

SDMvspecies: a software for creating virtual species for species distribution modelling

Ren-Yan Duan, Xiao-Quan Kong, Min-Yi Huang, Gan-Lin Wu and Zhi-Gao Wang

R.-Y. Duan, X.-Q. Kong, M.-Y. Huang (huang.m.y@163.com), G.-L. Wu and Z.-G. Wang, School of Life Science, Anqing Normal College, Anqing, Anhui 246011, PR China.

The following is a report on software developed to create virtual species for the study of species distribution modelling (SDM). SDMvspecies provides several methods to create virtual species. The package is designed to be simple and intuitive, even for users who are not familiar with the R language. SDMvspecies is available online free of charge from <http://cran.r-project.org/web/packages/sdmvspecies/>.

The use of virtual species in species distribution modelling (SDM) has increased in recent years (Barve et al. 2011, Peterson 2011, Barbet-Massin et al. 2012, Meynard and Kaplan 2013). They are an efficient way to allow complete control of the format and magnitude of input data (Li and Guo 2013, Varela et al. 2014). Virtual species can solve many problems that real species distribution data cannot do. For example, virtual species can insure the prediction results are not influenced by different species and peculiarities (Barbet-Massin et al. 2012). The method for creating virtual species is first reported by Hirzel et al. (2001). Since then, many methods of creating virtual species have been developed and used in recent studies on SDM (Jiménez-Valverde and Lobo 2007, Lobo and Tognelli 2011). However, most of these studies simply provide a description of how their method works, not a software that researchers can easily use. With the increasing use of virtual species in SDM, it is important to provide a simple, intuitive and standard software way to create them. Here we provide a report on software developed for creating virtual species with ease and simplicity.

Implementation and required input

SDMvspecies are implemented as a package in R. In theory, SDMvspecies can be used in any platform that supports the R language. In our study, SDMvspecies was implemented and tested on Linux platforms. SDMvspecies requires the installation of raster package in R which will be used to handle raster maps used in SDM analyses (Hijmans 2014). SDMvspecies allows all the raster formats available in the raster R-package.

Tests implemented in SDMvspecies

Current version of SDMvspecies (ver. 0.2.1) contains four methods to create virtual species. All those methods are

listed in Table 1. Pick median method is very similar with 'pick mean' method in algorithms and usage, so we only choose one of them (pick mean method) to make example. All environment variables are come from www.worldclim.org (Hijmans et al. 2005).

Niche syntheses method

The following method for virtual species creation was originally reported in Hirzel et al. (2001). This method is known as 'niche syntheses method' because it uses a function to simulate suitability of virtual species response to environmental variables. We combine the responses to synthesize suitability of virtual species, and then use a given threshold to make a virtual species distribution map.

Virtual species are generated by simulating ecological niches in multidimensional space. For i th-dimensional space, the ecological niche suitability is calculated as $H_i \times W_i$, where H_i is the virtual species' niche suitability index of this space and the W_i is the weight of this suitability ($H_i \in [0,1]$). The final (global) suitability is calculated by:

$$H = \frac{1}{\sum_{i=1}^n W_i} \sum_{i=1}^n W_i H_i + \varepsilon$$

Where H is the habitat suitability of a given cell. ε is random error term ($\varepsilon \in (-0.05, 0.05)$). Global habitat suitability is calculated as a weighted average of partial niche suitability adds a random error term. The partial niche suitability is mapping from environment by a function. This type of function is called 'niche response function', and its function is mapped the environment variable to niche suitability.

Table 1. Four methods to create virtual species in the SDMvspecies.

Method names	Algorithms	Reference
Niche syntheses method	Each environment variable is translated to partial suitability, then weighted average to total suitability.	Hirzel et al. (2001)
Pick mean method	Apply PCA on environment to obtain reduced non-correlated environment variables. The environment range falling within mean \pm SD of each factor is considered to be suitable for this layer. If all layers are suitable, this site is selected as the true distribution range of the virtual species.	Jiménez-Valverde and Lobo (2007)
Pick median method	Similar with 'pick mean method'. The environment variables are come from the principal components with the highest factor loadings. The environment range falling within the two central quartiles of each factor is considered to be suitable for this layer. If all layers are suitable, this site is selected as the true distribution range of the virtual species.	Lobo and Tognelli (2011)
Artificial bell-shaped response method	Artificial given a mean and SD of a bell shape curve for each of environment variables, then multiply them together.	Varela et al. (2014)

Tests of niche syntheses method

Here we demonstrate how to create virtual species by using the niche syntheses method. The environmental variable, response function and weight are listed in Table 2. After inputting these variables, SDMvspecies will output a raster map. We rescale the suitability by setting the range of suitability from 0 to 1. Then, a given threshold is chosen to create a species distribution map. In this example, 0.7 is used as the threshold. The resulting suitability and distribution map are shown in Fig. 1 and Fig. 2.

