

**NAME OYEDAYO OYELOWO    STUDENT NO: 014717208 DATE: 12<sup>TH</sup> DEC, 2017**  
**METHODS AND UNCERTAINTIES IN BIOCLIMATIC ENVELOPE MODELLING**  
**UNDER CLIMATE CHANGE.** (*Risto K. Heikkinen, Miska Luoto, Miguel B. Araújo,*  
*Raimo Virkkala, Wilfried Thuiller and Martin T. Sykes*)

It is infeasible to capture the distribution of all species as they exist in nature. This is the rationale behind species distribution model(SDM)- to predict the location, localisation, situation and distribution of organisms by utilising the environmental factors which affect them. The article dwells on a type of SDM - single-species bioclimatic envelope model - which relates suitable/unsuitable environmental variables with climatic variables, in order to predict and explain the distribution of species. Because, this is not the exact but simulated reality, it is salient to take into cognisance, the uncertainties that might arise, as such. Also, the pertinence of incorporating other important affecting factors such as landcover, CO<sub>2</sub> etc. was discussed.

With the projected increase in climate, several other kinds of models have been used to predict species distribution. However, the article adopted the species distribution models which are quite popular. Climatic variables are used based on present equilibrium to build a model and make predictions for the future. However, this has been critiqued since climate is not the singular factor affecting distribution, hence, not sufficient, especially for future prediction. Soil, fire, dispersal limitation, increasing atmospheric CO<sub>2</sub>, and other human-induced activities, pose limitations to bioclimatic envelope models). Effects of warming should be interpreted with caution.

Single species produce more realistic results while the combination of several might yield unrealistic results as they explore different niche, have different adaptation and reactions to climate. The article explored various methodologies that have been employed in bioclimatic models in recent years and the uncertainties that come with them. It also looks into the how selection of predictors can affect the result of an analysis. For instance, absence of mobile species is difficult to model. Nonetheless, presence only proves useful in remote areas with dearth of inventories. Absence data, when accessible can also be useful but it difficult to model absence

Different model approaches include Classification tree analysis(CTA), Generalised Boosted Model(GBM), Generalised linear model(GLM), Generalised additive model(GAM), artificial neural network(ANN) etc These all have their strength and weaknesses. CTA is good with handling complex interactions and non-additive behaviour. It can also handle quantitative

and qualitative variable. However, while it handles complex interactions properly, it can also create some complexities and misleading interpretations of results. GLM, GAM, GBM are quite similar but are respectively increasingly more detailed in delivering their predictions. They are all capable of handling not only gaussian distribution but also binomial and Poisson without having to transform the response variable via the identity function which uses logit and log for binomial and Poisson distributions respectively.

However, GAM has the advantage over GLM of being able to present some response curves from the models more readily. It offers more flexibility. In similar vein, GBM offers even more flexibility in dealing with the explanatory variables and their interactions. By recursive binary splitting also, it is able to produce model with improved predictive performance and less error. This is done with the number of internal iterations in it. It also presents the relative importance of the predictors.

ANN on the other hand, can handle predictors from different sources, such as Boolean and qualitative/categorical and can also determine the predictors with non-linear response. This is adopted the bioclimatic envelope models. However, it requires large quantity of data amongst others. Multivariate adaptive regression splines (MARS) is another kind of model which has not be used as much but proves to be useful. This is related to GBM.

The performance of these various models in species distribution are evaluated using several techniques. Notably is the Area Under Curve, which considers several thresholds for the performance of a model. Above 0.9 is considered to be excellent. This is good because it precludes the subjectiveness in selecting the threshold as it considers all possible thresholds to maximise the performance. This is a problem in kappa statistics or classification matrix. However, some methods such as prevalence have been suggested for selecting this threshold.

In validating the model, there are several approaches. One of them is resubstitution whereby same dataset is used for calibration and evaluation. Bootstrapping, data splitting and leave-one-out cross validation have also been adopted. Resubstitution is the least preferred because of its overly optimistic/pessimistic performance in predicting. The latter are more preferred, however, not sufficient enough. It has been recommended to use independent evaluation data to avoid the spatial autocorrelation bias. This evaluation data could be from other region or time.

Models building comes in several folds, such as, a priori selection of a set of explanatory variables, manual model building, and automated model calibration. A priori selection might be good when empirical information about the variables is available. However, the variables chosen which could show strong correlation with the response variable might

show a weak response in another region. Also, some of the variables might have to be excluded in the case of multicollinearity. Hence, this approach has little flexibility.

The automated model calibration, on the other hand, shows greater flexibility and are readily done in many software packages. However, it can also produce some unrealistic models. It also has little transparency which is where the manual model building comes into play. This helps one to connect logically to the workflow of the model building and able to identify the key things that might be unnoticed in the automated way.

Furthermore, model selection is a critical aspect of model building. Balance has to be created to avoid overfitting or underfitting. Studies have shown that including too many predictors might yield less performing models and also introduce noise into the models. A conservative approach to selection of predictors include, forward, backward (or both) stepwise regression which uses the significance of variables at every step to eliminate the least significant. Bayesian information criterion (BIC) and Akaike's information criterion (AIC) have similarly been employed but sometimes could produce underfitting. Care needs to be taken to consider the theoretical framework when selecting these predictors to produce realistic results.

Similarly, the sample size should be taken into account and should be representative. Multicollinearity should be treated with care. It might be useful to eliminate spatial autocorrelations. This can be done manually, or by using some other techniques such as the principal component analysis(PCA). It is also useful to eliminate outliers and overdispersion.

From the foregoing, it can be deduced that no method of modelling is perfect, as they all have their pros and cons. As such, one needs to consider and apply them, case-by-case. Importantly, one needs to consider the aim of the research, the data at hand, the geographical scale, amongst others. For instance, bioclimatic envelope model, can give insight into climatic influences on a species and the effect of potential climate change. In this case, it might be useful to adopt it. Despite this, it is critical to understand the underlying principles and shortcomings of the models. As the correlative performance of the model is evaluated, so, should the error be put into consideration.

Nonetheless, several researches have demonstrated the importance of considering other factors that are integral into the complex ecosystem. Such include the landcover, CO<sub>2</sub> etc. The other factors also play important roles. However, one needs to clearly define the scope of the conceptual framework, as the entire parts cannot be included in a model.