MaxEnt brings comparable results when the input data is being completed; model parameterization and background manipulation of four species distribution models

Mohsen Ahmadi[1], Mahmoud-Reza Hemami[1], Mohammad Kaboli[2], and Farzin Shabani[3]

[1]Isfahan University of Technology
[2]University of Tehran
[3]Qatar University College of Arts and Sciences

September 19, 2022

Abstract

Species distribution models (SDMs) are practical tools to assess the habitat suitability of species with numerous applications in environmental management and conservation planning. The manipulation of the input data to deal with their spatial bias is one of the advantageous methods to enhance the performance of SDMs. However, the development of a model parameterization approach covering different SDMs to achieve well-performing models has never been implemented. We integrated input data manipulation and model tuning for four commonly-used SDMs; generalized linear model (GLM), gradient boosted model (GBM), random forest (RF), and maximum entropy (MaxEnt), and compared their predictive performance to model geographically imbalanced biased data of a rare species complex of mountain vipers. Models were tuned up based on a range of model-specific parameters considering two background selection methods; random and background weighting schemes. The performance of the fine-tuned models was assessed based on a recently identified localities of the species. The results indicated that although the fine-tuned version of all models shows great performance in predicting training data (AUC > 0.9 and TSS > 0.5), they produce different results in classifying out-of-bag data. The GBM and RF with higher sensitivity of training data showed more different performances. The GLM, despite having high predictive performance for test data, showed lower specificity. It was only the MaxEnt model that showed high predictive performance and comparable results for identifying test data in both random and background weighting procedures. Our results highlight that while GBM and RF are prone to overfitting training data and GLM over-predict non-sampled areas MaxEnt is capable of producing results that are both predictable (extrapolative) and complex (interpolative). We discuss the assumptions of each model and conclude that MaxEnt could be considered as a practical method to cope with imbalanced-biased data in species distribution modeling approaches.

Title:

MaxEnt brings comparable results when the input data is being completed; model parameterization and background manipulation of four species distribution models

Authors:

Mohsen Ahmadi[1], Mahmoud-Reza Hemami[1], Mohammad Kaboli[2], Farzin Shabani[3]

Affiliations:

[1] *Department of Natural Resources, Isfahan University of Technology, 84156-83111 Isfahan, Iran*

[2] *Department of Environmental Sciences, Faculty of Natural Resources, University of Tehran, Karaj, Iran*

1

[3] *Department of Biological and Environmental Sciences, College of Arts and Sciences, Qatar University, P.O. Box 2713, Doha, Qatar*

Correspondence

Mohsen Ahmadi, Department of Natural Resources, Isfahan University of Technology, Isfahan, Iran.

Email: mahmadi@iut.ac.ir

Abstract

Species distribution models (SDMs) are practical tools to assess the habitat suitability of species with numerous applications in environmental management and conservation planning. The manipulation of the input data to deal with their spatial bias is one of the advantageous methods to enhance the performance of SDMs. However, the development of a model parameterization approach covering different SDMs to achieve well-performing models has never been implemented. We integrated input data manipulation and model tuning for four commonly-used SDMs; generalized linear model (GLM), gradient boosted model (GBM), random forest (RF), and maximum entropy (MaxEnt), and compared their predictive performance to model geographically imbalanced biased data of a rare species complex of mountain vipers. Models were tuned up based on a range of model-specific parameters considering two background selection methods; random and background weighting schemes. The performance of the fine-tuned models was assessed based on a recently identified localities of the species. The results indicated that although the fine-tuned version of all models shows great performance in predicting training data (AUC > 0.9 and TSS > 0.5), they produce different results in classifying out-of-bag data. The GBM and RF with higher sensitivity of training data showed more different performances. The GLM, despite having high predictive performance for test data, showed lower specificity. It was only the MaxEnt model that showed high predictive performance and comparable results for identifying test data in both random and background weighting procedures. Our results highlight that while GBM and RF are prone to overfitting training data and GLM over-predict non-sampled areas MaxEnt is capable of producing results that are both predictable (extrapolative) and complex (interpolative). We discuss the assumptions of each model and conclude that MaxEnt could be considered as a practical method to cope with imbalanced-biased data in species distribution modeling approaches.

Keywords: habitat suitability, model tuning, spatially imbalanced-biased, MaxEnt.

1. Introduction

Species distribution models (SDMs) by combining data of species occurrence and environmental variables are operative tools to understand the dynamics of biodiversity distribution in space and time. A wealth of literature exists on the utility of SDMs all aiming at explaining, predicting, and projecting species distribution (Araújo et al., 2019). In particular identifying geographic distribution and most effective variables in different geographic scales (Brito et al., 2009; Hemami et al., 2018; Vale et al., 2014), assessing conservation coverage and efficiency of protected areas (Farhadinia et al., 2015; Lentini and Wintle, 2015; Zupan et al., 2014), predicting the biological invasion of alien species (Thuiller et al., 2005; Tingley et al., 2014), climate change-induced range shifts (Thuiller et al., 2011; Waltari and Guralnick, 2009; Yousefi et al., 2017), and combining its results with phylogenetic analyses to explore species evolutionary history (Ahmadi et al., 2018; Ahmadzadeh et al., 2016; Boucher et al., 2015; Saladin et al., 2019) are among the most widely-used aspects of SDMs.

In general, a variety of SDMs with different algorithms has been developed which may lead to different results for a target species (Elith and Graham, 2009; Merow et al., 2014). Consequently, model manipulation and comparing their results have been the subject of a significant amount of debate and research (Elith et al., 2006; Shabani et al., 2016; Wisz et al., 2008). According to Araújo et al. (2019), four aspects of

SDMs determine the quality of the resulting model, including response variable (i.e. occurrence records of the species), predictor variables, model building, and model evaluation. An SDM is a process of modeling and prediction, thus, contains levels of uncertainty that rises from each of the above-mentioned aspects.

At each level, solutions have been proposed to increase the quality of the data and reduce the negative effects of uncertainty in the output models. For example, in the first step, improving sampling design can reduce bias and inaccuracy in the geographical distribution of the collected data (Araújo and Guisan, 2006). At this level, ensuring that the collected data correctly represent the actual distribution of the species (Guillera-Arroita et al., 2015; Tessarolo et al., 2014) and that the scale of modeling and independent variables are consistent with sampling precision (Guisan et al., 2007; Wiens et al., 2009), and reducing unbiased recognition of the taxonomy of the species (Hortal et al., 2008; Rocchini et al., 2011) improve results of an SDM analysis.

