisen and Frescino 2002, Guisan et al. 2002, Berg et al. 2004). And most of the techniques give the results as the probability of species presence, e.g. GLM, GAM and some algorithms of ANNs, or

environmental suitability for the target species, e.g. PCA (Robertson et al. 2003) and some algorithms of ANNs. However, in conservation and environmental management practice, the information presented as species presence/absence may be more practical than presented as probability or suitability. Therefore, a threshold is needed to transform the probability or suitability data to presence/absence data. A threshold is also needed when assessing model performance using the indices derived from the confusion matrix (Manel et al. 2001), which also facilitates the interpretation of modelling results. Before reviewing threshold determination approaches, we will review these model assessment indices first because some of these

indices are also the only or primary component of some threshold determination approaches.

Model assessment indices

Many indices can be used in the assessment of the predictions of species distributions, including sensitivity, specificity, overall prediction success (OPS), Kohen's kappa statistic, the odds ratio, and the normalized mutual information statistic (NMI). And some of them have been incorporated into the approaches to determining thresholds. Fielding and Bell (1997) gave a comprehensive review (Manel et al. 2001). All these indices (Table 1) need the information from the confusion matrix, which consists of four elements: true positive or presence (a), false positive or presence (b), false negative or absence (c) and true negative or absence (d). Since the value of an individual element in the confusion matrix may take zero, the odds ratio and NMI cannot be calculated in some cases. Precision, recall and F are three indices used in the field of information retrieval. Precision is the proportion of the retrieved items that are relevant, i.e. the proportion of predicted presences that are real presences, recall is the proportion of the relevant items that are retrieved, which is equal to sensitivity, and F is the harmonic average of precision and recall (Nahm and Mooney 2000). F varies from 0, when almost no relevant items are retrieved, i.e. almost no real presences are predicted as presences, to 1, when all and only the relevant items are retrieved, i.e. all and only the real presences are predicted as presences. α is a parameter, which gives weights (α and $1-\alpha$) to the two components of F. Moreover, when $\alpha = 0.5$, F is strongly towards the lower of the two values (precision and recall); therefore, this measure can only be high when both precision and recall are high.

Kappa and OPS are two widely used indices (Guisan et al. 1999, Manel et al. 1999, Hilbert and Ostendorf 2001, Luck 2002, Moisen and Frescino 2002). It should be noted that OPS can be deceptively high when frequencies of zeros and ones in binary data are very different (Fielding and Bell 1997, Pearce and Ferrier 2000, Moisen and Frescino 2002). However, Kappa measures the proportion of correctly predicted sites after the probability of chance agreement has been removed (Moisen and Frescino 2002).

Threshold determination approaches

There are many approaches to determining thresholds, which fall into two categories: subjective and objective. A representative in the first category is taking 0.5 as the threshold, which is widely used in ecology (Manel et al. 1999, 2001, Luck 2002, Stockwell and Peterson 2002, Bailey et al. 2002, Woolf et al. 2002). Sometimes 0.3 (Robertson et al. 2001) and 0.05 (Cumming 2000) are also used as thresholds. These choices are very arbitrary and lack any ecological basis (Osborne et al. 2001). Sometimes, a specific level, e.g. 95%, of sensitivity or specificity is desired or deemed acceptable, and it is predetermined (Cantor et al. 1999). Thus, the corresponding threshold can be found. This approach is also subjective because a specific level for some attribute (e.g. sensitivity or specificity, etc.) is predetermined by the researchers.

There are many objective approaches. With these approaches, thresholds are chosen to maximize the agreement between observed and predicted distributions. Cramer (2003) also realized the problem with fixed threshold approach, especially taking 0.5 as the threshold. He stated that with unbalanced samples, this gives nonsense results. Therefore, the sample frequency, i.e. the prevalence of species occurrence (which is defined as the proportion of species occurrences among all the sites), and the mean value of the predicted probabilities of species presence were recommended as the threshold.

Table 1. Indices for assessing the predictive performance of species distribution models, a is true positives (or presences), b is false positives (or presences), c is false negatives (or absences), d is true negatives (or absences), n ($=a+b+c+d$) is the total number of sites and α is a parameter between 0 and 1 (inclusive).

Index	Formula
Sensitivity (or Recall, R)	$a/(a+c)$
Specificity	$d/(b+d)$
Precision (P)	$a/(a+b)$
Overall prediction success (OPS)	$(a+d)/n$
Kappa	$\dfrac{(a+d) - [(a+c)(a+b) + (b+d)(c+d)]/n}{n - [(a+c)(a+b) + (b+d)(c+d)]/n}$
Odds ratio	$(ad)/(cb)$
Normalized mutual information statistic (NMI)	$\dfrac{-a\ln a - b\ln b - c\ln c - d\ln d + (a+b)\ln(a+b) + (c+d)\ln(c+d)}{n\ln n - [(a+c)\ln(a+c) + (b+d)\ln(b+d)]}$
F	$\dfrac{1}{\alpha/P + (1-\alpha)/R} \quad (0 \leq \alpha \leq 1)$

Fielding and Haworth (1995) used a threshold that was calculated as the mid-point between the mean probabilities of occupancy for the present and absent groups.

