

Optimal model selection for Maxent: a case of freshwater species distribution modelling in Bhutan, a data poor country

Tshering Dorji¹, Simon Linke², and Fran Sheldon²

¹Royal University of Bhutan College of Natural Resources

²Australian Rivers Institute Nathan campus

November 16, 2020

Abstract

Maxent is commonly used species distribution modelling (SDM) program due to its better performance over other SDM programs. But model complexity and selecting optimal models are two important concerns for Maxent users. In order to help advance the field we built 44 sets of models by combining 11 regularization multipliers and four feature classes for 10 fish and 28 odonate species of Bhutan with small occurrence data. We then selected optimal models using four sequential optimal model selection approaches: two OR_{TEST} approaches which combined threshold dependent test omission rate (OR) followed by area under receiver operating curve for test data (AUC_{TEST}), and two AUCDIFF approaches that combined OR followed by difference between training AUC and AUC_{TEST} (AUC_{DIFF}) and then AUC_{TEST}. We then screened for ecologically plausible binary suitable/unsuitable model for each species among the optimal models selected by the sequential approaches or from the remaining models using expert knowledge (EXP approach). We then compared different model features and the predicted binary habitat of the optimal models selected by the five approaches. Models selected by OR_{TEST} approaches matched better with ones selected by EXP approach despite them selecting more complex models compared to AUC_{DIFF} approaches. Further, models selected through AUCDIFF approaches overpredicted the habitat more often than the models selected through OR_{TEST} approaches when compared to models chosen by EXP approach. We recommend use of OR_{TEST} approaches for model selection either as the first line of model screening or by their own when less restrictive thresholds are used to produce binary habitat maps as we did here. First, this would reduce time required for expert screening of multiple models for ecologically plausible models when many species are studied. Second, when used alone, OR_{TEST} approaches can avoid either selecting models that under predict or over predict the suitable habitat.

Introduction

Species distribution models (SDMs) are used increasingly in different stages of conservation decision making (Guisan et al. 2013). Among numerous SDM methods, techniques that use presence-only occurrence data have become more popular with readily available georeferenced presence data (for example from GBIF) and environmental variables (Gomez et al. 2018). Among the presence only modelling methods Maxent has become very popular (Phillips et al. 2017, Morales, Fernández, and Baca-González 2017). Maxent identifies the suitable geographic areas for species given the set of environmental variables and known occurrence records by applying a maximum entropy model (Phillips and Dudík 2008). The increased use of Maxent has been ascribed to its better performance over other methods when used with low occurrence data (Elith et al. 2011, Coxen, Frey, Carleton, and Collins 2017) and also for its ease of use through its graphical user interface (GUI) (Phillips et al. 2006, Morales et al. 2017, Kass et al. 2018).

The Maxent output can be overfit or under fit to the occurrence localities which results in models that underpredict or overpredict the suitable area respectively (Shcheglovitova and Anderson 2013). Maxent output quality depends on model complexity (Shcheglovitova and Anderson 2013, Morales et al. 2017, Phillips 2017), run type (Phillips 2017), bias in occurrence data (Phillips et al. 2009, Syfert, Smith, and Coomes 2013) and its

correction methods (Kramer-Schadt et al. 2013), background selection methods (Merow, Smith, and Silander Jr 2013; Vollerling, Halvorsen, Auestad, and Rydgren 2019), and output types (Phillips 2017, Phillips et al. 2017). Once other factors are taken into consideration, model complexity becomes the most important issue (Syfert, Smith, and Coomes 2013). Model complexity is controlled by the types of feature class (hereafter: ‘FC’) and the value of the regularization multiplier (hereafter: ‘RM’) used (Radosavljevic and Anderson 2014, Morales et al. 2017, Phillips 2017). The current version of Maxent v 3.4.1 has linear (L), quadratic (Q), product (P), and hinge (H) features as default FCs and threshold (T) as optional FC (Phillips et al. 2017), while default RM used is 1. Though the models built using either larger or smaller RM compared with the default, are expected to be over-complex or over-simplistic respectively (Phillips et al. 2017), it was actually the models built using default RM and the FCs that were found to be either over-complex or over-simplistic (Shcheglovitova and Anderson 2013, Morales et al. 2017). Therefore, it is prudent to build multiple models using different combinations of RM values and FCs and then choose the optimal model for use in conservation decisions (Muscarella et al. 2014, Morales et al. 2017, Phillips et al. 2017, Galante et al. 2018).

Choosing the optimal model is another key concern for all Maxent users (Warren and Seifert 2011, Shcheglovitova and Anderson 2013, Muscarella et al. 2014, Radosavljevic and Anderson 2014, Galante et al. 2018). Maxent models have been selected using two model selection approaches: (i) information criteria, specifically Akaike information criteria corrected for small sample size (AICc) and (ii) performance in predicting withheld data (Galante et al. 2018). Out of the two approaches AICc was more robust when species occurrence had sampling bias, but both performed well when the bias was corrected (Galante et al. 2018). However, there is a question on the AICc’s fit for model selection in Maxent despite its better performance (Muscarella et al. 2014, Galante et al. 2018). Whereas, the withheld data selection approach is open to the use of multiple selection criteria (Muscarella et al. 2014) which are either used independently (e.g. Warren and Siefert 2011) or in various combinations (e.g. Shcheglovitova and Anderson 2013, Radosavljevic and Anderson 2014, Galante et al. 2018) to select the optimal model.