An essential hypothesis of statistical methods is that recorded data are independent (i.e. randomly allocated samples with independent distribution), requiring that the entire area of interest is randomly or systematically sampled. In practice, available data on the species distribution is spatially biased toward areas easily assessed and/or better surveyed (Araujo and Guisan, 2006; Boria et al., 2014). A different strategy and intensity of sampling cause uneven distribution of recorded data, inconsistent with the real spatial ecology of the target species. This spatial bias may result in spatial clumpiness, which in turn, leads to the over-representation of areas with a higher density of input data in the model. This can leads to spatial autocorrelation (SAC) of occurrence points (Dormann et al., 2007) that inflates model accuracy (Veloz, 2009), and misleads parameter estimates (Kramer-Schadt et al. 2013).

In general, manipulating the input data (Elith et al., 2010; Phillips et al., 2009), and parametrizing the modeling method (Fithian et al., 2015; Muscarella et al., 2014) are two strategies that have been used to take into account the bias in SDM efforts. In particular, the bias caused by spatial autocorrelation could be reduced by spatial filtering (Boria et al., 2014; Kramer-Schadt et al., 2013) and background weighting schemes, the later is also called 'target-group background' (Elith et al., 2010; Phillips et al., 2009). During spatial filtering, the severity of clumpiness is decreased by removing repeated occurrence points within a specific radius around them. The idea behind background weighting comes from the fact that presence-absence models are much less affected by sampling bias compared to presence-only models (Phillips et al., 2009). It is because in presence-absence models the spatial sampling bias is reflected in both presence and absence data. Accordingly, background weighting tries to select background data (e.g. pseudo-absences) with the same bias in occurrence points. This method reduces the bias in a way that favors areas densely sampled over sparsely sampled areas (Phillips et al., 2009; Shabani et al., 2016). Elith et al. (2010) recommended this method for invasive species experiencing range shifts in invaded areas, particularly, for those with more recent invasion.

On the contrary, parametrizing SDMs to obtain a fine-tuned model is an aspect that has been poorly considered. In almost all cases, the default setting are being used to perform SDMs, especially for complex machine learning ones (Kass et al., 2021). In addition to increase the possibility of overfitting caused by noisy data (Merow et al., 2014), default setting decreases model transferability during the projection to novel environment (Guevara et al., 2018). Applying different levels of the complexity and evaluating the balance between the bias and variance of models allows to find the optimal model with a justified level of complexinty (Araujo et al., 2019; Radosavljevic and Anderson, 2014). However, among the few attempts to parametrize SDMs are tuning the best combination of the primary models to a final ensemble model (Kindt, 2018; Thuiller et al., 2009) or applying a set of input parameters to fine-tune the MaxEnt model, e.g. the package ENMeval (Muscarella et al., 2014). The development of new tools, for example h2o platform (Candel et al., 2016) or caret package (Kuhn, 2021) can bring SDM parametrization into a new focus. However, a holistic effort in which a wider range of species distribution models are fine-tuned has so far, to our knowledge, not been implemented in this arena of research.

Using SDMs is particularly pragmatic for scarce species as the results of these methods provide valuable information for their conservation implementations (Farhadinia et al., 2015; Franklin, 2010) and for identifying target areas for future sampling (Galante et al., 2018). Notwithstanding, data on scarce species

mostly suffer spatial bias due to imbalanced sampling surveys (Rebelo and Jones, 2010). In this research, we evaluated the performance of different SDMs to identify new populations of the rare species of the genus *Montivipera* in the mountains of Iran, Turkey, and Armenia. From a phylo-geographic point of view, the species of this genus due to their rapid ratio of speciation in the recent evolutionary scales show interesting forms of neo-endemism in the Near and Middle East (Behrooz et al., 2018; Stumpel et al., 2016). This genus consists of two complex group of species, *M. xanthina* complex and *M. raddei* complex. In this research, we focused on *M. raddei* complex (MRC) distributed across mountainous landscapes of northeastern Turkey, Armenia, and Iran. The northern populations of the MRC are well-described and all their potential habitats are geographically well-sampled. On the contrary, the southern ranges in Iran across the Zagros Mountains have not proportionally been sampled and some new populations of these species plus a newly defined species have just recently been identified (Behrooz et al., 2018). Accordingly, the data of the species distribution due to the different intensity and quality of sampling is geographically imbalanced-biased. Here we integrated model parametrization and data manipulation to evaluate the proficiency of four correlative SDMs including generalized linear models (GLM), gradient boosting model (GBM), random forest (RF), and maximum entropy (MaxEnt) for locating recently discovered *Montivipera* populations. Models were fine-tuned based on their intrinsic parameters and the input data was bias-corrected by implementing a background weighting procedure. We then compared the results with models of random background procedure given the new populations as out-of-bag data to test the models. In addition to AUC and TSS as two commonly-used measures of model accuracy, we also depicted the accuracy of the models across the entire gradient of suitability thresholds to provie a better understanding of the models demeanor to classify spatially imbalanced-biased data of the species.

2. Materials and Methods

2.1. *Montivipera* occurrence points

In the current research, we used the most complete dataset of MRC distribution across mountainous landscapes of Iran, eastern Turkey, and southern Armenia (Fig. 1). We collected the presence points (n = 91) throughout the entire range of the species based on direct field research, other herpetologists' field studies, and literature review. All the presence points were assessed based on the IUCN range map and spatial outliers (especially those from the literature) were excluded. Moreover, to cope with the probable spatial autocorrelation of the occurrence points we removed repeated points within a buffer of 5-km radius which remained us 82 occurrence points. The coverage of all collected records indicates apparent signs of spatial bias toward the north and northwestern Iran, eastern Turkey, and Armenia, and sparse sampling across the Zagros Mountains in western Iran, where new localities were just recently recorded. Therefore, the occurrence points were split into training and testing data, using records from northern parts for training and keeping the newly identified presence data from southern parts as independent evaluation data.