For other objective approaches, usually, either a specific index, e.g. kappa, or the trade-off between two conflicting properties, e.g. sensitivity and specificity, is optimized in various ways. Kappa maximization approach is popular in ecology (Huntley et al. 1995, Lehmann 1998, Guisan et al. 1998, Collingham et al. 2000, Berry et al. 2001, Pearson et al. 2002). Similarly, OPS and F can also be used in the determination of thresholds (Shapire et al. 1998). The sum of sensitivity and specificity can be maximized to give the threshold (Manel et al. 2001), which is equivalent to finding a point on the ROC (receiver operating characteristics) curve (i.e. sensitivity against 1-specificity) whose tangent slope is equal to 1 (Cantor et al. 1999). The point at which sensitivity and specificity are equal can also be chosen to determine the threshold (Cantor et al. 1999). This approach can also be applied to precision and recall (Shapire et al. 1998). Another approach is to select the point on the ROC curve that is closest to the upper-left corner $(0,1)$ in the ROC plot since the point in this corner represents a perfect classification with 100% sensitivity and specificity (Cantor et al. 1999). Similarly, the point on the P-R (i.e. precision-recall) curve that is closest to the upper-right corner $(1,1)$ in the P-R plot can also be used to determine the threshold since the point in this corner represents a perfect classification with 100% precision and recall.

Some researchers went further to identify the appropriate threshold by incorporating the relative cost of FP (false positive) and FN (false negative) errors and prevalence (Zweig and Campbell 1993, Fielding and Bell 1997) or by incorporating the C/B ratio (the ratio of net FP cost and net true positive benefit) and prevalence (Metz 1978, Cantor et al. 1999). The threshold is corresponding to the point on the ROC curve at which the slope of the tangent is $(C/B) \times (1-p)/p$ or $(FPC/FNC) \times (1-p)/p$, where p is the prevalence (of species' presence) and FPC and FNC are the cost of false positive and false negative respectively.

Although there are so many approaches to determining the threshold, there is no comparative study on their

Table 2. Threshold-determining approaches studied in this paper.

Code	Approach	Definition	Reference
Subjective approach			
1	Fixed threshold approach	Taking a fixed value, usually 0.5, as the threshold	Manel et al. (1999), Bailey et al. (2002)
Objective approaches			
Single index-based approaches:			
2	Kappa maximization approach	Kappa statistic is maximized	Huntley et al. (1995), Guisan et al. (1998)
3	OPS maximization approach	Overall prediction success (OPS) is maximized	
Model-building data-only-based approach:			
4	Prevalence approach	Taking the prevalence of model-building data as the threshold	Cramer (2003)
Predicted probability/suitability-based approaches:			
5	Average probability/suitability approach	Taking the average predicted probability/suitability of the model-building data as the threshold	Cramer (2003)
6	Mid-point probability/suitability approach	Mid-point between the average probabilities of or suitabilities for the species' presence for occupied and unoccupied sites	Fielding and Haworth (1995)
Sensitivity and specificity-combined approaches:			
7	Sensitivity-specificity sum maximization approach	The sum of sensitivity and specificity is maximized	Cantor et al. (1999), Manel et al. (2001)
8	Sensitivity-specificity equality approach	The absolute value of the difference between sensitivity and specificity is minimized	Cantor et al. (1999)
9	ROC plot-based approach	The threshold corresponds to the point on ROC curve (sensitivity against 1-specificity) which has the shortest distance to the top-left corner $(0,1)$ in ROC plot	Cantor et al. (1999)
Precision and recall-combined approaches:			
10	Precision-recall break-even point approach	The absolute value of the difference between precision and recall is minimized	Shapire et al. (1998)
11	P-R plot-based approach	The threshold corresponds to the point on P-R (Precision-Recall) curve which has the shortest distance to the top-right corner $(1,1)$ in P-R plot	
12	F maximization approach	The index F is maximized. In this study, $\alpha = 0.5$ is used in F, i.e. there is no preference to precision and recall	Shapire et al. (1998)

behaviours, so we don't know their relative performance. In this paper we compared twelve different approaches to determining thresholds (Table 2), and investigated their behaviours in various situations, i.e. different prevalence for model-building data and test data, using artificial neural networks, which have been recognized by many researchers, e.g. Ozesmi and Ozesmi (1999), Brosse et al. (1999), Manel et al. (1999), Olden and Jackson (2001), Berry et al. (2002), Pearson et al. (2002, 2004) and Olden (2003), as a modeling technique better than other traditional techniques in modeling complex phenomena with non-linear relationships. We realized that the probability-based approaches were used for predicted probabilities, and our modeling result is predicted suitability for species presence. However, we believe this will not hinder our effort to use these approaches since the "suitabilities" we get are ranged from 0 to 1.