The commonly used model selection criteria under the ‘withheld data’ approach include the ‘area under the curve of the receiver operating characteristic’ plot for the training data (AUC_{TRAIN}) and AUC_{TEST} , the AUC_{DIFF} and the OR (Warren and Siefert 2011, Muscarella et al. 2014, Radosavljevic and Anderson 2014, Galante et al. 2018). When used independently AUC_{TRAIN} and AUC_{TEST} can provide model discriminatory power. However, use of AUC_{TRAIN} for model evaluation is criticized, with the use of AUC_{TEST} preferred (Radosavljevic and Anderson 2014). While, AUC_{DIFF} and OR can evaluate overfitting (Warren and Siefert 2011, Radosavljevic and Anderson 2014), they may also select over permissive models (Galante et al. 2018). Therefore, recent literature has either used a sequential combination of OR and AUC_{TEST} (OR_{TEST} approach) (Galante et al. 2018) or OR, AUC_{DIFF} and AUC_{TEST} (AUC_{DIFF} approach) (Radosavljevic and Anderson 2014) but their performances have not been compared directly. Since Maxent produces multiple ORs corresponding to multiple thresholding rules there is also a need to assess the performance of different thresholding rules and their corresponding ORs for optimal model selection along with the use of multiple taxonomic groups to help derive general patterns (Galante et al. 2018) if present.

In this study we developed multiple SDMs with different FC and RM combinations for two groups of freshwater organisms with different life history traits, namely 10 fish (which complete their whole lifecycle in water) and 28 odonate species (whose nymphal stage is aquatic but the adults are terrestrial (Bybee et al. 2016)) recorded from Bhutan. We then selected optimal models for each species using two OR_{TEST} and two AUC_{DIFF} approaches. We also selected optimal models through expert screening for ecologically plausible models using binary suitable habitat maps (hereafter ‘EXP approach’). Though it is sensible to tune models and then screen for ecologically plausible models in every study (Muscarella et al. 2014, Morales et al. 2017, Phillips et al. 2017, Galante et al. 2018), this option may be either very time consuming or outright impractical if multiple species are involved within a time bound project. Therefore, we aimed to assess which of the sequential approaches best matched the EXP approach in selecting the optimal models, and hence help reduce the time required for the optimal model selection. We did this by comparing (i) model complexity and (ii) the predicted suitable habitats of the optimal models selected through the five selection

approaches.

Materials and method

Species occurrence data

We collated and cleaned the occurrence data of the fish and odonate species of Bhutan (see Appendix 1 Table S1). We then rarefied the occurrence of all the fish and odonate species with five or more occurrences to Euclidean distance of 5km using the SDMtoolbox v2.4 (available from <http://sdmtoolbox.org/>, Brown 2014, Brown, Bennett and French 2017). We aimed to reduce the negative influence of the spatial autocorrelation among occurrences on the species distribution models (Brown 2014, Brown et al. 2017, Galante et al. 2018), but also to retain the maximum possible number of species with at least five occurrences. This resulted in 10 fish and 28 odonate species with small occurrences (n= 5 to 21) for which we built SDMs and used these for the current study.

Environmental variables

We used 19 bioclimatic variables with 30s resolution ($\sim 1\text{km}^2$ near the equator) that gives current climate data (Fick and Hijmans, 2017 available in <http://worldclim.org/version2>) as the environmental variables. Bioclimatic variables are derived from the monthly temperature and rainfall values to make them biologically more meaningful; for example, mean annual temperature, maximum temperature of the warmest month, annual precipitation and precipitation of the wettest quarter etc. (Fick and Hijmans, 2017). In addition, we also used elevation (AsterDEM_Version3_drukref03.im obtained from Watershed Management Division, Ministry of Agriculture and Forest) as an environmental variable. We used bioclimatic variables as environmental variables even for the fish as previous studies have found SDMs built using bioclimatic and hydrological variables did not differ for fish (McGarvey et al. 2018). We also did not reduce bioclimatic variables since our study is exploratory in nature, and collinearity among environmental variables was not an issue in a machine learning environment like Maxent (Elith et al. 2011, Marco Júnior and Nóbrega 2018) though some literature (e.g. Merow, Smith, and Silander Jr., 2013) suggests being cautious when interpreting SDMs resulting from the use of correlated environmental variables.

Bias files

The ‘presence only’ SDMs are affected by bias sampling, to overcome this one of the common methods employed is the use of a target group bias file to restrict background sampling area (Phillips et al. 2009; Syfert et al. 2013; Vollerling et al. 2019). We pooled all occurrence coordinates of the fish species and adult odonate species with 2 or more occurrence coordinates to produce target group bias files for both fish and odonate species respectively. However, we excluded *Epiophlebia laidlawi* from the odonate target group since its distribution in Bhutan is known only from its larval distribution while all other odonate species are mainly known from their adult occurrence (Appendix 1). Therefore, we developed a separate bias file for the *E. laidlawi* using all the sampling sites from Dorji (2015) along with occurrence coordinates from Brockhaus and Hartmann (2008) and Dupchu, S (Personal communication). We used Hydrosheds (hybas_as_lev12_v1c available in www.hydrosheds.org, Lehner and Grill, 2013) clipped to the Bhutan boundary to bound the occurrence instead of regular sized grid cells of the environmental variables (Phillips et al. 2009; Syfert et al. 2013) to develop sampling bias grids.

Model setting

We used Maxent v 3.4.1 (Phillips, Dudík and Schapire) to model the species distributions. We used Maxent’s current default output format Cloglog since it gives a better result over logistic when bias correction is used (Phillips, 2017; Phillips et al. 2017). We used four sets of FCs resulting from the use of individual FCs independently or in combination with other FCs, namely (i) linear (L), (ii) linear-quadratic (LQ), (iii) hinge (H) and (iv) linear-quadratic-hinge (LQH) after Galante et al. (2018) to build models given our species had only a small number of occurrence records (Table S1). Higher FCs were found to produce less complex (i.e., lesser number of parameters) and less overfitting (i.e., lower omission rates) models when occurrences were

small (Radosavljevic and Anderson 2014). Specifically, the hinge feature was better for species with small occurrence values (Galante et al. 2018).