2.2. SDM development

Models were performed using data of species presence points and explanatory variables. Two sets of variables were considered; climatic and vegetation. For climatic variables, we focused on four primary variables describing annual mean and variability of climatic conditions including mean annual temperature (meananultmp), annual precipitation (anulprc), temperature seasonality (tmpseas), and precipitation seasonality (precseas) all downloaded from WorldClim dataset (Hijmans et al., 2005). For the vegetation variable, we used the enhanced vegetation index (EVI) of the MODIS products (MOD13A3), and adopted the same variability in climatic variables for vegetation, e.i., mean annual EVI and EVI seasonality. To do so, we downloaded monthly-provided 1-kilometer-resolution MOD13A3 for 2015 from EarthExplorer dataset (https://earthexplorer.usgs.gov), extracted EVI bands in ENVI version 5.1, mosaicked them to cover the entire study area in one scene, and calculated the annual mean and standard deviation of the 12 monthly-EVI rasters. We used EVI instead of commonly-used NDVI, because of its potential to minimize canopy

4

background variations and maintain sensitivity over dense vegetation conditions (Jiang et al., 2008). The EVI also copes better with residual atmosphere contamination caused by smoke and sub-pixel thin clouds. Before SDM analysis, using variance inflation factor (VIF) in 'usdm' package (Naimi, 2015), we checked the multicollinearity of variables and found no variable with VIF greater than 6.

We focused on four SDM methods, GLM, GBM, RF, and MaxEnt, because of their prevalence, well-performance, and approval over other methods (Elith and Graham, 2009; Phillips et al., 2006; Shabani et al., 2016), all implemented in R environment v 3.5.2. We first splitted MRC occurrence points to training and test data and followed a crossvalidation scheme on the training datast to fit the models. This cross-validated scheme of training data was kept constant for tuning the preliminary models and running the final version of the models. GLM, GBM and RF were tuned up using the 'caret' package (Kuhn, 2021) with considering different model-specific parameters and the best-fitted model across the then folds was identified according to their ROC scores. The fine-tuned model with the highest accuracy was then used to generate the habitat suitability mpas and to predict the test dataset. The GLM was performed using simple and quadratic terms of explanatory variables and the model selection was based on a stepwise AIC selection procedure. GBM was fitted with allowing the maximum number of trees up to 2000, with three learning rates (i.e., shrinkage; 0.001, 0.01, 0.1), three interaction depths (i.e., complexity of the tree, maximum nodes per tree; 1, 3, and 5), and three values for subsampling fraction (i.e., bag fraction; 0.5, 0.65, and 0.8). The RF model was fitted with number of trees (ntrees) 500 and 1000, number of variables randomly selected at each split (i.e., mtry) 1 to 5, and node size 1 and 5. The MaxEnt model was tuned up using the package ENMeval (Muscarella et al., 2014) with allowing five combination of feature types (fc = L, LQ, LQH, LQHP, and LQHPT) and regularization multiplier (rm) of 0.5, 1, 1.5, and 2. The best fitted parameters for each model were then used to predict to the environmental layers and to generate the corresponding habitat suitability maps. Again the generation of habitat suitability maps was carried on given the constant 10-fold crossvalidation of the occurrence points, meaning that 10 habitat suitability map was predicted for each model. The final ensemble habitat suitability map of the four SDM algorithem was generated based on a proportionaly weighted average of the obtained AUC score of the then repetition.

To address the purpose of our study, background data were selected in two different ways including random and background weighting. For random we selected 10000 background points spatially at random leaving cells with MRC occurrence points. To create background weighting data by generating a weighting surface we gave prominence to those areas having less geographical proximity to others. Following Elith et al. (2010) we first generated a density raster map from the occurrence points and then allocated 10000 background points regarding its probability distribution (Fig. 1). This method copes with the bias caused by the spatially imbalanced-biased data in a way that favors occurrence points of severely sampled areas over those of sparsely sampled areas (Shabani et al., 2019). Of the 82 occurrence points of MRC, we used 12 newly sampled records as out-of-bag data to test models' performance. We used the area under the curve (AUC) of the receiver operating characteristic (ROC) plot to assess the discrimination capacity of models. AUC combines specificity and sensitivity (Fielding and Bell, 1997), thus, neglects the relative costs of errors of omission and commission (Jimenez-Valverde, 2012). Therefore, we also computed true statistic skill (TSS) as a threshold-dependent measure of classification accuracy calculated as sensitivity + specificity − 1. We used the package 'PresenceAbsence' to evaluate the performance of the models and the threshold '10 percintile of training suitability' was set to calculate the threshold-dependent measures. In Addition to these traditionally-used metrics which give an absolute measure of the model performance, we plotted sensitivity and specificity of the models against an ascending gradient of 100 thresholds to obtain more informative inferences on the models predictive performance.

3. Results

In total, for each of the random and background weighting schemes we fitted 128 preliminary models based on the 10 cross-validated folds of the training occurrence points. Since the GLM is inherently a simple

5

algorithm, for this method only one set of parameters was trained. Although for both traning and test dataset the AUC and TSS of this model in BkWt scheme decreased in comparison to random scheme (Table 1 & 2), it successfully classified all the test data, i.e., sensitivity = 1, by using the BkWt scheme. Totally, GLM obtained AUC of 0.92 and 0.89 and TSS of 0.66 and 0.65 for random and BkWt schemes, respectively (Table 1). From the multiple combination of the GBM parameters, for the random background selection a model characterized by shrinkage = 0.01, interaction depth = 5, and ntrees = 1800 showed highest ROC value (ROC = 1, sensitivity = 1, and specificity = 0.85). For BkWt, the fine-tuned GBM model (ROC = 1, sensitivity = 1, and specificity = 0.64) had shrinkage = 0.01, interaction depth = 5, and ntrees = 2000. Although for both training and test datasets the AUC and TSS of this model were almost equal (AUC = 0.976 and 0.971 and TSS = 0.59 and 0.58 for random and BkWt schemes, respectively), it lost the ability to truly predict presence points, i.e., sensitivity, compared to othe models (Table 2). For the RF model, the fine-tuned model of both random and BkWt schemes was charachterized by ntrees = 1000, and nodesize = 1. However, the mtry was 2 and 3 for random and Bkwt schemes, respectively. Similar to the GBM, RF obtained almost equal AUC and TSS scores for both training and test datasets (AUC = 0.97 and 0.96 and TSS = 0.45 and 0.58 for random and BkWt schemes, respectively), but the sensitivity of this model was low (Table 2). For the MaxEnt model the best-fitted model with the highest AIC_w of the ENMeval analysis obtained rm 0.5 and 1.5, and fc LQ and LQHP for random and BkWt schemes, respectively. In the final habitat suitability maps of the MaxEnt model, the test data had AUC 0.93 and 0.95 and TSS 0.66 and 0.80 for random and BkWt schemes, respectively (Table 2). Overall, the highest TSS score of the test data was obtained in the MaxEnt model that was fitted based on the BkWt background selection scheme (Table 2).

