



Prevalence, statistical thresholds, and accuracy assessment for species distribution models

B. B. Hanberry and H. S. He

School of Natural Resources, Univ. of Missouri, 203 Natural Resources Building, Columbia, MO 65211, USA

Correspondence to: B. B. Hanberry (hanberryb@missouri.edu)

Received: 16 July 2012 – Revised: 19 March 2013 – Accepted: 22 April 2013 – Published: 13 May 2013

Abstract. For species distribution models, species frequency is termed prevalence and prevalence in samples should be similar to natural species prevalence, for unbiased samples. However, modelers commonly adjust sampling prevalence, producing a modeling prevalence that has a different frequency of occurrences than sampling prevalence. The separate effects of (1) use of sampling prevalence compared to adjusted modeling prevalence and (2) modifications necessary in thresholds, which convert continuous probabilities to discrete presence or absence predictions, to account for prevalence, are unresolved issues. We examined effects of prevalence and thresholds and two types of pseudoabsences on model accuracy. Use of sampling prevalence produced similar models compared to use of adjusted modeling prevalences. Mean correlation between predicted probabilities of the least (0.33) and greatest modeling prevalence (0.83) was 0.86. Mean predicted probability values increased with increasing prevalence; therefore, unlike constant thresholds, varying threshold to match prevalence values was effective in holding true positive rate, true negative rate, and species prediction areas relatively constant for every modeling prevalence. The area under the curve (AUC) values appeared to be as informative as sensitivity and specificity, when using surveyed pseudoabsences as absent cases, but when the entire study area was coded, AUC values reflected the area of predicted presence as absent. Less frequent species had greater AUC values when pseudoabsences represented the study background. Modeling prevalence had a mild impact on species distribution models and accuracy assessment metrics when threshold varied with prevalence. Misinterpretation of AUC values is possible when AUC values are based on background absences, which correlate with frequency of species.

1 Introduction

Species distribution models predict the occurrence probability of species in space based on limited known locations where species are present. Although modeling with sampling prevalence, or frequency of samples that contain the species, may be an appropriate approach (Real et al., 2006; Albert and Thuiller, 2008; Ward et al., 2009; Li et al., 2011; Meynard and Kaplan, 2012), there also are potential limitations. Sampling prevalence may not reflect natural species prevalence, or frequency of the species over the entire study extent. That is, sampling prevalence observed in samples may be biased because for example, frequency may vary throughout the study area. Indeed, the study area often is a subsample of the entire distribution range and thus, artificially defined by

the modeler. In some cases, to maintain sampling prevalence it may be necessary to retain fewer samples, thus reducing sample size and accuracy of models (Hanberry et al., 2012a). Furthermore, if background data represent samples that do not contain the species, then prevalence will vary depending on resolution (i.e., number of samples) of the background data (Franklin et al., 2010). Lastly, adjustment from sampled prevalence that varies by species to a consistent modeling prevalence applied to all species may be necessary to standardize comparisons of model predictions among species or study areas or to map model predictions in similar and conventional classification units, such as equal intervals.

Modeling prevalence is the frequency of occurrence specifically selected for model training (Jiménez-Valverde et al., 2008). The most common recommendation is a balanced

modeling prevalence of 0.50 (McPherson et al., 2004; Liu et al., 2005), however researchers have suggested unbalanced modeling prevalence, such as 0.01 when sample sizes are limited (Lobo and Tognelli, 2011). Although adjustment of prevalence provides modeling flexibility, bias may reduce accuracy of models based on adjusted modeling prevalence more than sampling prevalence, which differ less from natural species prevalence.

Thresholds, defined by the researcher, specify the predicted probability above which a species is determined present. Often a threshold of 0.5 simply is used as a threshold for species presence, but may not to be the best choice (Liu et al., 2005; Freeman and Moisen, 2008). Liu et al. (2005) compared 12 different threshold selection methods and determined five methods that worked better than others to maximize both sensitivity (true positive rate) and specificity (true negative rate). One method was the use of modeling prevalence as the threshold. Another method was the use of mean predicted probability for the species as the threshold, but values a little lower than the mean also would be more likely to include the species than other threshold values. Three of the best five methods required sensitivity and specificity values, but specificity depends on known absences, which are uncertain for most datasets.