RMs with values less than default produce models which are overfit to occurrence data and are not well generalized, while larger RMs would produce spread out and less localized models (Phillips, 2017). Though Radosavljevic and Anderson (2014) also observed a slight peak in the model discriminatory ability around the default RM, they found substantial reduction in overfitting when RMs of two to four times that of the default were used. However, they also found both the model quality and the overall discriminatory power declined when RMs were above 4 (Radosavljevic and Anderson 2014). Hence, we used conservatively 11 different RM values, namely .25, .5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5 and 5.

We used Maxent’s default replication method of cross-validation since it is a better replicate option with small occurrence data. It randomly splits occurrence data into folds and uses all the folds in turn to build and evaluate models (Phillips, 2017). We set the number of iterations for each FC-RM combination equal to the number of occurrence (n) for each species thus making it equivalent to n-1 jackknife folds, which is a good approach for species with small occurrence data (Warren and Seifert 2011, Shcheglovitova and Anderson 2013, Radosavljevic and Anderson 2014, Galante et al. 2018). Crossing 11 RM values and four FC sets we built 44 sets of models for each species. Maxent generates (n+1) models including one composite (average) model for each set of RM-FC combination.

Selecting optimal models using sequential approaches

Out of the 11 threshold dependent omission rates Maxent produces in its output we used “10th percentile training presence test omission” (hereafter ‘percentile OR’) and “balance training omission, predicted area and threshold values test omission” (hereafter ‘balance OR’) for the sequential model selection approaches. We chose percentile OR (Radosavljevic and Anderson, 2014; Galante et al. 2018) over the “minimum training presence test omission” (Shcheglovitova and Anderson, 2013; Radosavljevic and Anderson, 2014) since the latter is more sensitive to extreme localities and over predicts when calibration localities are many (Radosavljevic and Anderson, 2014). We used balance OR to assess utility of a new thresholding rule and its OR in selecting optimal model. Through different sequential combinations of the two ORs, AUC_{TEST} and AUC_{DIFF} we formulated four sequential approaches. They were: (i) sequential combination of percentile OR followed by AUC_{TEST} (hereafter $OR_{TEST-PER}$), (ii) sequential combination of balance OR followed by AUC_{TEST} (hereafter $OR_{TEST-BAL}$), (iii) sequential combination of percentile OR followed by AUC_{DIFF} and then by AUC_{TEST} (hereafter $AUC_{DIFF-PER}$), and (iv) sequential combination of balance OR followed by AUC_{DIFF} and then by AUC_{TEST} (hereafter $AUC_{DIFF-BAL}$) approaches.

We used composite models instead of the jackknife iterations (Galante et al., 2018) for each RM-FC combination to select the optimal model. However, Maxent averages all the jackknife iterations to produce the composite model irrespective of whether some individual jackknife models have good model discrimination ($AUC_{TEST} > .5$), marginal discrimination ($AUC_{TEST} < .5$) or no discrimination at all ($AUC_{TEST} = .5$) (Figure S1). When the composite models are comprised of jackknife models with no discrimination they would have lower average ORs since Maxent assigns zero OR to the models with no discriminatory power, and thereby favours these as optimal models. Therefore, we first sorted composite models into four hierarchical groups beginning with (i) the composite models with all jackknife iterations with $AUC_{TEST} > .5$, (ii) followed by ones with some jackknife iteration models with $AUC_{TEST} < .5$, (iii) then with some jackknife iteration models with $AUC_{TEST} = .5$ and (iv) ended with composite models with all their jackknife iteration models having $AUC_{TEST} = .5$.

Following the above hierarchical groups, we then ranked the ORs, AUC_{DIFF} and AUC_{TEST} of the composite models. We accorded the highest rank to the models with the lowest OR since ORs higher than the theoretically expected value indicate overfitting (Radosavljevic and Anderson 2014). Similarly, we accorded the highest rank to the models with the lowest AUC_{DIFF} since less overfitting models are expected to have lower AUC_{DIFF} (Warren and Seifert 2011, Radosavljevic and Anderson 2014). Here, we also considered negative AUC_{DIFF} as equal to zero, the lowest AUC_{DIFF} for model selection (Muscarella et al. 2014), though

we used raw values for general analysis. For the AUC_{TEST} we accorded the highest rank to the models with the highest AUC_{TEST} since higher AUC_{TEST} means better model performance or discriminatory ability (Radosavljevic and Anderson, 2014).