The predicted suitability maps are shown in Fig. 2. We found a good consistency between the patterns of occurrence points and suitable habitats. Comparing the spatial pattern of suitable habitats in random and background weighting methods showed that all models represented different results except for MaxEnt model in which comparable results were obtained (Fig. 2). Accordingly, we calculated the correlation coefficient between the two background selection schemes of the four SDM methods revealing that the highest correlation was obtained for MaxEnt model ($r = 0.85$), followed by GLM ($r = 0.61$), GBM ($r = 0.45$), and RF ($r = 0.42$). These findings were also confirmed by sensitivity and specificity graphs (Fig. 3). We found that while the capacity of GBM and RF to predict the training and test background points (i.e. models' specificity) was maintained excellent even at higher thresholds, their capability to predict presence data (i.e. models' sensitivity) were reduced at lower thresholds. On the contrary, GLM and MaxEnt models showed good performances to predict presence data, but lost their capacity to classify background data at lower thresholds especially in the BkWt scheme (Fig. 3). The comparison of the response curves of the variables between the two bakhground selection schemes indicated identical pattern, however, for GBM and RF the response curves of the BkWt scheme were more rugged compared to the smoother variation in the random background selection scheme (Fig. 4).

4. Discussion

In the present study, the performance of different models in predicting spatially biased occurrence data was evaluated. For this purpose, modeling was performed based four SDMs. A variety of correlative distribution models have been emerged in recent decades and their performance has been compared in numerous studies (Elith and Graham, 2009; Merow et al., 2014). These methods, however, are challenged by a number of methodological problems that make their comparison controversial. Among these issues are providing a balance between goodness-of-fit and model complexity (Araujo et al., 2019; Warren and Seifert, 2011), and the spatial bias of the input data and manipulating them for evaluating model performance (Hijmans, 2012; Phillips et al., 2009). Generally, for most SDMs, particularly for complex machine learning ones, using a set of default parameters have been recommended based on a comprehensive model tuning [for example see Phillips & Dudik (2008) for the MaxEnt and Elith et al., (2008) for boosted regression trees]. However, a blindfold utilization of them may come up with a poorly performing model (Muscarella et al., 2014). Equally important, is the probable SAC in the input occurrence points which may result in an inflated score

6

of the metrics used for model valuation (Dormann et al., 2007; Radosavljevic and Anderson, 2014). Another problem is the paucity of operative procedures for tuning SDMs that are efficiently effortless and time-saving. Consequently, in most SDM efforts, while the species, scale, and area of interest are different, the modeling relies on default setting of the related model.

In the current research, we employed a novel approach that in one hand applyed model-specific parameterization, and on the other hand, manipulated a spatially imbalanced-biased input data to cope with the above-mentioned issues. In general, the results showed that the models have high predictive performance based on AUC and TSS values. However, the geographical pattern of the predicted suitable habitats among models was different. For example, the area of suitable habitats was greater in GLM in comparison with other SDMs. Accordingly, for this model lower values of specificity were observed across the different gradients of suitability thresholds (red curves in Fig. 3). This is intrinsically due to the simple regression-based nature of this model in which a basic assumption is the normality of the error distribution and its constant variance (Osborne and Waters, 2002). Thus, those data that are not used as training are well-fitted in the predicted model. On the other hand, this makes GLM prone to over-prediction across non-sampled areas leading to lower specificity, i.e., true negative rate, of this model.

On the contrary, we found the best training sensitivity and specificity for decision-tree-based methods, i.e., GBM and RF. These models, due to the automatic promotion caused by the model learner, attempt to improve task classification, as much as possible, leading to the highest discrimination capacity of training presence/absence dataset (De'ath and Fabricius, 2000; Friedman, 2017). In our case, results indicated that the RF models for both random and background weighting procedures, while excellently classified the training dataset, failed to predict out-of-bag ones. Fundamentally, RF and GBM depend on bagging and boosting algorithms of the tree learner, respectively (Elith et al., 2008; Friedman, 2017). In RF the bagging algorithm, also called bootstrap aggregation, allows the tree learner repeatedly select a random sample with replacement of the training set and fits trees to these samples (Breiman, 2001). The trees in RF are run in parallel and there is no interaction between them while the trees are built. Once all the trees are built, then an average is taken across all the trees' predictions (Cutler et al., 2012). Conversely, the trees in boosting algorithms, e.g. GBM, are trained sequentially, and accordingly, weaker results are boosted or reweighted over many iterations to have the learner focus more on areas it got wrong, and less on those observations that were correct (Elith et al., 2008; Friedman, 2017). The risk of deeply-grown trees in random forests comes at the expense of overfitting the training data set in which predictions have low bias but high variances. In our modeling approach, we grew the RF model to 1000 trees which is reasonable and comparable to the similar SDM efforts (Breiman, 2001; Shabani et al., 2016). Altough based on a cross-validated splitting of the training data we fine-tuned the RF model, it appears that this model still suffers from high variance. In fact, the strict classification algorithm of the RF model is prone to return low capabilities in predicting out-of-bag dataset (Elith et al., 2008; Merow et al., 2014). Accordingly, the use of this method in extrapolating predictions should be taken into account prudently.

Given the random background approach, our results highlight the excellent predictive performance of the GBM model in which true positive rate, i.e. sensitivity, and true negative rate, i.e. specificity, remained well-justified across accumulative thresholds. In contrast to the paralleled bagging method of the RF, it is believed that the boosting manner of the tree learner in GBM reduce the probability of overfitting and allows well-classifying out-of-bag samples (Elith et al., 2008; Shabani et al., 2016). However, our study revealed that fitting a GBM model based on a background weighting approach significantly reduces the predictive performance of this model for identifying out-of-bag data. This highlights the inefficiency of this method for being used in SDM efforts with imbalanced-biased data where the primary goal is finding probability distribution over areas that are not comparably sampled.