The most common accuracy assessment metric currently is the area under the curve (AUC) of the receiver operating characteristic (ROC), which plots true positive rate (sensitivity) against false positive rate (1-specificity; commission error). The AUC values are threshold-independent, which may not necessarily be a beneficial attribute if (1) the accuracy metric is meant to measure presence and absence at a specified threshold dividing species presence from absence, (2) omission and commission errors are not equal in importance, and (3) there is inclusion of irrelevant prediction ranges (Lobo et al., 2007; Jiménez-Valverde, 2011). Furthermore, AUC values appear to be sensitive to the extent and location of species distributions (Elith et al., 2006; Tarmansen, 2006; Lobo et al., 2007; Raes and ter Steege, 2007). Pseudoabsences replace absences for most datasets, introducing uncertainty (Jiménez-Valverde, 2011). Therefore, researchers must reinterpret the AUC curve in the context of pseudoabsences selected from the background (Peterson et al., 2008; Jiménez-Valverde, 2011) or select pseudoabsences from surveyed plots rather than pseudoabsences from the unknown background extent (Hanberry et al., 2012b). Another option is to use the area of predicted presence as a proxy for commission errors (Engler et al., 2004; Hernandez et al., 2006) to match researcher-specified thresholds of species presence.

Modeling prevalence, along with the selected threshold, may affect some accuracy metrics, perhaps due to bias away from natural species prevalence (Manel et al., 2001; Jiménez-Valverde and Lobo, 2006). Contradiction about the effects and bias of prevalence (Fielding and Bell, 1997; Manel et al., 2001; McPherson et al., 2004; Jiménez-Valverde and

Lobo, 2006; Meynard and Kaplan, 2012) may be due to use of alternative types of prevalence (i.e., sampling prevalence vs. adjusted modeling prevalence) and varying threshold designations and accuracy assessment metrics. Therefore, we explored some unresolved issues about species distribution models, that is, (1) whether sampled prevalence produces less biased models than adjusted modeling prevalence, (2) whether modeling prevalence should be balanced at 0.5 or unbalanced (i.e., sampling prevalence or an alternative modeling prevalence), (3) whether threshold selection alone can account for sampling and adjusted modeling prevalence, and (4) the effects of prevalence, thresholds, and pseudoabsence type on accuracy assessment. It was not possible to maintain the sampling prevalence and keep a constant sample size of present cases; however, models were weighted averages of numerous models (e.g., ensemble trees in random forests) and the cumulative model had access to a greater modeling sample of present and absent cases. We adjusted modeling prevalence from 0.33 to 0.83, while keeping sample size of present cases constant. To examine effect of thresholds, we used a constant threshold of 0.50 and thresholds equal to prevalence. We based accuracy assessment on true positive rates, AUC values and true negative rates using surveyed and background pseudoabsences, and areal extent of predicted species presence.

2 Methods

2.1 Tree surveys

The USDA Forest Service Forest Inventory and Analysis (FIA) surveys fixed plots (composed of four subplots that are each 7.3 m in radius), which are located uniformly across the landscape, during a five year cycle. The latest complete cycle for Minnesota's Laurentian Mixed Forest (Fig. 1) was during 2004–2008 and contained 2666 plots. Because the available FIA plot locations are perturbed to protect landowner privacy, the USDA Forest Service joined a set of predictor variables (described below) to plots to provide a table without revealing locations but based on accurate spatial locations for modeling and prediction.