Materials and methods

Species and environmental data

Two species, *Fagus sylvatica* (beech) and *Puccinellia maritima* (common salt marsh grass), with differing European distributions were used in this study. *Fagus sylvatica* is widespread in Europe and extends northwards to the edge of the boreal zone and eastwards into Poland and Romania. *Puccinellia maritima* is a maritime species that is found around the coast of Europe, although it is absent from parts of southern Spain and the Adriatic coast. Their current European distributions were obtained as presence/absence data and mapped to a $0.5°$ latitude $\times 0.5°$ longitude grid (Fig. 1a, b). Five bioclimatic variables were selected as predictors, which are absolute minimum temperature expected over a 20-yr period, annual maximum temperature, growing degree days above a base temperature of $5°C$, mean soil water availability for the summer half year (May–September), and accumulated annual soil water deficit. These data are also at the scale of $0.5° \times 0.5°$. They were described in detail by Berry et al. (2001) and Pearson et al. (2002).

Design of modelling experiment

Multilayer feed-forward ANNs with back-propagation algorithm were trained with SAS software (release 8.1). The networks contained one input layer, one hidden layer and one output layer. There were 5 neurons in the input layer, which correspond to the 5 input variables, 5 neurons in the hidden layer, and 1 neuron in the output layer. This architecture was chosen after many modelling experiments with varying neurons in the hidden layer. The five environmental variables were
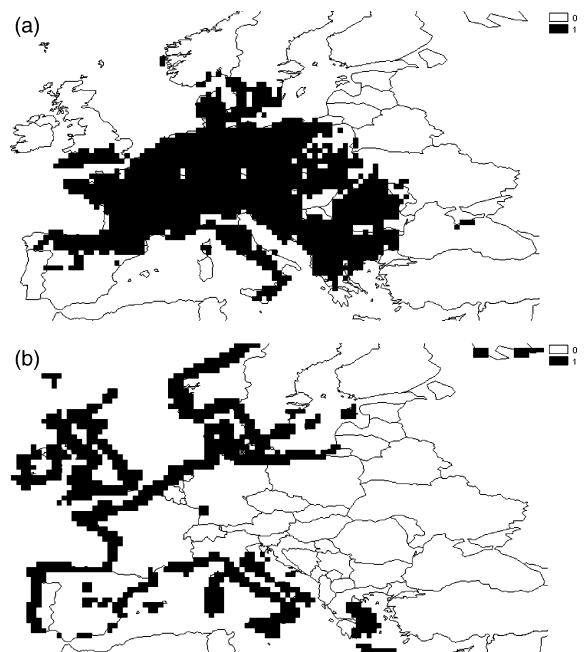

Fig. 1. The observed distributions of *F. sylvatica* (a) and *P. maritima* (b).

standardized to have zero mean and unit standard deviation.

In order to investigate the performance of the threshold-determining approaches in varying situations, we set seven levels of prevalence for both model-building data (including training data and validation data) and test data, i.e. 5, 10, 25, 50, 75, 90 and 95%. The sample size is 100 for each of the training, validation and test datasets. For each level of the prevalence for model-building data, one dataset for training was created by randomly sampling specified numbers of presences and absences without replacement from the original presences pool and the original absences pool respectively; then another dataset for validation and two datasets for testing were created sequentially from the left-over data without replacement. An ANN was trained using the created training dataset and validation dataset, and the resulting model was applied to the two test datasets for each of the seven levels of prevalence. This procedure was repeated five times for each level of prevalence for the model-building data. There are 10 sets of predictions for each combination of the levels of prevalence for model-building data and test data, 70 sets of predictions for each level of the prevalence for model building data, and 490 sets of predictions in total for each species.

For each set of model-building data, the threshold was determined by each of the twelve approaches (Table 2 for details). Then, these thresholds were applied to each testing dataset, and the four assessment indices,

including sensitivity, specificity, OPS and Kappa, were calculated.