Unlike the decision-tree-based methods, MaxEnt resulted in a comparable prediction based on both random and background weighting approaches. This was similar to the results of the GLM model except that for the latter true negative rates were misplaced at lower thresholds. As mentioned before, the normally-distributed errors and no trends in residuals relative to the fitted values (Hardin et al., 2007) allow the GLM to be

interpretably efficient for predicting out-of-bag data. In MaxEnt, as a density estimator algorithm, the species distribution is represented by a probability distribution that is closest to uniform (Phillips et al., 2006). This probability distribution is bounded by a set of constraints that are simple functions of the explanatory variables, called 'features', and derived from the species occurrence locations. The primary assumption of the MaxEnt model is that the mean of each feature is required to be close (within same error bounds) to the empirical average over the presence sites (Phillips and Dudik, 2008). From a general point of view, this constraint in MaxEnt could be assumed equivalent to the consistency of error variance in the GLM method, and as a consequence, bringing consistent results for out-of-bag data is also expectable in the MaxEnt model. In both GLM and MaxEnt models a maximum likelihood is used to estimate a parametric exponential distribution of linear combination of features (Phillips et al. 2004). Although GLM could be fitted by considering quadratic and interactive terms of the explanatory variables, more variation of feature types in MaxEnt allows fitting more complex models (Phillips et al. 2006). More importantly, while GBM and RF as complex machine learning methods are more prone to overfitting, the regularization multiplier in MaxEnt prevents the model to match the input data too closely. Altogether, our results highlight the efficiency of the MaxEnt model in using the benefits of different modeling methods to produce results that are both predictable (extrapolative) and complex (interpolative).

Asymmetry in spatially biased model predictions also highlights the need to evaluate model performance using threshold-dependent sensitivity (true positive rate) and specificity (true negative rate) in addition to threshold-independent AUC. As criticized by Lobo et al. (2008) and Jimenez-Valverde (2012), modeling goals and setting highly influence the appropriateness of the AUC for measuring the performance of a model. AUC inflates the number of false absence data (Lobo et al., 2008), and accordingly, over-represents predictive performance for rare species (Phillips et al., 2009; Stolar and Nielsen, 2015), as in the case of *Montivipera* species. Moreover, being only as a discrimination measure, AUC doesn't show goodness-of-fit, i.e. classification accuracy of the model, and consequently, a model with high AUC value is not necessarily a well-fitted one (Jimenez-Valverde, 2012). For threshold dependent measures a critical trick is selecting the best suitability threshold at which the sensitivity and specificity of the resulted model are well-balanced. Although several thresholds have been suggested to do this, for example see Liu et al. (2005) as a review, a single suitability threshold provides only a cross-section of the model performance and doesn't provide a comprehensive perception of the classification accuracy of the model across a gradient of suitability threshold. For example, the threshold at which the sensitivity is equal to the specificity, i.e., where their curves cross on the Fig. 3, is among the widely used suitability thresholds, nevertheless, the corresponding accuracy measure doesn't inclusively specify the performance of the model to classify presence and background data.

There are several ways to correct sampling bias, some of which cannot be used in cases where data is scarce. Spatial filtering may not be helpful when there are only a few presence points (Phillips et al., 2009). Decreasing clumpiness reduces training sample size and, depending on the heterogeneity of the surrounding environment and the selected spatial resolution, it may drop some of the information on the species occupation sites. In addition to spatial filtering and background weighting, a third method called model-based bias correction has been used (El-Gabbas and Dormann, 2018a) to address spatial bias in occurrence data. In this method other environmental variables, used as bias covariates, characterize potential sources of sampling bias. Although this method is confirmed to be useful when dealing with sparse datasets (El-Gabbas and Dormann, 2018b), it is highly dependent on the selected bias variable that, in turn, intensifies upstream uncertainty caused by assisting covariates.

It is worth mentioning that bias adjustment and model parameterization depending on the prevalence of the target species result in varying spatial predictions (Araujo et al., 2019; Pottier et al., 2013). This variation is most noticeable for common species, and thus, distribution models of rare species that are habitat specialists may not be very sensitive to spatially-biased occurrence data (Stolar and Nielsen, 2015). Being limited to a narrow gradient of environmental conditions, specialist species are thus more predictable as well as more distinguishable, i.e. high values of AUC of their SDMs, because of the high distinctiveness between their occurrence points and background space. In our case, this tendency was more obvious where narrow-ranged mountain vipers (Ahmadi et al., 2019) obtained high scores of AUC and TSS. This, to a high extend, justify

our SDM approach where, due to the sparse data of the mountain vipers, their occurrence points were pooled into one set. Since they belong to distinct taxonomic levels, e.g. species or sub-species levels, the resulted SDMs might challenge niche equilibrium assumption (Wiens et al., 2009) and be prone to a inflated niche breadth (Pearman et al., 2010) where the resulted distribution models show higher levels of over-estimation. However, niche inflation is more challenging for general species with abundant data (Randin et al., 2006). Moreover, the narrow-ranged mountain vipers, in general, and the species/sub-species belonging to the Raddei clade, in specific, show low rates of niche evolution and high degrees of niche conservatism (Ahmadi et al., 2021) that leads to the occupation of similar ecological conditions in this species.

5. Conclusion

Spatial bias of the input data is one of the main sources of uncertainty in the species distribution modeling approaches. This issue is particularly important for scarce species with geographically imbalanced biased data of their distribution ranges. Despite the great emphasis on the importance of model tuning and input data manipulation in improving SDMs, the performance of different models in using such an approach has not received much attention. In this research we evaluated the performance of four commonly-used SDMs to predict imbalanced biased occurrence points based on two methods of background data selection including random and background weighting. Our result reveals that different models produced dissimilar results for two background selection schemes. Complex GBM and RF models, due to their interpolative conception, showed inefficiency in predicting test points, especially for the background weighting mode. The GLM over-predicted presence areas due to its extrapolative nature. In spite of being a machine learning method, MaxEnt show a comparable performance in predicting test points in two background selection schemes. The results of the present study emphasize the proficiency of MaxEnt model in generating reproducible comparisons particularly when the input data is being completed.

