

## Species Distribution Models

### Introduction

Species distribution models (SDMs) (Franklin, 2009) have shown great potential in helping to achieve conservation planning goals by refining our knowledge of species distributions (Jetz et al., 2012). SDMs extrapolate species distribution data in space and time, usually based on a statistical model. These models identify areas that are ecologically suitable for the presence of species (Soberon & Peterson, 2005; Hirzel et al., 2002; Franklin, 2009). Use of SDMs can help to support management decisions with regard to biodiversity (Pawar et al., 2007; Baldwin, 2009; Franklin, 2009). Many examples can be cited that have made extensive use of SDMs for different applications, for example, assessing global impacts, prioritising or targeting areas for protected status, assessing threats to those areas, predicting species distributions in unsurveyed areas and designing reserves (Araújo & Williams, 2000; Pearce & Ferrier, 2000; Thuiller, 2003; Araújo et al., 2004; Elith et al., 2006; Romero-Cacerrada & Luque, 2006; Elith et al., 2010). SDMs can be applied to vegetation or animal distribution modelling; several examples exist in Europe of their application to species, species groups, guilds, alliances or communities (for vegetation).

There are a wide variety of SDM methods, each with their own characteristics. The BIOMOD methods are discussed here specifically because they are used in the OpenNESS case study 5 (forest planning in the Vercors Mountain Range, France), (Thuiller, 2003; Thuiller et al., 2009; <http://cran.r-project.org/web/packages/biomod2/index.html>). BIOMOD is a platform for ensemble forecasting of species distributions, enabling the explicit treatment of model uncertainties and the examination of species-environment relationships (Thuiller et al., 2009). It includes the ability to model species distributions with several techniques (see a summary in Figure 1), test models with a wide range of approaches, project species distributions into the future using different climate scenarios and dispersal functions, assess species temporal turnover, plot species response curves, and test the strength of species interactions with predictor variables. Computationally, BIOMOD is a collection of functions running within the R (CRAN) software (programmed in the R language) and allows the user to apply a range of statistical models to several dependent variables using a set of independent variables. Thus, BIOMOD attempts to span the different approaches that can be used in habitat suitability modelling. It does not aim to be exhaustive, but it presents the most commonly used modelling approaches and the ones considered to be the most interesting and robust and which are implemented in R (see <http://cran.r-project.org/bin/windows/base/>).

Model		Concept	Technique	Environmental variables types	Key references
SRE	Rectilinear Envelope	Environmental envelope	Equivalent to Bioclim. Climatic Envelope Model is a GARP-simulation	Cont	Busby 1991, Nix 1986 ; Walker & Cocks, 1991; McMahon et al. 1996
CTA	Classification Tree Analysis	Decision tree	Classification and regression	Cont/Cat	Breiman et al., 1984
RF	Random Forest		Classification and regression	Cont/Cat	Breiman, 2001
GBM	Generalized Boosting Model	Regression and decision tree	Combination of regression decision Trees & "boosting"(method combining several simple models to improve predictions performance)	Cont/Cat	Jerome H. Friedman, 1999
FDA	Flexible Discriminant Analysis	Regression analysis	Classification method by using Friedman's (1991) multivariate adaptative regression spline, using the MARS function for the regression part of the model.	Cont/Cat	Hastie, T., Tibshirani, R and Buja, A.,1994 Manel, D., Dias, J. M., Buckton, S. T. and Ormerod, S. J.,1999
MARS	Multivariate Adaptive Regression Splines		Linear model	Cont/Cat (only Cont in Biomod2)	Friedman, 1991
GAM	Generalised Additive Models		Additives model	Cont/Cat	Guisan et al., 2002, Pearce & Ferrier, 2000
GLM	Generalised Linear Models		Linear models / additives models/least square fitting	Cont/Cat	Guisan et al., 2002, Pearce & Ferrier, 2000
ANN	Artificial Neural Networks	Machine learning	Neural nets	Cont/Cat	Pearson et al., 2002
Maxent	Maximum entropy		Maximum entropy	Cont/Cat	Phillips et al., 2006

Figure 1. Summary of the models that can be used within BIOMOD (Le Roux, 2013). 'Cont/Cat' = Continuous/Categoric.

## Keywords

Species distributions; Climate change; Suitable climate space; Biodiversity; Spatial approach.

## Why would I chose this approach?

If it was important to understand where the climate might be appropriate for species in the future. Information on climatic suitability for the species that make up a habitat may provide a valuable first step towards understanding the ecosystem services that habitat may provide under different future climatic conditions.