Once thus ranked, we followed the steps outlined in Figure 1. We chose the composite model or subset of composite models with the highest OR rank (corresponding to Step 1 of Figure 1). Since we used OR as the first criteria to select the optimal models if only a single composite model had the highest OR rank (i.e., the lowest ORs among the models) we considered it the optimal model for both OR_{TEST} and AUC_{DIFF} approaches (Figure 1). If Step 1 resulted in a subset of composite models we chose either a composite model or subset of models with the highest AUC_{TEST} rank for the two OR_{TEST} approaches (corresponding to Step 2b, Figure 1) (Shcheglovitova and Anderson 2013, Galante et al. 2018). Whereas, for AUC_{DIFF} approaches we chose the model or models with the best ranked AUC_{DIFF} (corresponding to Step 2a, Figure 1) followed by Step 2b (Radosavljevic and Anderson, 2014) depending on the outcome of Step 2a (Figure 1). After Step 2b, depending on the outcome, we followed Steps 3 to 5 for both the OR_{TEST} and AUC_{DIFF} approaches (Figure 1). In Step 3 we chose the models with the lowest average number of parameters since models with lower numbers of parameters are considered less complex and better models (Galante et al. 2018). We derived the average number of parameters for each candidate composite optimal model by dividing the sum of the number of parameters with non-zero lambda coefficients for each individual model extracted from the LAMBDA text file (Galante et al. 2018) by the number of iterations used for building SDM since Maxent does not provide directly the average number of parameters in the result for the composite models unlike it does for the threshold values and ORs. Further, when multiple optimal models had equal average number of parameters, we then chose models with the lower average lambda coefficients obtained by dividing the sum of the absolute value of lambda coefficients by the total number of parameters. However, for some species multiple optimal models with same RM values but different FCs had equal average numbers of parameters as well as the average absolute lambda coefficients. In such cases we used the composite models with simpler FC as the final optimal model since lower FCs are considered better for species with smaller occurrence in Maxent (Shcheglovitova and Anderson 2013).

Selecting optimal models using expert approach

We visually screened and compared which of the suitable/unsuitable binary maps produced from the optimal models selected using the four sequential approaches were ecologically plausible. If none of the optimal models selected through the four sequential approaches resulted in ecologically plausible binary maps we then moved to the next best choice and continued to screen till we arrived at the best possible ecologically plausible map using expert knowledge (Galante et al. 2018). We qualitatively assessed ecological plausibility by considering the known elevation range and the predicted suitable habitat. We also compared the predicted map with any available literature, such as IUCN distribution maps or authors' field experience. However, we acknowledge here that given the relatively high number of fish and odonate species modelled, compounded by very poor literature on the species from Bhutan, some of the chosen optimal models could be subjective. We developed the binary maps using the "balance training omission, predicted area and threshold value Cloglog threshold" (hereafter 'balance threshold') out of 11 Cloglog thresholds generated by Maxent for each model iteration. Though lower thresholds can overpredict suitable habitat, they are better for species with few occurrence data (Pearson et al. 2007, Radosavljevic and Anderson 2014) and can also uncover potentially informative distribution areas (Pearson et al. 2007). We also found using the 10th percentile threshold restricted the predicted suitable habitats around occurrence points, or predicted the whole study area as unsuitable, in some cases (See 'Predicted Habitat' in Result) (Pearson et al. 2007, Radosavljevic and Anderson, 2014; Galante et al. 2018). We produced binary maps using ArcGIS10.2.2 (ESRI 2014).

Data analysis

We calculated the percentages of optimal models sharing the same RM-FC combination among all the five model selection approaches, and between different pairs, to check for any general trend in RM-FC combination among optimal models. We also statistically tested the correlation among the optimal models chosen by the selection approaches for the model features, viz., ORs, AUC_{DIFF} , AUC_{TEST} , and average number of

parameters, and also the area of suitable habitat predicted. We interpreted the strength of correlation based on Mukaka (2012). We used nonparametric Spearman’s correlation since our variables were not normally distributed, except for few (Table S2) when we performed a preliminary analysis. We performed all the statistical analysis in IBM SPSS Statistics 23 (IBM Corp.).

Results

Model complexity

Model setting

The most common model setting or regularization multiplier and feature class (RM-FC) combinations for fish was 5xH (Figure 2a, Table S3), while for odonates it was 5xL (Figure 2b, Table S3). Further, greater number of optimal models had RM-FC combinations of larger RM values and more complex feature classes (Figure 2a-b, Table S3). For three fish and four odonates all the five selection approaches selected same optimal models (Table 1). Whereas, OR_{TEST_PER} agreed best with EXP for fish and AUC_{DIFF_PER} and OR_{TEST_BAL} for the odonates (Table 1; Table S3).

Omission rates

We found all selection approaches chose optimal models with percentile ORs equal to or more than theoretically expected 10% for both fish (Figure 3a) and odonates (Figure 3b). On the other hand, except for four fish and 10 odonates, all had optimal models with 0 balance ORs (Figures 3a-b). Generally, there was a very high and statistically significant correlation for the percentile ORs of the optimal models chosen by the EXP approach to the optimal models chosen by all four sequential approaches for fish, while for the odonate there was statistically significant but very low correlations only with OR_{TEST_PER} , AUC_{DIFF_PER} and OR_{DIFF_BAL} (Table S4). However, for the balance OR EXP approach had statistically significant correlations only with OR_{TEST_PER} and AUC_{DIFF_PER} for fish and with AUC_{DIFF_PER} for odonates (Table S4).

AUC Difference

AUC differences (AUC_{DIFF}), as expected, were comparatively lower for the optimal models chosen by AUC_{DIFF} approaches over OR_{TEST} approaches for both fish and odonates (Figures 3c-d). While, the EXP approach chose a greater number of optimal models with comparatively larger AUC_{DIFF} over any of the four sequential approaches for both fish and odonates (Figure 3c-d). However, there were statistically significant correlations for AUC_{DIFF} of the optimal models selected by the EXP approach to all the sequential selection approaches (Table S4). The highest positive correlation was found with OR_{TEST_PER} for fish and with OR_{TEST_BAL} for odonates though correlation strength was only moderate for the latter (Table S4).

Number of parameters

The greater number of optimal models selected by the EXP approach had a greater number of parameters among all the selection approaches for both fish and odonates (Figures 3e-f). Among the four sequential approaches, the OR_{TEST} approaches selected a greater number of optimal models with a greater number of parameters over the AUC_{DIFF} approaches (Figures 3e-f). But all approaches selected optimal models with the number of parameters greater than the number of occurrences used for model building for at least some species (Figures 3e-f). Optimal models chosen by the EXP approach had a statistically significant correlation for the number of parameters only with optimal models chosen by OR_{TEST_PER} for fish (Table S4).