Acknowledgments

Conflict of Interest

The authors declare no conflict of interest.

Author contribution

Mohsen Ahmadi: Conceptualization (lead), Formal analysis (lead), Investigation (lead), Methodology (lead), Writing—original draft (lead), Writing—review & editing (lead). **Mahmoud-Reza Hemami:** Conceptualization (supporting), Methodology (supporting), Writing—original draft (supporting), Writing—eview & editing (equal). **Mohammad Kaboli:** Conceptualization (supporting), Methodology (supporting), Writing—original draft (supporting), Writing—eview & editing (equal). **Farzin Shabani:** Conceptualization (supporting), Formal analysis (supporting), Investigation (supporting), Methodology (supporting), Writing—original draft (supporting), Writing—review & editing (equal).

Data Availability Statement

The data and the R code that support the findings of this study are openly available in github at https://github.com/mahmadi-iut/SDM_Tuning.git.

References

Ahmadi, M., Hemami, M.-R., Kaboli, M., Malekian, M., Zimmermann, N.E., 2019. Extinction risks of a Mediterranean neo-endemism complex of mountain vipers triggered by climate change. Scientific Reports 9, 1-12.

Ahmadi, M., Hemami, M.-R., Kaboli, M., Nazarizadeh, M., Malekian, M., Behrooz, R., Geniez, P., Alroy, J., Zimmermann, N.E., 2021. The legacy of Eastern Mediterranean mountain uplifts: rapid disparity of phylogenetic niche conservatism and divergence in mountain vipers. BMC Ecology and Evolution 21, 1-13.

Ahmadi, M., Naderi, M., Kaboli, M., Nazarizadeh, M., Karami, M., Beitollahi, S.M., 2018. Evolutionary applications of phylogenetically-informed ecological niche modelling (ENM) to explore cryptic diversification over cryptic refugia. Molecular Phylogenetics and Evolution 127, 712-722.

Ahmadzadeh, F., Flecks, M., Carretero, M.A., Bohme, W., Ihlow, F., Kapli, P., Miraldo, A., Rodder, D., 2016. Separate histories in both sides of the Mediterranean: phylogeny and niche evolution of ocellated lizards. Journal of Biogeography 43, 1242–1253.

Araujo, M.B., Anderson, R.P., Barbosa, A.M., Beale, C.M., Dormann, C.F., Early, R., Garcia, R.A., Guisan, A., Maiorano, L., Naimi, B., 2019. Standards for distribution models in biodiversity assessments. Science Advances 5, eaat4858.

Araujo, M.B., Guisan, A., 2006. Five (or so) challenges for species distribution modelling. Journal of Biogeography 33, 1677-1688.

Behrooz, R., Kaboli, M., Arnal, V., Nazarizadeh, M., Asadi, A., Salmanian, A., Ahmadi, M., Montgelard, C., 2018. Conservation Below the Species Level: Suitable Evolutionarily Significant Units among Mountain Vipers (the Montivipera raddei complex) in Iran. Journal of Heredity 109, 416-425.

Boria, R.A., Olson, L.E., Goodman, S.M., Anderson, R.P., 2014. Spatial filtering to reduce sampling bias can improve the performance of ecological niche models. Ecological modelling 275, 73-77.

Boucher, F.C., Zimmermann, N.E., Conti, E., 2015. Allopatric speciation with little niche divergence is common among Alpine Primulaceae. Journal of Biogeography 43, 591-602.

Breiman, L., 2001. Random forests. Machine learning 45, 5-32.

Brito, J.C., Acosta, A.L., Alvares, F., Cuzin, F., 2009. Biogeography and conservation of taxa from remote regions: an application of ecological-niche based models and GIS to North-African Canids. Biological Conservation 142, 3020-3029.

Candel, A., Parmar, V., LeDell, E., Arora, A., 2016. Deep learning with H2O. H2O. ai Inc, 1-21.

Cutler, A., Cutler, D.R., Stevens, J.R., 2012. Random forests, Ensemble machine learning. Springer, pp. 157-175.

De'ath, G., Fabricius, K.E., 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. Ecology 81, 3178-3192.

Dormann, C.F., McPherson, J.M., Araujo, M.B., Bivand, R., Bolliger, J., Carl, G., Davies, R.G., Hirzel, A., Jetz, W., Kissling, W.D., 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. Ecography 30, 609-628.

El-Gabbas, A., Dormann, C.F., 2018a. Improved species-occurrence predictions in data-poor regions: using large-scale data and bias correction with down-weighted Poisson regression and Maxent. Ecography 41, 1161-1172.

El-Gabbas, A., Dormann, C.F., 2018b. Wrong, but useful: regional species distribution models may not be improved by range-wide data under biased sampling. Ecology and evolution 8, 2196-2206.

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

Elith, J., H. Graham, C., P. Anderson, R., Dudik, M., Ferrier, S., Guisan, A., J. Hijmans, R., Huettmann, F., R. Leathwick, J., Lehmann, A., Li, J., G. Lohmann, L., A. Loiselle, B., Manion, G., Moritz, C., Nakamura,

M., Nakazawa, Y., McC. M. Overton, J., Townsend Peterson, A., J. Phillips, S., Richardson, K., Scachetti-Pereira, R., E. Schapire, R., Soberon, J., Williams, S., S. Wisz, M., E. Zimmermann, N., 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography 29, 129-151.

Elith, J., Kearney, M., Phillips, S., 2010. The art of modelling range-shifting species. Methods in ecology and evolution 1, 330-342.

Elith, J., Leathwick, J.R., Hastie, T., 2008. A working guide to boosted regression trees. Journal of Animal Ecology 77, 802-813.

Farhadinia, M.S., Ahmadi, M., Sharbafi, E., Khosravi, S., Alinezhad, H., Macdonald, D.W., 2015. Leveraging trans-boundary conservation partnerships: Persistence of Persian leopard (Panthera pardus saxicolor) in the Iranian Caucasus. Biological Conservation 191, 770-778.

Fielding, A.H., Bell, J.F., 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environmental conservation 24, 38-49.

Fithian, W., Elith, J., Hastie, T., Keith, D.A., 2015. Bias correction in species distribution models: pooling survey and collection data for multiple species. Methods in Ecology and Evolution 6, 424-438.

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

Friedman, J.H., 2017. The elements of statistical learning: Data mining, inference, and prediction. springer open.