We selected tree species with at least 2500 individuals, to ensure that models did not lose accuracy because of small samples, and we also created two mixed species groups. The species were American basswood (*Tilia americana*), balsam fir (*Abies balsamea*), balsam poplar (*Populus balsamifera*), black ash (*Fraxinus nigra*), black spruce (*Picea mariana*), bur oak (*Quercus macrocarpa*), jack pine (*Pinus banksiana*), northern white cedar (*Thuja occidentalis*), paper birch (*Betula papyrifera*), red maple (*Acer rubrum*), red pine (*Pinus resinosa*), sugar maple (*A. saccharum*), tamarack (*Larix laricina*), and quaking aspen (*Populus tremuloides*). The species groups were aspens (*Populus tremuloides*, *P. balsamifera*) and maples (*Acer rubrum*, *A. saccharum*).

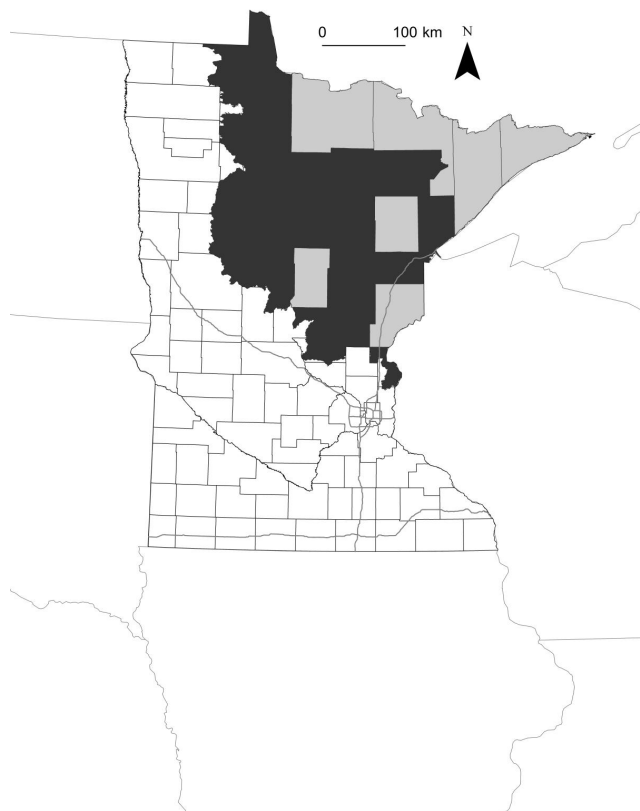


Figure 1. Study area (shaded black; contains soils surveys) in Minnesota's Laurentian Mixed Forest (shaded black and grey).

2.2 Spatial units and environmental variables

The spatial mapping units or grain were Soil Survey Geographic (SSURGO) Database (Natural Resources Conservation Service, <http://soildatamart.nrcs.usda.gov>) polygons. Soil surveys have not been completed in Cook, Crow Wing, Isanti, Koochiching, Lake, Pine, and St. Louis counties, leaving a study extent of about 4 895 238 ha (Fig. 1). After removal of polygons that were water or otherwise miscellaneous areas (e.g., mines, pits, dumps), there were 310 000 polygons with a mean polygon area of 16 ha ($SD = 92$).

We used sixteen predictor variables that are important for tree presence. For soil variables, we determined values based on polygons with similar characteristics by county (soil map units; 2364 map units total). Soil variables were (1) drainage class (very poorly drained to excessively drained), (2) hydric soil presence class, (3) water holding capacity (cm cm^{-1}), (4) pH, (5) organic matter (%), (6) clay (%), and (7) sand (%). We intersected two more categorical variables to each soil polygon: (8) subsection, which is an ecological division of the study area (ECOMAP, 1993) and (9) bedrock geology. We created a unique unit based on soil map unit, land type association (an ecological classification), and bedrock geology, which contained spatially distinct soil polygons that averaged about 210 ha. We determined mean values

for the unit, which became the unit for predicted probabilities, from a 30 m DEM (digital elevation model), for (10) elevation (m), (11) slope (%), (12) transformed aspect ($1 + \sin(\text{aspect}/180 \times 3.14 + 0.79)$; Beers et al., 1966), (13) solar radiation (07:00 to 19:00 in 4 h intervals on summer solstice for re-sampled 60 m DEM), (14) topographic roughness (Sappington et al., 2007), (15) wetness convergence (T. Dilts, <http://arcscrips.esri.com>), and (16) topographic position index.