Results

Overall assessment

The twelve approaches to determining thresholds were assessed using all the 490 sets of predictions from the 490 combinations of training datasets and test datasets with various prevalences for each species (Fig. 2). It can be seen that the trend for the two species is similar. The only exception is that approach 2 is better than approach 1 for *F. sylvatica*, but it is worse than approach 1 for *P. maritima* in specificity. The ranking of the twelve approaches given by specificity is different from those given by the other three indices (sensitivity, OPS and kappa), the latter three give similar ranking and OPS and kappa give almost the same ranking for the twelve approaches. It can be seen that approaches 4, 5, 7, 8, and 9 are relatively better than the other approaches (1, 2, 3, 6, 10, 11 and 12) according to the four indices, especially OPS and kappa. In the following two sections, we will take approaches 4 and 2 as representatives for the above two groups to investigate their further behaviours.

Assessment using model-building datasets with different prevalence

The twelve approaches were investigated using the training data with different prevalence, and the trends
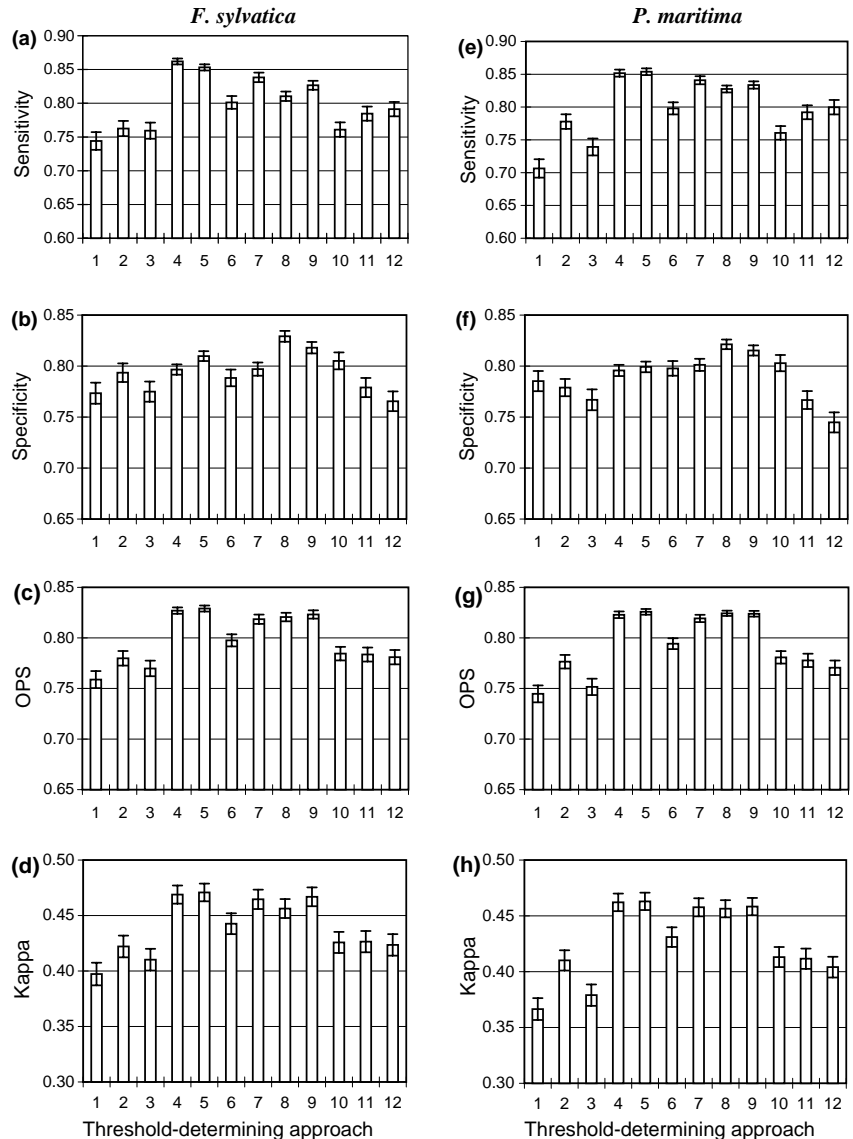