AUC test

A greater number of optimal models selected by the EXP approach for fish had a greater AUC_{TEST} over other approaches, while for the odonates the OR_{TEST} approaches had greater number of optimal models with greater AUC_{TEST} followed by the EXP approach (Figures 3g-h). Overall the AUC_{TEST} values of optimal models chosen by all five approaches were high (Figures 3g-h) except for one fish species *Schizothorax progas-tus* which had an AUC_{TEST} value of .382 for the optimal models selected by OR_{TEST_PER} and AUC_{DIFF_PER} approaches (Figures 3g-h). There was also a high to very high statistically significant positive correlation for

the AUC_{TEST} among the optimal models chosen by all the approaches (Table S4) suggesting all five selection approaches may select equally the optimal models with good discriminatory power.

Predicted habitat

The EXP approach selected the maximum number of optimal models with comparatively smaller predicted habitat area among all the approaches; this was expected as the expert approach tried to avoid over prediction while selecting the optimal models (Figure 4a-b). In comparison, AUC_{DIFF} approaches predicted a comparatively larger suitable habitat area for a greater number of both fish and odonate species over the OR_{TEST} approaches (Figure 4a-b). Most often suitable habitats were over predicted by AUC_{DIFF} approaches and also for some species by OR_{TEST} approaches when compared to the area predicted by the EXP approach (Figure 4a-b). For instance, AUC_{DIFF_PER} and AUC_{DIFF_BAL} predicted habitat area above 38,000 km² for four of the odonate species and for one fish species by AUC_{DIFF_BAL} (Figure 4a-b).

The choice of threshold used to derive binary suitable habitat maps from the optimal models chosen may explain some of the variation in the area of habitat predicted. Though we did not assess for the effect of thresholds used on the habitat areas predicted we present the binary maps derived using percentile and balance thresholds for four species with varying occurrence records used for the model building as an example (Figure 5). In general, we observed the restrictive ‘percentile threshold’ seemed to restrict the predicted suitable habitat around occurrence data used for model building, while the less restrictive ‘balance threshold’ seemed to overpredict the habitat for most of the optimal models selected by sequential approaches (Figure 5).

However, there were statistically significant correlations between the areas of the optimal models chosen by the EXP approach, with optimal models chosen by all four sequential approaches (Table S4). Comparatively, the strongest correlation was with OR_{TEST_PER} and OR_{TEST_BAL} for fish and with OR_{TEST_PER} and AUC_{DIFF_PER} for odonates, though correlation strength differed with very high correlations for the former and moderate for the latter (Table S4).

Discussion

Our study compared for the first time the model complexity and predicted suitable habitat of the optimal models selected by OR_{TEST} and AUC_{DIFF} approaches using two omission rates derived from both restrictive and less restrictive thresholds. We also used the less restrictive threshold to derive predicted suitable habitats and to assess for their ecological plausibility. Further, we used two taxonomic groups of freshwater species with different life history traits. The focus of this study, Bhutan, is also interesting as freshwater species distributions are poorly documented (National Biodiversity Centre, 2014).

Model complexity

Overall, our results showed 5xH was the most common RM-FC combinations for fish which agreed well with the earlier findings of optimal models with higher RM values combined to more complex FCs like H as better model setting for species with small occurrence data (Shcheglovitova and Anderson 2013, Galante et al. 2018). While, for the odonates the most common RM-FC combination was 5xL which also agrees with the need to use RM values greater than the default setting (Galante et al., 2018). However, our result also showed other optimal models had different RM-FC combinations for both fish and odonates (Figure 2a-b). Further, our results also showed poor agreement among the five model selection approaches with regard to RM-FC combinations of the optimal models selected by them for both the fish and odonate. The EXP approach had the highest agreement with OR_{TEST_PER} for fish, while for odonates the EXP approach agreed the most with AUC_{DIFF_PER} and OR_{TEST_BAL} . When these findings are considered together, they may suggest the need for taxon-specific model tuning, thus agreeing with the recognized need for model tuning for the specific species in a given study (Galante et al., 2018).

Our results showed that all approaches chose models with percentile ORs equal to or greater than the theoretically expected 10% suggesting generally overfit models (Galante et al., 2018). But other studies have also found percentile ORs greater than 10% (see Muscarella et al. 2014, Radosavljevic and Anderson

2014, Galante et al. 2018) when a small number of occurrences were used. However, the OR_{TEST} and EXP approaches not only chose a larger number of optimal models with larger percentile ORs over that by AUC_{DIFF} approaches, but they also chose models with larger AUC_{DIFF} and a larger number of parameters for both fish and odonates. These findings suggest the OR_{TEST} and EXP approaches might have selected overfit and over-parameterized optimal models (Muscarella et al. 2014, Radosavljevic and Anderson 2014, Galante et al., 2018). However, an earlier study found over-parameterization a lesser issue than under-parameterization (Warren and Seifert 2011).

Our use of a relaxed balance threshold to first generate binary suitable/unsuitable habitat area and then choosing the EXP optimal model might have overcome model overfitting and under predicting for EXP approach (Pearson et al., 2007, Radosavljevic and Anderson 2014). Whereas, models chosen solely based on smaller percentile ORs, AUC_{DIFF} and number of parameters may choose overly relaxed (over predicting) models (Galante et al., 2019) which can be aggravated by our use of a relaxed balance threshold to generate the binary habitat map. Therefore, though AUC_{DIFF} approaches chose the optimal models with small AUC_{DIFF} and smaller number of parameters over OR_{TEST} and EXP approaches, in our context these optimal models are not necessarily the best. Further, optimal models chosen by AUC_{DIFF} approaches had comparatively lower AUC_{TEST} values over the EXP approach followed by OR_{TEST} approaches. Therefore, AUC_{DIFF} approaches might have chosen a greater number of over predicting optimal models with lower model discriminatory power (Warren and Seifert 2011).