2.3 Statistical analyses and prevalence

We applied random forests (Breiman, 2001; Cutler et al., 2007), a classification method based on bootstrap aggregation (bagging) by the majority vote of many trees grown using random samples of both predictor variables and modeling data. We used the randomForest package (Liaw and Wiener, 2002) in R statistical software (R Development Core Team, 2010). We set the number of classification trees at 2000 and the number of variables randomly sampled at each split as the square root of the number of predictors.

We randomly selected a set containing 66% of plots, to a maximum of 2500, which contained each species for modeling, reserving the rest for accuracy assessment. For pseudoabsences, we selected up to 2500 plots that did not contain the species. Because random forests classification is an ensemble method that averages many trees, we worked around the limitations of sampled prevalence by allowing each classification tree iteration to draw from the complete modeling set. For sampled prevalence, we maintained the sampling prevalence for each species of 0.09–0.69, but we adjusted modeling prevalence with the sample size option (which is sampled without replacement). We varied the prevalence, using a moderate range from 0.33 to 0.83, representing ratios of pseudoabsence cases to present cases of 2 : 1 (prevalence of 0.33), 1 : 1 (0.50), 1 : 2 (0.67), 1 : 3 (0.75), 1 : 4 (0.8), and 1 : 5 (0.83). We held the present case sample size at 335 while varying the total sample size from 400 to 1000, i.e., we set the modeling prevalence at (1) 335 presences/1000 total cases (0.33 modeling prevalence), (2) 335 presences/700 total cases (0.50 modeling prevalence), (3) 335 presences/500 total cases (0.67 modeling prevalence), (4) 335 presences/446 total cases (0.75 modeling prevalence), (5) 335 presences/419 total cases (0.8 modeling prevalence), and (6) 335 presences/400 total cases (0.83 modeling prevalence). It was not possible to achieve a sample size of 335 present cases and maintain the sampling prevalence. Instead, sample sizes of present cases ranged from 164 to 1236 with a mean of 504 samples.

2.4 Comparisons, thresholds, and accuracy assessment

We compared predicted probabilities for the same species for each prevalence using correlation (Proc Corr; SAS software, Version 9.2, Cary, North Carolina, USA). We also

Table 1. Correlation (all species combined) among predicted probabilities for varying modeling prevalences at a constant sample size of present cases and sampling prevalence (mean sampling prevalence = 0.28, range = 0.09–0.69).

prevalence	0.33	0.5	0.67	0.75	0.8	0.83	sampling
0.33	1.00	0.98	0.94	0.90	0.88	0.86	0.96
0.5	0.98	1.00	0.98	0.96	0.94	0.93	0.98
0.67	0.94	0.98	1.00	0.99	0.98	0.97	0.91
0.75	0.90	0.96	0.99	1.00	1.00	0.99	0.88
0.8	0.88	0.94	0.98	1.00	1.00	1.00	0.85
0.83	0.86	0.93	0.97	0.99	1.00	1.00	0.83
sampling	0.96	0.98	0.91	0.88	0.85	0.83	1.00

determined the mean of predicted probabilities for each prevalence. For thresholds to accept presence of a species, we used a constant 0.50 threshold and also thresholds that matched the modeling prevalence. We computed true positive rate, AUC, and true negative rate values (ROCR package in R; Sing et al., 2005) using reserved present samples and surveyed pseudoabsences from plots that did not have records of the species and also background pseudoabsences where the entire study extent was coded as absent (but we only modeled using surveyed pseudoabsences). We calculated area predicted as present and converted area from an absolute value to the fraction of the total area (4 895 238 ha). We examined the relationship between area predicted as present and true negative rate and AUC values (Proc Reg; SAS software, Version 9.2, Cary, North Carolina, USA).