## What are the main advantages of the approach?

- Can identify areas where climate and/or habitat is appropriate for a given species;
- Can be used to explore multiple future scenarios;
- Spatial outputs produced with accompanying goodness-of-fit statistics;
- Freely available.

## What are the constraints/limitations of the approach?

- Some species are very hard to model as the factors driving their present-day distributions are unclear;
- As with any modelling, some species fit better with the driving variables and produce projections that are more statistically significant than others;
- Relatively advanced statistical process underly the models; mathematical and technical expertise are required to interpret the results;

- The projections reflect the climate, environmental characteristics and/or habitat niche that a species could potentially use – it does not usually take into consideration other factors such as predation, competition or disease, or changes over time in factors such as habitat distribution.

## What types of value can the approach help me understand?

SDMs can provide information to support biophysical valuation. They also provide information about impacts on future habitats and species and so information that can support assessment of other values that depend on these species: for example suitability for key crop/timber species or those with particular regulatory (e.g. key carbon-sequestering or soil regulating species) or cultural significance (e.g. spiritually significant or emblematic species). With a temporal perspective they can be used to illustrate the potential impacts on intrinsic, existence and bequest values.

## How does the approach address uncertainty?

The approaches will usually provide some metric of match to baseline conditional (a goodness-of-fit for the model). Sensitivity testing can be applied to explore the impacts of uncertainty around the input variables on the robustness of the results.

## How do I apply the approach?

The majority of SDMs (e.g. most of the modelling approaches within BIOMOD) require data on the presence and absence of species, but it is possible to work only with presence data<sup>1</sup>. The presence/absence species data may be related to a wide variety of environmental variables, including habitat parameters, temperature, soil moisture, NDVI<sup>2</sup>, slope, aspect, distance to wetlands or rivers, and evapotranspiration index. The environmental variables that should be included depend on the knowledge of the species or groups of species to be modelled. The information which is entered into the model should relate in some way to the distribution of the species being modelled (e.g. they should limit or control the distribution of the species in some manner). Some of the models used, such as Maxent, were specifically designed for presence-only data, and to overcome problems of small samples.

Instructions to implement BIOMOD2 are freely available in Thuiller (2012), Thuiller et al. (2012) and Georges & Thuiller (2013). The steps in the BIOMOD2 modelling process are:

1. Gather all available and meaningful GIS information (as outlined above). All GIS layers have to share the same projection system (e.g. WGS 84). GIS layer resolution depends on the original data, but may be degraded to speed calculation if fine-resolution layers are not crucial for the species studied. The spatial extent needs to be specified (the calculation time will depend on its surface area). All layers need to be supplied as rasters (using conversion tools if necessary in GIS software). Raw GIS layers may need to be adapted, e.g. by first producing a map of distance to a river from an original river map and then converting this to a distance-to-river map as a raster.

<sup>1</sup> Presence/absence data maps in detail where a species is present and also where it is absent; presence data only maps roughly where it is known to be present.

<sup>2</sup> Normalised difference vegetation index – an indicator of photosynthetic activity and hence vegetation productivity.

2. Data on observed species distributions can be provided either as .csv data with three columns providing geographical coordinates and presence/absence data, or as a raster from GIS software. Note that more than three columns can be provided if you are modelling more than one species.
3. Install BIOMOD2, R-Cran, the latest version of Java and Maxent (following the instructions provided in Georges & Thuiller, 2012); the following links will be needed:
  - <http://cran.r-project.org/web/packages/biomod2/index.html>
  - <http://www.cs.princeton.edu/~schapire/maxent/>
  - <http://cran.r-project.org/bin/windows/base/>
  - <http://www.oracle.com/technetwork/java/javase/downloads/index.html>

Different R packages need to be installed: biomod2, abind, sp, raster, rastervis, lattice, latticeExtra, RColorBrewer, hexbin, grid, nnet, gbm, survival, splines, gbm, mda, class, randomForest, rpart, pROC, plyr, rgdal, zoo (for further information see [http://cran.r-project.org/web/packages/available\\_packages\\_by\\_name.html](http://cran.r-project.org/web/packages/available_packages_by_name.html)). The tutorials listed in the links above explain how to carry out the analyses.