Pick mean method

This method was originally reported in Jiménez-Valverde and Lobo (2007). The workflow is as follows: 1) choose environment variables of interest; 2) apply principal component analysis (PCA) to obtain reduced non-correlated environment variables (the user can choose to apply PCA or other filtering methods); 3) virtual species will be considered present in cells that fall within the mean \pm SD of each variable and absent if the fall outside the mean \pm SD.

Tests of pick mean method

To test the pick mean method, we chose annual mean temperature (Bio1), max temperature of the warmest month (Bio5),

Table 2. Variable name, variable means, niche response function and weight of niche syntheses method.

Variable name	Variable means	Niche response function	Weight
Bio1	Annual mean temperature ($^{\circ}\text{C} \times 10$)	Bell-shaped function	2
Bio5	Max temperature of warmest month ($^{\circ}\text{C} \times 10$)	Linear decrease function	1
Bio7	Temperature annual range ($^{\circ}\text{C} \times 10$)	Truncated linear (increasing) function	2
Bio12	Annual precipitation (mm)	Linear increase function	2

annual temperature range (Bio7), and annual precipitation (Bio12) as environment variables. After PCA analysis, we extract two principal components. We inputted these principal component variables into SDMvspecies, and the resulting distribution map (no suitability map) is shown in Fig. 3.

Artificial bell-shaped response method

This method was originally reported in Varela et al. (2014). This method defines virtual species that are not resistant to extreme environment conditions, and virtual species limited by extreme environments. The workflow is as follows: 1)

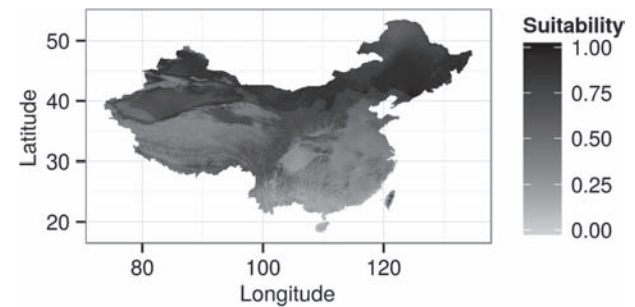


Figure 1. Suitability map of niche syntheses method.

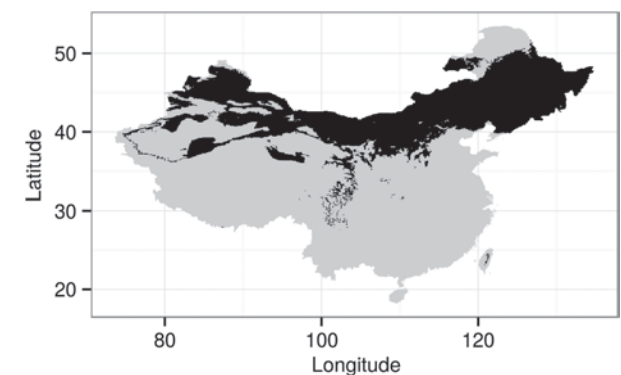


Figure 2. Distribution map of niche syntheses method.

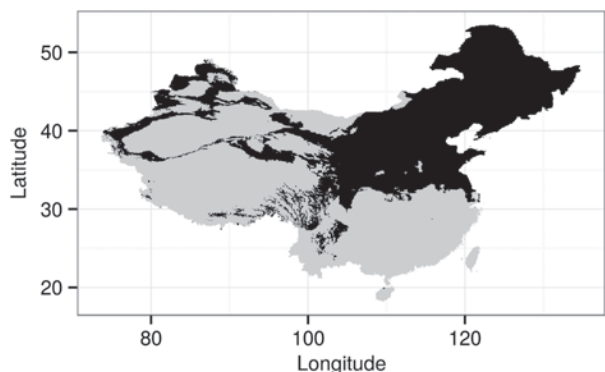


Figure 3. Distribution map of pick mean method.

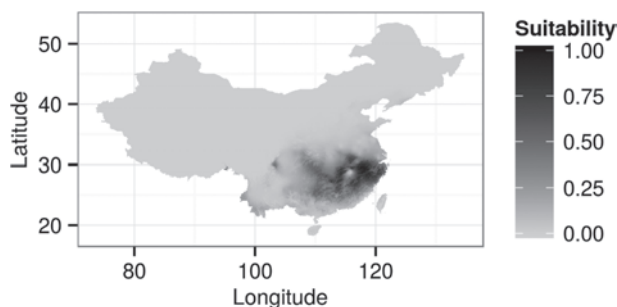


Figure 4. Suitability map of artificial bell-shaped response method.

environment variables that limit the virtual species distribution are chosen; 2) for each of environment variables, users choose a mean and standard deviation to build bell curves (normal curves); 3) the overall species suitability is defined as the multiplication of each variable's suitability; 4) The users then chooses a threshold to make a species distribution map.