3 Results

Independent of sampling prevalence and thresholds, correlations indicated that the relationship of the predicted probabilities differed only slightly with modeling prevalence values. Correlation values among models based on modeling prevalence (all species combined) at a constant sample size for present cases ranged from 0.86 (correlation between the most distant modeling prevalences of 0.33 and 0.83) to 1.00 (correlation between the closest modeling prevalences; Table 1). Models based on sampling prevalences were slightly less similar to models based on modeling prevalences, in part due to varying sample size. Correlation values ranged from 0.96 at the lower modeling prevalences to 0.83 at the greatest modeling prevalence. In addition, predicted probabilities increased with modeling prevalence. Mean predicted probabilities for all species combined increased from 0.25 for the modeling prevalence of 0.33 to 0.60 for the modeling prevalence of 0.83 (Table 2).

Because predicted probabilities varied with prevalence, thresholds modified the trade-off between true positive rate and true negative rate due to prevalence, as balance in commission (false positive rate) and omission (false negative rate) errors changed (Fig. 2; mean values for all species com-

bined). As prevalence (and predicted probabilities) increased and the threshold stayed constant, mean true positive rate and species prediction area increased while true negative rate decreased, creating large differences in values among prevalence ratios. Changing the threshold to match the modeling prevalence resulted in relatively similar true positive rate, true negative rate, and area for all prevalence values (Fig. 2). Models based on sampling prevalence performed similarly to models based on adjusted modeling prevalence.

Mean AUC values (all species combined) for the modeling prevalence models ranged from 0.92 to 0.94 using surveyed pseudoabsences (Table 2). The AUC values reflected true positive and true negative rates at the prevalence threshold ($R^2 = 0.97$), with a slightly greater influence by true negative rate ($R^2 = 0.90$) compared to true positive rate ($R^2 = 0.79$). Area was not correlated with AUC values ($R^2 = 0.07$) and true negative rate at the prevalence threshold ($R^2 = 0.06$). When using the background to provide pseudoabsences, AUC and true negative rate values were lower (Table 2). Area explained more of the variance in AUC values ($R^2 = 0.35$) and true negative rate at the prevalence threshold ($R^2 = 0.41$).

4 Discussion

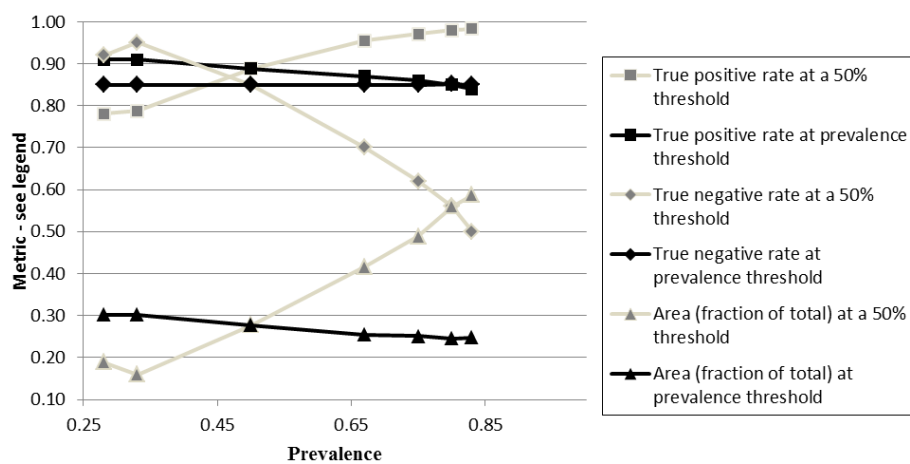
There appears to be only a minor influence by prevalence, whether based on sampling or adjusted for modeling, on species distribution models for common species (and for uncommon species, Jiménez-Valverde et al., 2009). Models with a prevalence ratio ranging from (lower than) 0.33 to 0.83 will be highly correlated. Modeling prevalence therefore does not need to be balanced (as recommended by McPherson et al., 2004; Liu et al., 2005). However, better models may result from increased representation by the present case, rather than models with increased representation of the unknown case.