Predicted habitat

The EXP approach selected optimal models with generally smaller predicted habitat among all the approaches given we avoided over prediction while selecting ecologically plausible models. Among the four sequential approaches AUC_{DIFF} approaches generally predicted a larger area as suitable habitat for both the taxonomic groups compared to the two OR_{TEST} approaches when the balance threshold was used. For some species the habitat predicted was several times bigger than that predicted by the EXP approach, thus over predicting. When the less restricted thresholds, considered robust for the species with small occurrence records (Pearson et al. 2007, Warren and Seifert 2011), are used to develop binary suitable habitat maps, as we did here, use of OR_{TEST} approaches may be a better choice over AUC_{DIFF} approaches for optimal model selection. While the use of AUC_{DIFF} approaches may not be useful when a restrictive threshold, like the percentile threshold (Galante et al. 2018), is used, since optimal models selected through AUC_{DIFF} approaches under predicted for some species (Figure 5). Further studies may be required to confirm if there is a true relationship between the thresholds used to derive binary suitable maps and the corresponding model selection approaches used to select the optimal models.

Though the expert approach selected more complex models judged using RM values, ORs, AUC_{DIFF} and average number of parameters, the use of a less restrictive threshold value for producing binary suitable habitat maps helped select ecologically plausible models (Pearson et al. 2007). We suggest the use of OR_{TEST} approaches over AUC_{DIFF} approaches as either the first line of optimal model screening or by their own to select the final optimal models since optimal models chosen by these approaches had higher correlation with the optimal models chosen by the EXP approach for almost all the parameters we tested. Further, though AUC_{DIFF} approaches selected less overfit models over OR_{TEST} and EXP approaches, AUC_{DIFF} approaches over predicted the habitat area when the balance threshold (a less restrictive threshold) was used in our case. However, we feel there is a need for further studies using other thresholding rules available in the Maxent result as well as including a broader range of taxonomic groups to assess the generality of our findings.

Author contributions

First author conceptualised, gathered data, did analysis and wrote first draft and final version, co-authors provided key conceptual ideas, revised the drafts and approved the final version.

Data availability: Raster data used to derive Figure 4 will be submitted to DRYAD if the manuscript is accepted.

Reference

- Brockhaus, T. and Hartmann, A. (2009). New records of *Epiophlebia laidlawi* Tillyard, 1921 in Bhutan with notes on its biology, ecology, distribution, zoogeography and threat status (Anisozygoptera: Epiophlebiidae). *Odonatologica* , 38: 203-215.
- Brown J. L. (2014). SDMtoolbox: a python-based GIS toolkit for landscape genetic, biogeographic, and species distribution model analyses. *Methods in Ecology and Evolution* , DOI: 10.1111/2041-210X.12200.
- Brown J. L., Bennett J. R. and French, C. M. (2017). SDMtoolbox 2.0: the next generation Python-based GIS toolkit for landscape genetic, biogeographic and species distribution model analyses. *PeerJ* 5:e4095; DOI 10.7717/peerj.4095.
- Bybee, S., Córdoba-Aguilar, A., Duryea, M. C., Futahashi, R., Hansson, B., Lorenzo-Carballa, M. O... Wellenreuther, M. (2016). Odonata (dragonflies and damselflies) as a bridge between ecology and evolutionary genomics. *Frontiers in Zoology*, 13: 46. DOI 10.1186/s12983-016-0176-7.
- Coxen, C. L., Frey, J. K., Carleton, S. A. and Collins, D. P. (2017). Species distribution models for a migratory bird based on citizen science and satellite tracking data. *Global Ecology and Conservation* , 11: 298-311. <https://doi.org/10.1016/j.gecco.2017.08.001>
- Darwall, W., Bremerich, V., De Wever, A., Dell A. I., Freyhof, J., Gessner, M. O., ... Weyl, O. (2018). The *Alliance for Freshwater Life* : A global call to unite efforts for freshwater biodiversity science and conservation. *Aquatic Conservation: Marine and Freshwater Ecosystems* , 28: 1015-1022. <https://doi.org/10.1002/aqc.2958>
- Dorji, T. (2015). New distribution records of *Epiophlebia laidlawi* Tillyard, 1921 (Insecta: Odonata) in Bhutan. *Journal of Threatened Taxa* , 7:7668-7675.
- Elith, J., Graham, C.H., Anderson, R.P., Dudík, M., Ferrier, S., Guisan, A., ... Zimmermann, N.E. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography* , 29: 129-151.
- Elith, J., Phillips, S.J., Hastie, T., Dudík, M., Chee, Y.E. and Yates, C.J. (2011). A statistical explanation of Maxent for ecologists. *Diversity and Distribution* , 17: 43-57.
- Fick, S. E. and Hijmans, R. J. (2017). worldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* , 37: 4302-4315. DOI: 10.1002/joc.5086
- Galante, P. J., Alade, B., Muscarella, R., Jansa, S. A., Goodman, S. M. and Anderson, R. P. (2018). The challenge of modelling niches and distributions for data-poor species: a comprehensive approach to model complexity. *Ecography* , 41: 726-736. DOI:10.1111/ecog.02909
- Gomez, V. H. F., IJff, S. D., Raes, N., Amaral, I. L., Salomão, R. P., Coelho, L. d. S... Steege, H. t. (2018). Species distribution modelling: contrasting presence-only models with plot abundance data. *Scientific Reports* , 8: 1003. DOI:10.1038/s41598-017-18927-1
- Guisan, A., Tingley, R., Baumgartner, J.B., Naujokaitis-Lewis, I., Sutcliffe, P.R., Tulloch, A.I.T., ... Buckley, Y.M. (2013). Predicting species distributions for conservation decisions. *Ecology Letters* , 16: 1424-1435.
- Gyeltshen, C., Tobgay, K., Gyeltshen, N., Dorji, T., and Dema, S. (2018). New species discoveries and records in Bhutan Himalaya. In M. Hartmann, M. V. L. Barclay and J. Weipert. (Eds.), *Biodiversität und Naturausstattung im Himalaya / Biodiversity and Natural Heritage of the Himalaya*, vol. VI (pp. 59-82). Erfurt: Naturkundemuseum Erfurt.
- Kass, J. M., Vilela, B., Aiello-Lammens, M. E., Muscarella, R., Merow, C. and Anderson, R. P. (2017). WAL-LACE: a flexible platform for reproducible modelling of species niches and distributions built for community expansion. *Methods in Ecology and Evolution* , 9: 1151-1156. DOI:10.1111/2041-210x.12945.