## Requirements

<i>Data</i>	<input checked="" type="checkbox"/> Data is available <input checked="" type="checkbox"/> Need to collect some new data <input type="checkbox"/> Need to collect lots of new data	Depending on the availability of data within the case study and the species in question. The resolution of the case study will also determine the extent to which suitable data are available both in terms of species and contextual datasets. Collecting primary species data is a considerable task: in most cases SDMs depend on secondary data collation rather than collection of primary data.
<i>Type of data</i>	<input type="checkbox"/> Qualitative <input checked="" type="checkbox"/> Quantitative	
<i>Expertise and production of knowledge</i>	<input checked="" type="checkbox"/> Work with researchers within your own field <input type="checkbox"/> Work with researchers from other fields <input type="checkbox"/> Work with non-academic stakeholders	In general, SDMs require expertise from the ecology/biodiversity field, but input from non-academic stakeholders can be useful to validate the results.
<i>Software</i>	<input checked="" type="checkbox"/> Freely available <input type="checkbox"/> Software licence required <input checked="" type="checkbox"/> Advanced software knowledge required	Depends on the species distribution model in question. Some are freely available for download, others are embedded in particular institutions. BIOMOD is implemented in R statistical coding language and is a freeware, open source, package. SPECIES is implemented as standalone interface.
<i>Time resources</i>	<input checked="" type="checkbox"/> Short-term (< 1 year) <input type="checkbox"/> Medium-term (1-2 years) <input type="checkbox"/> Long-term (more than 2 years)	Depending on the level of available data can be performed in less than a year. Will depend on the level of skill of the programmer and the level of pre-processing required to create the driving variables.
<i>Economic resources</i>	<input type="checkbox"/> < 6 person-months	

	<input checked="" type="checkbox"/> 6-12 person-months <input type="checkbox"/> > 12 person-months	
Other requirements		

## Where do I go for more information?

- Baldwin, R.A. (2009). Use of Maximum Entropy Modeling in Wildlife Research. *Entropy*, 11: 854-866.
- Elith, J., Graham, C.H., Anderson, R.P., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., McC. Overton, J., Peterson, A.T., Phillips, S.J., Richardson, K., Scachetti-Pereira, R., Schapire, R.E., Soberón, J., Williams, S., Wisz, M.S. & Zimmermann, N.E. (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.
- Franklin, J. (2009). *Mapping Species Distribution: Spatial Inference and Prediction (Ecology, Biodiversity and Conservation Series)*, Saint Diego State: Cambridge University Press. 320pp.
- Georges, D. & Thuiller W. (2013). *An example of species distribution modeling with BIOMOD2*.
- Harrison PA, Berry PM, Butt N & New M (2006). Modelling climate change impacts on species' distributions at the European scale: Implications for conservation policy. *Environmental Science and Policy*, 9: 116-128.
- Hirzel, A.H., Haussez, J., Chessel, D. & Perrin, N. (2002). Ecological-Niche Factor Analysis: How to compute habitat suitability maps without absence data? *Ecology*, 7: 2027-2036.
- Jetz, W., McPherson, J. & Guralnick, R. (2012). Integrating biodiversity distribution knowledge: toward a global map of life. *Trends in ecology & evolution (Personal edition)*, 27(3): 151-159.
- Pawar, S., Koo, M., Kelley, S.C., Ahmed, M., Choudhury, F.S. & Sarkar, S. (2007). Conservation assessment and prioritization of areas in Northeast India: priorities for amphibians and reptiles. *Biological Conservation*, 136: 346-361.
- Pearce, J. & Ferrier, S. (2000). An evaluation of alternative algorithms for fitting species distribution models using logistic regression. *Ecological Modelling*, 128: 127-147.
- Pearson RG, Dawson TP, Berry PM, Harrison PA (2002). SPECIES: A spatial valuation of climate impact on the envelope of species. *Ecological Modelling*, 154: 289-300.
- Romero-Calcerrada, R. & Luque, S. (2006). Habitat quality assessment using Weights-of-Evidence based GIS modelling: The case of *Picoides tridactylus* as keystone species indicator of the biodiversity value of the Finnish forest. *Ecological Modelling*, 196: 62-76.
- Soberon, J. & Peterson, A. (2005). Interpretation of models of fundamental ecological niches and species distribution areas. *Biodiversity Informatics*, 2: 1-10.
- Thuiller, W. (2003). BIOMOD – optimizing predictions of species distributions and projecting potential future shifts under global change. *Global Change Biology*, 9: 1353-1362.
- Thuiller, W., Lafourcade, B., Engler, R. & Araújo, M.B. (2009). *BIOMOD – a platform for ensemble forecasting of species distributions*, pp. 369-373.

*Factsheet prepared by Sandra Luque, Marie Le Roux, Frederci Archaux, Paula Harrison & Robert Dunford*