Galante, P.J., Alade, B., Muscarella, R., Jansa, S.A., Goodman, S.M., Anderson, R.P., 2018. The challenge of modeling niches and distributions for data-poor species: a comprehensive approach to model complexity. Ecography 41, 726-736.

Guevara, L., Gerstner, B.E., Kass, J.M., Anderson, R.P., 2018. Toward ecologically realistic predictions of species distributions: A cross-time example from tropical montane cloud forests. Global change biology 24, 1511-1522.

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

Guisan, A., Graham, C.H., Elith, J., Huettmann, F., Group, N.S.D.M., 2007. Sensitivity of predictive species distribution models to change in grain size. Diversity and distributions 13, 332-340.

Hardin, J.W., Hardin, J.W., Hilbe, J.M., Hilbe, J., 2007. Generalized linear models and extensions. Stata press.

Hemami, M.-R., Esmaeili, S., Brito, J.C., Ahmadi, M., Omidi, M., Martinez-Freiria, F., 2018. Using ecological models to explore niche partitioning within a guild of desert felids. Hystrix, the Italian Journal of Mammalogy 29, 216-222.

Hijmans, R.J., 2012. Cross-validation of species distribution models: removing spatial sorting bias and calibration with a null model. Ecology 93, 679-688.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A., 2005. Very high resolution interpolated climate surfaces for global land areas. International journal of climatology 25, 1965-1978.

Hortal, J., Jimenez-Valverde, A., Gomez, J.F., Lobo, J.M., Baselga, A., 2008. Historical bias in biodiversity inventories affects the observed environmental niche of the species. Oikos 117, 847-858.

Jiang, Z., Huete, A.R., Didan, K., Miura, T., 2008. Development of a two-band enhanced vegetation index without a blue band. Remote sensing of Environment 112, 3833-3845.

Jimenez-Valverde, A., 2012. Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modelling. Global Ecology and Biogeography 21, 498-507.

Kass, J.M., Muscarella, R., Galante, P.J., Bohl, C.L., Pinilla-Buitrago, G.E., Boria, R.A., Soley-Guardia, M., Anderson, R.P., 2021. ENMeval 2.0: Redesigned for customizable and reproducible modeling of species' niches and distributions. Methods in Ecology and Evolution 12, 1602-1608.

Kindt, R., 2018. Ensemble species distribution modelling with transformed suitability values. Environmental Modelling & Software 100, 136-145.

Kramer-Schadt, S., Niedballa, J., Pilgrim, J.D., Schroder, B., Lindenborn, J., Reinfelder, V., Stillfried, M., Heckmann, I., Scharf, A.K., Augeri, D.M., Cheyne, S.M., Hearn, A.J., Ross, J., Macdonald, D.W., Mathai, J., Eaton, J., Marshall, A.J., Semiadi, G., Rustam, R., Bernard, H., Alfred, R., Samejima, H., Duckworth, J.W., Breitenmoser-Wuersten, C., Belant, J.L., Hofer, H., Wilting, A., 2013. The importance of correcting for sampling bias in MaxEnt species distribution models. Diversity and Distributions 19, 1366–1379.

Kuhn, M., 2021. caret: Classification and Regression Training. R package version 6.0-90. https://CRAN.R-project.org/package=caret.

Lentini, P.E., Wintle, B.A., 2015. Spatial conservation priorities are highly sensitive to choice of biodiversity surrogates and species distribution model type. Ecography 38, 1101-1111.

Liu, C., Berry, P.M., Dawson, T.P., Pearson, R.G., 2005. Selecting thresholds of occurrence in the prediction of species distributions. Ecography 28, 385-393.

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

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

Muscarella, R., Galante, P.J., Soley-Guardia, M., Boria, R.A., Kass, J.M., Uriarte, M., Anderson, R.P., 2014. ENM eval: An R package for conducting spatially independent evaluations and estimating optimal model complexity for Maxent ecological niche models. Methods in ecology and evolution 5, 1198-1205.

Naimi, B., 2015. usdm: uncertainty analysis for species distribution models. R package version 1.1-15, R Documentation http://www.rdocu-mentation.org/packages/usdm.

Osborne, J.W., Waters, E., 2002. Four assumptions of multiple regression that researchers should always test. Practical assessment, research, and evaluation 8, 2.

Pearman, P.B., D'Amen, M., Graham, C.H., Thuiller, W., Zimmermann, N.E., 2010. Within-taxon niche structure: niche conservatism, divergence and predicted effects of climate change. Ecography 33, 990-1003.

Phillips, S.J., Anderson, R.P., Schapire, R.E., 2006. Maximum entropy modeling of species geographic distributions. Ecological modelling 190, 231-259.

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

Phillips, S.J., Dudik, M., Elith, J., Graham, C.H., Lehmann, A., Leathwick, J., Ferrier, S., 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. Ecological applications 19, 181-197.

Pottier, J., Dubuis, A., Pellissier, L., Maiorano, L., Rossier, L., Randin, C.F., Vittoz, P., Guisan, A., 2013. The accuracy of plant assemblage prediction from species distribution models varies along environmental gradients. Global Ecology and Biogeography 22, 52-63.

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

Randin, C.F., Dirnbock, T., Dullinger, S., Zimmermann, N.E., Zappa, M., Guisan, A., 2006. Are niche-based species distribution models transferable in space? Journal of biogeography 33, 1689-1703.

Rebelo, H., Jones, G., 2010. Ground validation of presence-only modelling with rare species: a case study on barbastelles Barbastella barbastellus (Chiroptera: Vespertilionidae). Journal of Applied Ecology 47, 410-420.

Rocchini, D., Hortal, J., Lengyel, S., Lobo, J.M., Jimenez-Valverde, A., Ricotta, C., Bacaro, G., Chiarucci, A., 2011. Accounting for uncertainty when mapping species distributions: the need for maps of ignorance. Progress in Physical Geography 35, 211-226.

Saladin, B., Thuiller, W., Graham, C.H., Lavergne, S., Maiorano, L., Salamin, N., Zimmermann, N.E., 2019. Environment and evolutionary history shape phylogenetic turnover in European tetrapods. Nature communications 10, 249.