Threshold selection alone can account for changing values of prediction probabilities in species distribution models due to modeling prevalence. Using a constant threshold for different prevalence values clearly will affect accuracy metrics by minimizing either omission (at greater prevalence) or commission (at lower prevalence) errors. Retaining a threshold that is similar to prevalence will maintain fairly constant error rates, no matter the selected prevalence, similar to findings reported by Liu et al. (2005).

Summary statistics and thresholds provide ways to assess species distribution models. True positive rate, true negative rate, and AUC values are basic measures of whether the model is able to predict presence and absence of species. One AUC value may be more helpful than two summarized values of true positive rate and true negative rate, even though for most datasets, true absent cases are uncertain. Any research that uses AUC (or true negative rate) values must adjust commission error so that it does not reflect merely frequency of

Table 2. Mean values for predicted probability, true positive rate, AUC, and true negative rate for varying modeling prevalences and sampling prevalence (mean sampling prevalence = 0.28, range = 0.09–0.69).

Prevalence	0.33	0.50	0.67	0.75	0.80	0.83	Sampling
	Mean value						
Predicted probability	0.25	0.34	0.45	0.51	0.56	0.60	0.25
True positive rate at 0.5 threshold	0.79	0.89	0.95	0.97	0.98	0.98	0.78
True positive rate at prevalence threshold	0.91	0.89	0.87	0.86	0.85	0.84	0.91
AUC with surveyed absences	0.94	0.94	0.93	0.92	0.92	0.92	0.95
AUC with background absences	0.90	0.90	0.89	0.89	0.89	0.88	0.91
True negative rate at 0.5 threshold	0.95	0.85	0.70	0.62	0.56	0.50	0.92
True negative rate at prevalence threshold	0.85	0.85	0.85	0.85	0.85	0.85	0.85
True negative rate at prevalence threshold with background absences	0.68	0.73	0.75	0.77	0.77	0.77	0.67

**Figure 2.** True positive rate, true negative rate, and area (fraction of total area) at 0.50 thresholds and prevalence thresholds. The prevalence threshold maintained true positive rate, true negative rate, and area at a relatively constant value among modeling prevalence ratios. Sampling prevalence (points located at the 0.28 x-axis, the mean sampling prevalence) did not affect performance.

species when the background is coded as absent. If the background is coded as absent, AUC values indicate whether a species is widespread or restricted in range within the study extent; models for species with small or restricted ranges will have greater AUC values due to the match between low predicted probabilities and background areas scored as absent (Elith et al., 2006; Stokland et al., 2011). When we coded the background rather than surveyed areas as absent for assessment, albeit using surveyed pseudoabsences in modeling, the relationship between AUC values and area predicted as present increased. That is, R^2 increased from 0.07 to 0.35 in our study because R^2 values between true negative rate and area increased from 0.06 to 0.41. Additionally, it appeared that no matter the extent of commission error, if the true positive rate was high then the AUC value also will be high (Wisn et al., 2008).

Greater predicted probabilities should be meaningful indicators about the probability of species presence, however it is important to realize factors can increase and decrease

predicted probabilities. Increasing modeling prevalence increased predicted probabilities. Even though predicted probability values are consequential for a species, it is not certain what the predicted probabilities mean in terms of frequency, particularly when compared to other species (Lobo et al., 2007). For some conservation or monitoring goals, it may be useful to select areas where the species is very likely to be present at the expense of excluding potential areas of presence. Under those circumstances, setting the threshold at the mean predicted probability for known presences will target locations with greater probabilities of presence.