- Kramer-Schadt, S, Niedballa, J., Pilgrim, J.D., Schroder, B., Lindenborn, J., Reinfelder, V... Wilting, A. (2013). The importance of correcting for sampling bias in MaxEnt species distribution models. *Diversity and Distribution* , 19:1366-1379.
- Lehner, B. and Grill, G. (2013). *Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems*. *Hydrological Processes*, 27: 2171–2186. Data is available at www.hydrosheds.org.
- Liu, C., Newell, G. and White, M. (2016). On the selection of thresholds for predicting species occurrence with presence-only data. *Ecology and Evolution* , 6: 337–348.
- Marco Junior, P. D. and Nóbrega, C. C. (2018). Evaluating collinearity effects on species distribution models: an approach based on virtual species simulation. *PLoS ONE* , 13, e0202403. <https://doi.org/10.1371/journal.pone.0202403>
- McGarvey, D.J., Menon, M., Woods, T., Tassone, S., Reese, J., Vergamini, M. and Kellogg, E. (2018). On the use of climate covariates in aquatic species distribution models: are we at risk of throwing out the baby with the bath water? *Ecography* , 41: 695-712.
- Merow, C., Smith, M.J. and Silander, Jr J.A. (2013). A practical guide to Maxent for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography* , 36:1058-1069.
- Morales, N. S., Fernández, I. C., & Baca-González, V. (2017). MaxEnt's parameter configuration and small samples: are we paying attention to recommendations? A systematic review. *PeerJ* , 5 , e3093. doi:10.7717/peerj.3093
- Mukaka, M. M. (2012). Statistics corner: a guide to appropriate use of correlation coefficient in medical research. *Malawi Medical Journal* , 24: 67-71.
- Muscarella, R., Galante, P. J., Soley-Guardia, M., Boria, R. A., Kass, J. M., Uriarte, M. and Anderson, R. P. (2014). ENMeval: 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. Doi:10.1111/2041-210X.12261.
- National Biodiversity Centre. (2014). *National biodiversity strategies and action plan of Bhutan 2014* . NBC, Ministry of Agriculture and Forests, Royal Government of Bhutan, Thimphu.
- Norris, D. (2014). Model thresholds are more important than presence location type: understanding the distribution of lowland tapir (*Tapirus terrestris*) in a continuous Atlantic forest of southeast Brazil. *Tropical Conservation Science* , 7: 529-547.
- Pereira, D. G., Afonso, A. and Medeiros, F. M. (2015). Overview of Friedman's test and post-hoc analysis. *Communications in Statistics - Simulation and Computation* , 44:10, 2636-2653, DOI: 10.1080/03610918.2014.931971
- Pearson, R. G., Raxworthy, C. J., Nakamura, M., & Peterson, A. T. (2007). Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *Journal of Biogeography*, 34 , 102-117.
- Phillips, S. J. (2017). A brief tutorial on Maxent. Available from <http://biodiversityinformatics.amnh.org/open-source/maxent/>. Accessed on 10. 02.2018.
- Phillips, S. J., Anderson, R. P., Dudík, M., Schapire, R. E. and Blair, M. E. (2017). Opening the black box: an open-source release of Maxent. *Ecography*, 40: 887-893.
- Phillips, S.J., Anderson, R.P. and Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling* , 190: 231–259.