Shabani, F., Ahmadi, M., Peters, K.J., Haberle, S., Champreux, A., Saltre, F., Bradshaw, C.J., 2019. Climate-driven shifts in the distribution of koala-browse species from the Last Interglacial to the near future. Ecography 42, 1587-1599.

Shabani, F., Kumar, L., Ahmadi, M., 2016. A comparison of absolute performance of different correlative and mechanistic species distribution models in an independent area. Ecology and evolution 6, 5973-5986.

Stolar, J., Nielsen, S.E., 2015. Accounting for spatially biased sampling effort in presence-only species distribution modelling. Diversity and Distributions 21, 595-608.

Stumpel, N., Rajabizadeh, M., Avcı, A., Wüster, W., Joger, U., 2016. Phylogeny and diversification of mountain vipers (Montivipera, Nilson et al., 2001) triggered by multiple Plio–Pleistocene refugia and high-mountain topography in the Near and Middle East. Molecular phylogenetics and evolution 101, 336-351.

Tessarolo, G., Rangel, T.F., Araújo, M.B., Hortal, J., 2014. Uncertainty associated with survey design in Species Distribution Models. Diversity and Distributions 20, 1258-1269.

Thuiller, W., Lafourcade, B., Engler, R., Araújo, M.B., 2009. BIOMOD–a platform for ensemble forecasting of species distributions. Ecography 32, 369-373.

Thuiller, W., Lavergne, S., Roquet, C., Boulangeat, I., Lafourcade, B., Araujo, M.B., 2011. Consequences of climate change on the tree of life in Europe. Nature 470, 531-534.

Thuiller, W., Richardson, D.M., PYŠEK, P., Midgley, G.F., Hughes, G.O., Rouget, M., 2005. Niche-based modelling as a tool for predicting the risk of alien plant invasions at a global scale. Global Change Biology 11, 2234-2250.

Tingley, R., Vallinoto, M., Sequeira, F., Kearney, M.R., 2014. Realized niche shift during a global biological invasion. Proceedings of the National Academy of Sciences 111, 10233-10238.

Vale, C.G., Tarroso, P., Brito, J.C., 2014. Predicting species distribution at range margins: testing the effects of study area extent, resolution and threshold selection in the Sahara–Sahel transition zone. Diversity and Distributions 20, 20-33.

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

Waltari, E., Guralnick, R.P., 2009. Ecological niche modelling of montane mammals in the Great Basin, North America: examining past and present connectivity of species across basins and ranges. Journal of Biogeography 36, 148-161.

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

Wiens, J.J., Stralberg, D., Jongsomjit, D., Howell, C.A., Snyder, M.A., 2009. Niches, models, and climate change: assessing the assumptions and uncertainties. Proceedings of the National Academy of Sciences 106, 19729-19736.

Wisz, M.S., Hijmans, R., Li, J., Peterson, A.T., Graham, C., Guisan, A., Group, N.P.S.D.W., 2008. Effects of sample size on the performance of species distribution models. Diversity and distributions 14, 763-773.

Yousefi, M., Ahmadi, M., Nourani, E., Rezaei, A., Kafash, A., Khani, A., Sehhatisabet, M.E., Adibi, M.A., Goudarzi, F., Kaboli, M., 2017. Habitat suitability and impacts of climate change on the distribution of wintering population of Asian Houbara Bustard Chlamydotis macqueenii in Iran. Bird Conservation International 27, 294-304.

Zupan, L., Cabeza, M., Maiorano, L., Roquet, C., Devictor, V., Lavergne, S., Mouillot, D., Mouquet, N., Renaud, J., Thuiller, W., 2014. Spatial mismatch of phylogenetic diversity across three vertebrate groups and protected areas in Europe. Diversity and Distributions 20, 674-685.

Tables

Table 1 The average performance of the models across 10 crossvalidated folds of the training dataset calculated for random and background weighting (BkWt) schemes. The threshold-dependent measures were calculated given a 10 percentile of suitability score at the training occurrence points.

	Sensitivity	Sensitivity	Specificity	Specificity	AUC	AUC	TSS	TSS
	Random	BkWt	Random	BkWt	Random	BkWt	Random	BkWt
GBM	0.9	0.9	0.955	0.921	0.98	0.964	0.855	0.821
GLM	0.9	0.9	0.894	0.681	0.947	0.838	0.794	0.581
MaxEnt	0.9	0.9	0.887	0.683	0.947	0.828	0.787	0.583
RF	0.9	0.9	0.984	0.956	0.994	0.979	0.884	0.856

Table 2 Results of the model evaluation for random and background weighting (BkWt) schemes. Models were evaluated based on the 12 newly sampled occurrence points of the MRC. The threshold-dependent measures were calculated given a 10 percentile of suitability score at the training occurrence points.

	Sensitivity	Sensitivity	Specificity	Specificity	AUC	AUC	TSS	TSS
	Random	BkWt	Random	BkWt	Random	BkWt	Random	BkWt
GBM	0.636	0.636	0.953	0.947	0.976	0.971	0.589	0.583
GLM	0.818	1	0.84	0.657	0.917	0.893	0.658	0.647
MaxEnt	0.818	1	0.843	0.804	0.927	0.949	0.661	0.804
RF	0.455	0.636	0.991	0.945	0.97	0.957	0.446	0.581

Figure legends

Figure 1 Geographic distribution of MRC occurrence and background-weighted points. Background-weighted points were allocated based on a probability distribution of a density raster of training occurrence points.

Figure 2 Habitat suitability map of the tuned-up models derived from random (Rndm) and background weighting (BkWt) schemes.

Figure 3 Variation of sensitivity (Sens); the proportion of correctly predicted presence data, and specificity (Spec); the proportion of correctly predicted pseudo-absence data, across a gradient of suitability thresholds. Dashed lines indicates test data. Rndm: random and BkWt: background weighting schemes. The black diamonds indicate the corresponding threshold of the 10 percentile suitability of the training dataset.

Figure 4 Response curve of the explanatory variables derived from models obtained based on the random (top) and background weighting (bottem) schemes. Models were fitted in biomod2 package.

(GBM-Rndm)

(GBM-BkWt)

(GLM-Rndm)

(GLM-BkWt)

(MaxEnt-Rndm)

(MaxEnt-BkWt)

(RF-Rndm)

(RF-BkWt)

0 150 300 km

Habitat suitability
High : 1

Low : 0