Fig. 2. Overall assessment of the twelve threshold-determining approaches using four indices: sensitivity (a, e), specificity (b, f), OPS (c, g) and kappa (d, h) for *F. sylvatica* (a, b, c, d) and *P. maritima* (e, f, g, h). The bars show the ±1 standard errors.
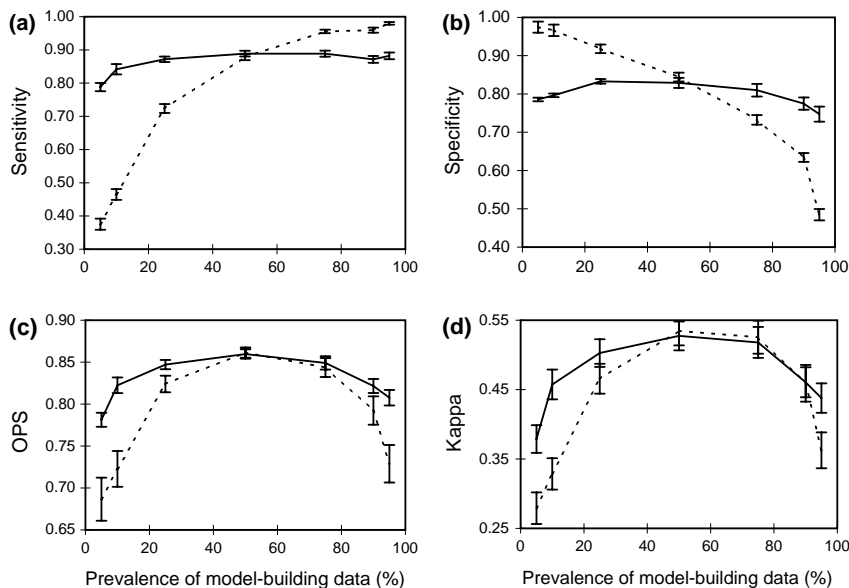
Fig. 3. Assessment of two threshold-determining approaches, i.e. approach 2 (dashed line) and 4 (solid line), which represent bad ones and good ones respectively, using sensitivity (a), specificity (b), OPS (c) and Kappa (d) for *F. sylvatica* when the prevalence of model-building data is different. The bars show the ± 1 standard errors.

for the two species are the same. The results of the two approaches (2 and 4) for *F. sylvatica* are shown in Fig. 3. The sensitivity and specificity for approach 2 are severely affected by the prevalence of model-building data, and those for approach 4 are not. The ranking of the approaches changes when the prevalence of model-building data varies according to sensitivity and specificity. But according to OPS and kappa, the ranking of the approaches remains relatively stable, i.e. approach 4 is almost always better than approach 2. Detailed investigation shows that among the five good approaches (4, 5, 7, 8 and 9), approach 7 is relatively more sensitive to the prevalence of model-building data and approaches 4 and 5 are more robust when the prevalence of model-building data changes. It is obvious that there is the least difference among different approaches when the prevalence of model-building data is 50%.

Assessment using test datasets with different prevalence

The twelve approaches were further investigated using the test datasets with different prevalence when the prevalence of model-building data is fixed, and the results for *P. maritima* are consistent with those for *F. sylvatica*. The results of the two representative approaches (2 and 4) for *F. sylvatica* are shown in Fig. 4. When the prevalence of test data changes, ranking of different approaches keeps relatively stable according to sensitivity and specificity; but it varies according to OPS and kappa. In addition, approach 4 is less severely affected by the prevalence of test data than approach 2 according to OPS and

kappa. In this respect, approach 4 is also better than approach 2.

Discussion

Finding a threshold and making the presence/absence prediction is the final step in species distribution modeling, and it is necessary in, for example, the estimation of species range and the assessment of the impact of climate change. It is important to give an accurate presence/absence prediction in these situations. However, in other situations, other considerations should be included. For example, in species reintroduction programs, we may limit the reintroduction sites to the most suitable areas; but in some conservation planning programs, we may take a less restrictive strategy, that is, purposely including some less suitable areas in protection. It is expected that the larger the predicted probability/suitability of presence at a site, the more suitable is the site to the reintroduction of the species. However, since the prevalence of model-building data has significant effect on the predicted probability/suitability of presence, i.e. the higher the prevalence, the bigger the predicted probability/suitability (Cramer 2003, Liu et al. unpubl.), this makes it difficult to decide the more suitable or less suitable sites. Therefore, even in those applications with some subjective decision-makings involved, it is still necessary to find the appropriate threshold and take the "objective" presence/absence prediction as a reference.

In this study, we treated two kinds of errors, e.g. false positive and false negative, as equally important and gave no preference to either side. But approach 12 can be