5 Conclusions

Modeling prevalence had a mild impact on species distribution models if thresholds for accepting species presence varied with prevalence and sample size was removed from modeling prevalence. We did not examine the effects of changing modeling prevalence for other statistical methods, and effects may differ if the method is not an ensemble

method. However, Liu et al. (2005) had similar results for threshold selection using artificial neural networks as the statistical method. Modelers certainly should (1) specify the modeling prevalence and threshold for species presence and accuracy assessment and (2) standardize threshold values among modeled species so that models have similar balance between omission and commission errors. The AUC (and true negative rate) values appeared to be meaningful as measures of whether models produce greater predicted probabilities where a species was present than where it was unknown (Phillips et al., 2006). However, interpretation is misleading because with known commission error of scoring the entire background as absent, AUC (and true negative rate) values reflected the area predicted as present, that is, the frequency of species, rather than simply error. Any AUC values that trend inversely with extent of species distributions may be artifacts of pseudoabsence selection.

Edited by: M. Bezemer

Reviewed by: G. Hengeveld and one anonymous referee

References

- Albert, C. H. and Thuiller, W.: Favourability functions versus probability of presence: advantages and misuses, *Ecography*, 31, 417–422, 2008.
- Beers, T. W., Dress, P. E., and Wensel, L. C.: Aspect transformation in site productivity research, *J. Forest.*, 64, 691–692, 1966.
- Breiman, L.: Random Forests, *Mach. Learn.*, 40, 5–32, 2001.
- Cutler, D. R., Edwards Jr., T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., and Lawler, J. J.: Random forests for classification in ecology, *Ecology*, 88, 2783–2792, 2007.
- ECOMAP: National hierarchical framework of ecological units, USDA Forest Service, Washington, DC, 1993.
- Engler, R., Guisan, A., and Rechsteiner, L.: An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data, *J. Appl. Ecol.*, 41, 263–274, 2004.
- 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. M., Peterson, A. T., Phillips, S. J., Richardson, K., Scachetti-Pereira, R., Schapire, R. E., Soberón, J., Williams, S., Wisz, M. S., and Zimmermann, N. E.: Novel methods improve prediction of species' distributions from occurrence data, *Ecography*, 29, 129–151, 2006.
- Fielding, A. H. and Bell, J. F.: A review of methods for assessment of prediction errors in conservation presence/absence models, *Environ. Conserv.*, 24, 38–49, 1997.
- Franklin, J.: Mapping species distributions: Spatial inference and predictions, Cambridge University Press, New York, 2010.
- Freeman, E. A. and Moisen, G. G.: A comparison of the performance of threshold criteria for binary classification in terms of predicted prevalence and kappa, *Ecol. Model.*, 217, 48–58, 2008.
- Hanberry, B. B., He, H. S., and Dey, D. C.: Sample sizes and model comparison metrics for species distribution models, *Ecol. Model.*, 227, 29–33, 2012a.
- Hanberry, B. B., He, H. S., and Palik, B. J.: Pseudoabsence generation for species distribution models, *PLoS ONE*, 7, e44486, doi:10.1371/journal.pone.0044486, 2012b.
- Hernandez, P. A., Graham, C. H., Master, L. L., and Albert, D. L.: The effect of sample size and species characteristics on performance of different species distribution modeling methods, *Ecography*, 29, 773–785, 2006.
- Jiménez-Valverde, A.: Insights into the area under the receiver operating characteristic curve AUC as a discrimination measure in species distribution modeling, *Global Ecol. Biogeogr.*, 21, 498–507, 2011.
- Jiménez-Valverde, A. and Lobo, J. M.: The ghost of unbalanced species distribution data in geographical model predictions, *Divers. Distrib.*, 12, 521–524, 2006.
- Jiménez-Valverde, A., Lobo, J. M., and Hortal, J.: Not as good as they seem: the importance of concepts in species distribution modelling, *Divers. Distrib.*, 14, 885–890, 2008.
- Jiménez-Valverde, A., Lobo, J. M., and Hortal, J.: The effect of prevalence and its interaction with sample size on the reliability of species distribution models, *Community Ecol.*, 10, 196–205, 2009.
- Li, W., Guo, Q., and Elkan, C.: Can we model the probability of presence of species without absence data?, *Ecography*, 34, 1096–1105, 2011.
- Liaw, A. and Wiener, M.: Classification and regression by random Forest, *R News*, 2, 18–22, 2002.
- Liu, C., Berry, P. M., Dawson, T. P., and Pearson, R. G.: Selecting thresholds of occurrence in the prediction of species distributions, *Ecography*, 28, 385–393, 2005.
- Lobo, J. M. and Tognelli, M. F.: Exploring the effects of quantity and location of pseudo-absences and sampling bias on the performance of distribution models with limited point occurrence data, *J. Nat. Conserv.*, 19, 1–7, 2011.
- Lobo, J. M., Jiménez-Valverde, A., and Real, R.: AUC: a misleading measure of the performance of predictive distribution models, *Global Ecol. Biogeogr.*, 17, 145–151, 2007.
- Manel, S., Williams, H. C., and Ormerod, S. J.: Evaluating presence-absence models in ecology: the need to account for prevalence, *J. Appl. Ecol.*, 38, 921–931, 2001.
- McPherson, J. M., Jetz, W., and Rogers, D. J.: The effects of species' range sizes on the accuracy of distribution models: ecological phenomenon or statistical artifact?, *J. Appl. Ecol.*, 41, 811–823, 2004.
- Meynard, C. N. and Kaplan, D. M.: The effect of a gradual response to the environment on species distribution modeling performance, *Ecography*, 35, 499–509, 2012.
- Peterson, A. T., Papeş, M., and Soberón, J.: Rethinking receiver operating characteristic analysis applications in ecological niche modeling, *Ecol. Model.*, 213, 63–72, 2008.
- Phillips, S. J., Anderson, R. P., and Schapire, R. E.: Maximum entropy modeling of species geographic distributions, *Ecol. Model.*, 190, 231–259, 2006.
- R Development Core Team: R: A language and environment for statistical computing, R Foundation for Statistical Computing, 2010.
- Raes, N. and ter Steege, H.: A null-model for significance testing of presence-only species distribution models, *Ecography*, 30, 727–736, 2007.