Tests of artificial bell-shaped response method

We chose annual mean temperature (Bio1; mean = 15°C and standard deviation = 5°C), max temperature of the warmest month (Bio5; mean = 30°C and standard deviation = 10°C), annual temperature range (Bio7; mean = 40°C and standard deviation = 10°C), and annual precipitation (Bio12; mean = 2000 mm and standard deviation = 500 mm) as environment variables. We inputted the environment

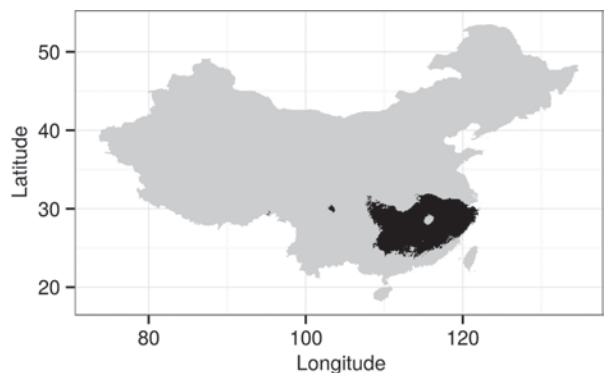


Figure 5. Distribution map of artificial bell-shaped response method.

variables and settings into *SDMvspecies*, and the resulting suitability map is shown in Fig. 4. We rescale the suitability to range 0 and 1, then chose 0.3 as threshold to gain species distribution map (Fig. 5).

Package installation

The *SDMvspecies* package is free and open source, that meaning users can use the software free of charge and obtain the source code. *SDMvspecies* is under the GPL v2 license. The *SDMvspecies* software, manual, and sample data are available at the *SDMvspecies* website <<http://cran.r-project.org/web/packages/sdmvspecies/>>.

To cite *SDMvspecies* or acknowledge its use, cite this Software note as follows, substituting the version of the application that you used for 'version 0':

Duan, R.-Y., Kong, X.-Q., Huang, M.-Y., Wu, G.-L. and Wang, Z.-G. 2014. *SDMvspecies*: a software for creating virtual species for species distribution modelling. – *Ecography* 37: 000–000 (ver. 0).

Acknowledgements – We thank Sara Varela of Charles Univ., Prague, Czech Republic for her important suggestions. This work was supported by the Special Public Welfare Forest Foundation of State Forestry Administration (201304314), National Natural Science Foundation of China (31100311, 31300342), and Anhui Provincial National Science Foundation (1308085QC65). R-YD and X-QK contributed equally to this work.

References

- Barbet-Massin, M. et al. 2012. Selecting pseudo-absences for species distribution models: how, where and how many? – *Methods Ecol. Evol.* 3: 327–338.
- Barve, N. et al. 2011. The crucial role of the accessible area in ecological niche modeling and species distribution modeling. – *Ecol. Model.* 222: 1810–1819.
- Hijmans, R. J. 2014. Raster: geographic data analysis and modeling. – <<http://cran.r-project.org/web/packages/raster/>>.
- Hijmans, R. J. et al. 2005. Very high resolution interpolated climate surfaces for global land areas. – *Int. J. Climatol.* 25: 1965–1978.
- Hirzel, A. et al. 2001. Assessing habitat-suitability models with a virtual species. – *Ecol. Model.* 145: 111–121.
- Jiménez-Valverde, A. and Lobo, J. M. 2007. Threshold criteria for conversion of probability of species presence to either-or presence-absence. – *Acta Oecol.* 31: 361–369.
- Li, W. and Guo, Q. 2013. How to assess the prediction accuracy of species presence-absence models without absence data? – *Ecography* 36: 788–799.
- Lobo, J. M. and Tognelli, M. F. 2011. Exploring the effects of quantity and location of pseudo-absences and sampling biases on the performance of distribution models with limited point occurrence data. – *J. Nat. Conserv.* 19: 1–7.
- Meynard, C. N. and Kaplan, D. M. 2013. Using virtual species to study species distributions and model performance. – *J. Biogeogr.* 40: 1–8.
- Peterson, A. T. 2011. Ecological niche conservatism: a time-structured review of evidence. – *J. Biogeogr.* 38: 817–827.
- Varela, S. et al. 2014. Environmental filters reduce the effects of sampling bias and improve predictions of ecological niche models. – *Ecography* 37: 1084–1091.