Fig. 4. Assessment of the two threshold-determining approaches, i.e. approach 2 (dashed line) and 4 (solid line), which represent bad ones and good ones respectively, using sensitivity, specificity, OPS and Kappa for *F. sylvatica* when the prevalence of test data is different and the prevalence of model-building data is fixed at 10 and 90%. The bars show the ± 1 standard errors.
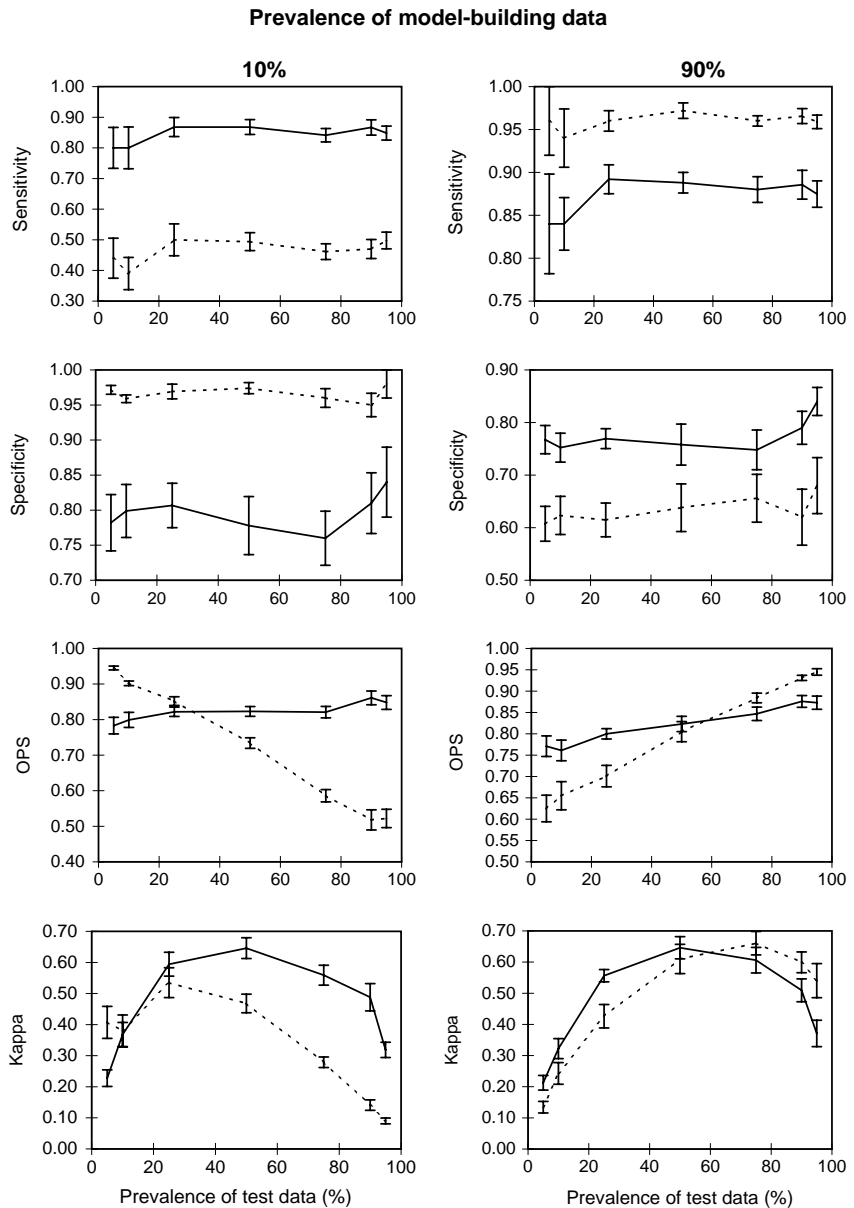
adapted to the situation in which one of two conflicting sides is emphasized by changing the parameter α from 0 to 1. A smaller α emphasizes recall, and bigger α emphasizes precision. However, it is difficult to say to what degree one side is emphasized. In this situation, the subjective approach may be suitable, e.g., a "minimum acceptable error" could be defined that depended on the intended application of the model. For example, compared with false negatives, we could tolerate more false positives when we set up a conservation area for a particularly endangered species. If the purpose of the model was to identify experimental sites, where we could find a species, we should minimize the false positive error rate (Fielding and Bell 1997).

When some kind of cost for the false positive or false negative and/or benefit for true positive or true negative needs to be taken into account, Metz's (1978) approach can be adopted because the cost of false positive and the benefit of true positive as well as prevalence were explicitly considered. However, the relevant cost and benefit are difficult to determine in environmental and ecological practice. Zweig and Campbell (1993) suggested that if FPC > FNC, the threshold should favor specificity, while sensitivity should be favored if FNC > FPC (Fielding and Bell 1997). Because estimation of the cost and benefit may add more uncertainty to the problem, caution must be taken when this approach is adopted.

It is interesting to note that among the twelve approaches we studied, both sensitivity and specificity for approaches 4, 5, 7, 8 and 9 are high (>0.8) and are higher than those for the other approaches. Since the bigger the sensitivity, the smaller the false negatives rate, and the bigger the specificity, the smaller the false positives rate, therefore, both false positives rate and false negatives rate are low (<0.2) for the approaches 4, 5, 7, 8 and 9. These approaches are recommended to use. The other approaches either have low false negatives rate and high false positives rate (e.g. approach 12), or have high false negatives rate and low false positives rate (e.g. approach 10), or have both high false positives rate and high false negatives rate (e.g. approaches 1 and 3), therefore, these approaches are not recommended.