- Real, R., Barbosa, A. M., and Vargas, J. M.: Obtaining environmental favourability functions from logistic regression, *Environ. Ecol. Stat.*, 13, 237–245, 2006.
- Santika, T.: Assessing the effect of prevalence on the predictive performance of species distribution models using simulated data, *Global Ecol. Biogeogr.*, 20, 181–192, 2011.
- Sappington, J. M., Longshore, K. M., and Thompson, D. B.: Quantifying landscape ruggedness for animal habitat analysis: a case study using bighorn sheep in the Mojave Desert, *J. Wildlife Manage.*, 71, 1419–1426, 2007.
- Sing, T., Sander, O., Beerenwinkel, N., and Lengauer, T.: ROCr: visualizing classifier performance, *R. Bioinformatics*, 21, 3940–3941, 2005.
- Stokland, J. N., Halvorsen, R., and Støa, B.: Species distribution modelling – Effect of design and sample size of pseudo-absence observations, *Ecol. Model.*, 222, 1800–1809, 2011.
- Termansen, M., McClean, C. J., and Preston, C. D.: The use of genetic algorithms and Bayesian classification to model species distributions, *Ecol. Model.*, 192, 410–424, 2006.
- Ward, G., Hastie, T., Barry, S., Elith, J., and Leathwick, J. R.: Presence-only data and the EM algorithm, *Biometrics*, 65, 554–563, 2009.
- Wisz, M. S., Hijmans, R. J., Li, J., Peterson, A. T., Graham, C. H., and Guisan, A.: NCEAS Predicting Species Distributions Working Group: Effects of sample size on the performance of species distribution models, *Divers. Distrib.*, 14, 763–773, 2008.