- Phillips, S.J. and Dudík, M. (2008). Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* , 31:161-175.
- Phillips, S. J., Dudík, M., Elith, J., Graham, C. H., Lehmann, A... Ferrier, S. (2009). Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications* , 19: 181-197.
- Phillips, S.J., Dudík, M. and Schapire, R.E. [Internet] Maxent software for modeling species niches and distributions (Version 3.4.1). Available from url:http://biodiversityinformatics.amnh.org/open_source/maxent/. Accessed on 12.2.2018.
- Radosavljevic, A. and Anderson, R.P. (2014). Making better Maxent models of species distributions: complexity, overfitting and evaluation. *Journal of Biogeography* , 41:629-643.
- Shcheglovitova, M. and Anderson, R. P. (2013). Estimating optimal complexity for ecological niche models: a jackknife approach for species with small sample sizes. *Ecological Modelling*, 269: 9-17. <http://dx.doi.org/10.1016/j.ecolmodel.2013.08.011>
- Syfert, M. M., Smith, M. J. and Coomes, D. A. (2013). The effects of sampling bias and model complexity on the predictive performance of MaxEnt species distribution models. *PLoS ONE* , 8: e55158. Doi:10.1371/journal.pone.0055158.
- Vollering, J., Halvorsen, R., Auestad, I. and Rydgren, K. (2019). Bunching up the background betters bias in species distribution models. *Ecography* , 42: 1717-1727. Doi:10.1111/ecog.04503
- Warren, D. L. and Seifert, S. N. (2011). Ecological niche modelling in Maxent: the importance of model complexity and the performance of model selection criteria. *Ecological Applications* , 21: 335-342.
- Young, N., Carter, L. and Evangelista, P. (2011). A maxent model v3.3.3e tutorial (ArcGIS v10). Last modified on September 1, 2011. Natural Resources Ecology Laboratory at Colorado State University and the National Institute of Invasive Species Science. Available in http://ibis.colostate.edu/webcontent/ws/coloradoview/tutorialsdownloads/a_maxent_model_v7.pdf. Accessed on 12.1.2018.

Figure legends

Figure 1. Schematic diagram of selecting optimal models. Steps with red arrows are unique to AUC_{DIFF} approaches while all other steps are common to all the four sequential approaches. Green boxes with “Single model selected” are the optimal models for either OR_{TEST} or AUC_{DIFF} approaches. Step 2a is followed then one will derive optimal models for AUC_{DIFF} after Step 2, otherwise optimal models derived will be for OR_{TEST} approaches. While purple box with “Single model selected” is the optimal model for only AUC_{DIFF} approaches. Step 5 is the ultimate step wherein models with lower feature class is chosen as the optimal model when multiple models have same numbers of parameters and average absolute value.

Figure 2. Regularization multiplier and feature class (RM-FC) combinations for (a) fish and (b) odonate of the optimal models chosen by five model selection approaches.

Figure 3. Summary of different features of the optimal models selected through five optimal model selection approaches for the fish and odonate species of Bhutan. 10 percentile training presence test omission (1) and balance training omission, predicted area and threshold value test omission (2) for the (a) fish and (b) odonate; AUC difference for the (c) fish and (d) odonate; average number of parameters for the (e) fish and (f) odonate; test AUC for the (g) fish and (h) odonate species of Bhutan. Expert approach based on ecological plausibility of binary suitable/unsuitable model (EXP), sequential approaches using 10 percentile training presence test omission and test AUC (OR_{TEST_PER}), 10 percentile training presence test omission, AUC difference and test AUC (AUC_{DIFF_PER}), balance training omission, predicted area and threshold value test omission and test AUC (OR_{TEST_BAL}) and balance training omission, predicted area and threshold value test omission, AUC difference and test AUC (AUC_{DIFF_BAL}).

Figure 4. Area of the predicted habitats of the optimal models selected through the five model selection approaches for the (a) fish and (b) odonate species of Bhutan.

Figure 5. Examples of binary suitable habitat maps derived using 10 percentile training presence Cloglog threshold (Left panel) and balance training omission, predicted area and threshold value Cloglog threshold (Right panel) among the optimal models chosen by the Expert and the four sequential optimal model selection approaches. (a) *Cyprinion semiplotum* (5 occurrence; EXP, OR_{TEST_PER}, OR_{TEST_BAL}, AUC_{DIFF_PER} and AUC_{DIFF_BAL}), (b) *Neolissochilus hexagonolepis* (12 occurrence; EXP, OR_{TEST_PER}); (c) *Neolissochilus hexagonolepis* (12 occurrence; OR_{TEST_BAL}, AUC_{DIFF_PER} and AUC_{DIFF_BAL}); (d) *Aristocypha quadrimaculata* (8 occurrence; EXP, OR_{TEST_PER}, OR_{TEST_BAL}, AUC_{DIFF_PER} and AUC_{DIFF_BAL}); (e) *Diplacodes trivialis* (21 occurrence; EXP); (f) *Diplacodes trivialis* (21 occurrence; OR_{TEST_PER}); (g) *Diplacodes trivialis* (21 occurrence; OR_{TEST_BAL}, AUC_{DIFF_PER} and AUC_{DIFF_BAL}).

Tables

Table 1. Number of optimal models with same Regularization Multiplier-Feature Class (RM-FC) combination among all the five model selection approaches (ALL) and between pairs of the five model selection approaches for the fish (n=10) and odonate (n=28) species of Bhutan. Expert approach based on ecological plausibility of binary suitable/unsuitable model (Expert); sequential approaches using 10 percentile training presence test omission and test AUC (OR_{TEST_PER}), 10 percentile training presence test omission, AUC difference and test AUC (AUC_{DIFF_PER}), balance training omission, predicted area and threshold value test omission and test AUC (OR_{TEST_BAL}) and balance training omission, predicted area and threshold value test omission, AUC difference and test AUC (AUC_{DIFF_BAL}).

Selection approaches	Fish	Odonata
All	3	4
EXP-OR _{TEST_PER}	7	6
EXP-AUC _{DIFF_PER}	3	10
EXP-OR _{TEST_BAL}	5	10
EXP-AUC _{DIFF_BAL}	3	8
OR _{TEST_PER} -AUC _{DIFF_PER}	6	11
OR _{TEST_PER} -OR _{TEST_BAL}	6	14
OR _{TEST_PER} -AUC _{DIFF_BAL}	5	10
AUC _{DIFF_PER} -OR _{TEST_BAL}	7	12
AUC _{DIFF_PER} -AUC _{DIFF_BAL}	8	25
OR _{TEST_BAL} -AUC _{DIFF_BAL}	8	13