It is not unexpected that the fixed threshold approach (threshold =0.5) is one of the worst. Guisan and Theurillat (2000) found that the threshold histogram is not centered on 0.5 with symmetric tails in each opposite direction (toward 0 and 1), rather all values range between 0.05 and 0.65 with a mean at 0.35 and an asymmetric shape. In fact, the prevalence of model-building data affects all the results. The output is biased towards the larger of the two groups (Fielding and Bell 1997, Cramer 2003), occupied sites and unoccupied sites. When the prevalence is small, a 0.5 threshold would classify most of the sites as unoccupied (Cumming 2000). However, the prevalence approach is one of the most robust, i.e. although it is not the best in every situation, it is good, at least not bad, even in the worst situation. Actually it is one of the best as assessed using sensitivity, OPS and kappa. This is also not unexpected. In fact, in another study we found that the probabilities corresponding to the maximum OPS and maximum kappa for the test data are positively correlated to the prevalence of model-building data (Liu et al. unpubl.). We suggested that a good presence/absence prediction would be obtained by taking the prevalence of model-building data as the threshold. This hypothesis was verified by this study.

From this study we also found that when the prevalence of model-building data is 50%, there is little difference among the twelve approaches as measured by the four indices (Fig. 3), and the relative difference (=difference of the maximum and minimum among the twelve values for each index divided by the maximum) is <5% for all the four indices and the two species. Furthermore, in addition to approaches 1 and 4, approaches 2, 3 and 7 (and also 11 for *P. maritima*), 8 and 10, and 5 and 6 reach exactly the same result respectively for each of the two species. The convergence of approaches 1 and 4 is obvious. The convergence of approaches 5 and 6 can be easily deduced because there are equal number of occupied sites and unoccupied sites. The convergence of approaches 2, 3 and 7 means that kappa, OPS and the sum of sensitivity and specificity

were maximized at the same time. The convergence of approaches 8 and 10 means that specificity =sensitivity (=recall) =precision at the same time. This means that many conditions are satisfied or nearly satisfied at the same time in this situation. Therefore, the best result will most probably be obtained by any approach, even the poor ones, which is verified by this study (Fig. 3c, d). This is encouraging since it supports our recommendation that it is better to use model-building data with prevalence of 50% in species distribution modeling (Liu et al. unpubl.).

Conclusion

The prevalence approach and average probability/suitability approach are simple and effective, and they are at least as good as the more complicated approaches, i.e. sensitivity-specificity sum maximization approach, sensitivity-specificity equality approach and the ROC plot-based approach. These five approaches fall into the group of good ones. Unfortunately, one of the widely used approaches, i.e. fixed threshold approach, is the worst one, which is not therefore recommended. Another popular approach, i.e. kappa maximization approach, is also not a good one. We also recommend that if possible, using datasets with prevalence of 50% to build models since in addition to other advantages, it is easier to find the optimal threshold.

References

Austin, M. P. 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modeling. – Ecol. Modell. 157: 101–118.

Bailey, S.-A., Haines-Young, R. H. and Watkins, C. 2002. Species presence in fragmented landscapes: modeling of species requirements at the national level. – Biol. Conserv. 108: 307–316.

Berg, Å, Gärdenfors, U. and von Proschwitz, T. 2004. Logistic regression models for predicting occurrence of terrestrial mollusks in southern Sweden – importance of environmental data quality and model complexity. – Ecography 27: 83–93.

Berry, P. M. et al. 2001. Impacts on terrestrial environments. – In: Harrison, P. A., Berry, P. M. and Dawson, T. P. (eds), Climate change and nature conservation in the Britain and Ireland: modelling natural resource responses to climate change (the MONARCH project). UKCIP Tech. Rep., pp. 43–150.

Berry, P. M. et al. 2002. Modelling potential impacts of climate change on the bioclimatic envelope of species in Britain and Ireland. – Global Ecol. Biogeogr. 11: 453–462.

Brosse, S. et al. 1999. The use of artificial neural networks to assess fish abundance and spatial occupancy in the littoral zone of a mesotrophic lake. – Ecol. Modell. 120: 299–311.

Cantor, S. B. et al. 1999. A comparison of C/B ratios from studies using receiver operating characteristic curve analysis. – J. Clin. Epidemiol. 52: 885–892.

Collingham, Y. et al. 2000. Predicting the spatial distribution of non-indigenous riparian weeds: issues of spatial scale and extent. – J. Appl. Ecol. 37 (Suppl. 1): 13–27.

Cramer, J. S. 2003. Logit models: from economics and other fields. – Cambridge Univ. Press, pp. 66–67.

Cumming, G. S. 2000. Using habitat models to map diversity: pan-African species richness of ticks (Acri: lxodida). – J. Biogeogr. 27: 425–440.

D'heygere, T., Gorthals, P. L. M. and De Pauw, N. 2003. Use of genetic algorithms to select input variables in decision tree models for the prediction of benthic macroinvertebrates. – Ecol. Modell. 160: 291–300.

Fielding, A. H. and Haworth, P. F. 1995. Testing the generality of bird-habitat models. – Conserv. Biol. 9: 1446–1481.

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.

Guisan, A. and Theurillat, J.-P. 2000. Equilibrium modeling of alpine plant distribution: how far can we go. – Phytocoenologia 30: 353–384.

Guisan, A. and Zimmermann, N. E. 2000. Predictive habitat distribution models in ecology. – Ecol. Modell. 135: 147–186.

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.

Guisan, A., Weiss, S. B. and Weiss, A. D. 1999. GLM versus CCA spatial modeling of plant species distribution. – Plant Ecol. 143: 107–122.

Guisan, A., Edwards, T. C. Jr and Hastie, T. 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. – Ecol. Modell. 157: 89–100.

Hilbert, D. W. and Ostendorf, B. 2001. The utility of artificial neural networks for modeling the distribution of vegetation in past, present and future climates. – Ecol. Modell. 146: 311–327.

Huntley, B. et al. 1995. Modelling present and potential future ranges of some European higher plants using climate response surfaces. – J. Biogeogr. 22: 967–1001.

Lehmann, A. 1998. GIS modeling of submerged macrophyte distribution using Generalized Addition Models. – Plant Ecol. 139: 113–124.

Luck, G. W. 2002. The habitat requirements of the rufous treecreeper (Climacteris rufa). 2. Validating predictive habitat models. – Biol. Conserv. 105: 395–403.

Manel, S., Dias, J.-M. and Ormerod, S. J. 1999. Comparing discriminant analysis, neural networks and logistic regression for predicting species distributions: a case study with a Himalayan river bird. – Ecol. Modell. 120: 337–347.

Manel, S., Williams, H. C. and Ormerod, S. J. 2001. Evaluating presence-absence models in ecology: the need to account for prevalence. – J. Appl. Ecol. 38: 921–931.

Metz, C. E. 1978. Basic principles of ROC analysis. – Seminar in Nuclear Medicine 8: 283–298.

Moisen, G. G. and Frescino, T. S. 2002. Comparing five modeling techniques for predicting forest characteristics. – Ecol. Modell. 157: 209–225.

Nahm, U. Y. and Mooney, R. J. 2000. Using information extraction to aid the discovery of prediction rules from text. – Proc. of the KDD (Knowledge Discovery in Databases) – 2000 Workshop on Text Mining, pp. 51–58.

Olden, J. D. 2003. A species-specific approach to modelling biological communities and its potential for conservation. – Conserv. Biol. 17: 854–863.

Olden, J. D. and Jackson, D. A. 2001. Fish-habitat relationships in lakes: gaining predictive and explanatory insight by using artificial neural networks. – Trans. Am. Fish. Soc. 130: 878–897.

Osborne, P. E., Alonso, J. C. and Bryant, R. G. 2001. Modelling landscape-scale habitat use using GIS and remote sensing: a case study with great bustards. – J. Appl. Ecol. 38: 458–471.

Ozesmi, S. L. and Ozesmi, U. 1999. An artificial neural network approach to spatial habitat modelling with interspecific interaction. – Ecol. Modell. 116: 15–31.

Pearce, J. and Ferrier, S. 2000. Evaluating the predictive performance of habitat models developed using logistic regression. – Ecol. Modell. 133: 225–245.

Pearson, R. et al. 2002. SPECIES: A Spatial Evaluation of Climate Impact on the Envelope of Species. – Ecol. Modell. 154: 289–300.

Pearson, R., Dawson, T. P. and Liu, C. 2004. Modelling species distributions in Britain: a hierarchical integration of climate and land-cover data. – Ecography 27: 285–298.

Robertson, M. P., Caithness, N. and Villet, M. H. 2001. A PCA-based modeling technique for predicting environmental suitability for organisms from presence records. – Div. Distribut. 7: 15–27.

Robertson, M. P. et al. 2003. Comparing models for predicting species' potential distributions: a case study using correlative and mechanistic predictive modeling techniques. – Ecol. Modell. 164: 153–167.

Schapire, R. E., Singer, Y. and Singhal, A. 1998. Boosting and Rocchio applied to text filtering. – Proc. ACM SIGIR, pp. 215–223.

Stockwell, D. R. B. and Peterson, A. T. 2002. Effects of sample size on accuracy of species distribution models. – Ecol. Modell. 148: 1–13.

Woolf, A. et al. 2002. Statewide modeling of bobcat, Lynx rufus, habitat in Illinois, USA. – Biol. Conserv. 104: 191–198.

Zweig, M. H. and Campbell, G. 1993. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. – Clin. Chem. 39: 561–577.