submit to biorxiv as preprint

submit to Ecological Modelling for peer-review

**Title:** Shifting from close-source graphical-interface to open-source programming: a brief tutorial on runing Maxent in R

Authors: Xiao Feng and coauthors

Abstract:

Keywords: Maxent, R, open-source, Rmarkdown, Github, clamping

**Introduction**

With the rapid development of GIS techniques and accumulation and digitization of museum specimens, ecological niche modeling (ENM) is increasingly used in quantifying the relationship between species’ presences and environmental conditions (Peterson *et al.*, 2011). In the field of ENM, many modeling algorithms have been developed [e.g., GARP (Stockwell, 1999), ENFA (Hirzel *et al.*, 2002), and Maxent (Phillips *et al.*, 2004)], and recently there is a trend of development toward an open-source algorithm/software (e.g., Maxent; Phillips *et al.*, 2017), particularly in a programming environment (i.e., R; Hijmans *et al.*, 2013; Naimi & Araújo, 2016; Thuiller *et al.*, 2016).

The trend of open-source in a programing environment is particularly important and useful in the era of big data, when high volume, high velocity, and/or high variety of scientific data are accumulated. The input data for ENM are mainly species’ occurrence data and environmental data. Species’ occurrences data are rapidly increasing, benefited from digitization of museum specimens and citizen science projects (\*\*cite review paper), and for example, the number of records on Global Biodiversity Information Facility (gbif.org) has reached 0.79 billion. The other major input data for ENM, environmental data (i.e., GIS data and remote sensing data), also increase in unprecedented speed, benefitted from advanced techniques in sensors, satellites, … (\*\*need citations). The traditional software (e.g., using graphical user interface by clicking) will be inadequate in handling such large amount of data, whereas in a programing environment a simple loop function can efficiently process a task thousands of times based on an established procedure, with further advantage of repeatability and accuracy.

It is not only the data are increasingly accumulated, but also new methods and tools are more rapidly being developed (Muscarella *et al.*, 2014; Qiao *et al.*, 2015; Kass *et al.*, 2017; Feng *et al.*, in press). The traditional way of scientific software development will be insufficient in keeping pace with newly developed scientific methods, because of the limited efforts and insights in a closed developing environment compare to a collective and open source developing environment (e.g. Github.com).

In the field of ENM, Maxent (Elith *et al.*, 2006; Phillips & Dudík, 2008) is probably the most used ENM algorithm and its publications have been cited collectively thousands of times, probably because of the user-friendly graphical-interface (Joppa *et al.*, 2013). Maxent was originally a close-source software and recently became open-source (Phillips *et al.*, 2017), though the graphical user interface persists as the major way of using it. Here, we identified several gaps of knowledge of using Maxent in an open-source programming environment (i.e., R).

There has been an effort in bridging the use of Maxent in R by Hijmans *et al.* (2013), but detailed demonstration of running Maxent in R is lacking. Maxent implements 67 parameters (\*\*cite Maxent help document), which are easy to manipulate in the graphical-interface but are not straight forward to implement in R. There are scattered pieces of information on the Internet (e.g. Maxent-Google Groups, R mailing lists, or programmer communities), but there has been no detailed documentation that explains or demonstrates implementation of those Maxent parameters in R.

Additionally, ENM scholars found that three settings in Maxent have major influence on model performance, which are features, regularization parameters, and clamping (Phillips & Dudík, 2008; Moreno-Amat *et al.*, 2015; Qiao *et al.*, under review), thus manipulations of those parameters will be important in model comparison and evaluation. Muscarella *et al.* (2014) insightfully provided a function that automatically implements models using different combinations of features and regularization parameters, though different features are always coupled with the same regularization value, which could potentially be set as different (Phillips & Dudík, 2008).

Inspired by such gaps of knowledge, we identified three objectives which were expected to be important in promoting the user community of Maxent shifting from graphical-interface to open-source programming: 1) we would demonstrate the general use of Maxent in R based on a common used ENM procedure, 2) we would provide a function that bridges the Maxent algorithm and R computing environment for easer use, and 3) we would demonstrate the manipulation of features, regularization parameters, and clamping in R and explain potential tips and caveats of such implementations in R.

**Methods and Results**

Objective 1:

We demonstrated the general use of Maxent in R based on a common used ENM procedure (Appendix 1-workshop document; code at Github).

Objective 2: we would provide a function that bridges the Maxent algorithm and R computing environment for easer use. We only implemented 26 parameters that more commonly used by Maxent user. Detailed explanation of implemented parameters is in Appendix 2-details of parameters; the source code of this function is at Github.

Objective 3, we demonstrated the manipulation of features, regularization parameters, and clamping in R and explain potential tips and caveats of such implementations in R (Appendix 3-new markdown document; code at Github).

**Discussion**

Maxent is undoubtfully a robust modeling algorithm that will attract more research interests in invasive species distribution, climate change and conservation, and outbreaks of epidemics (\*\*\*cite a few). We hope our demonstration and function will help Maxent users learn skills necessary to use Maxent in R, as well as strength the efficiency of future research in the field of ENM. We hope that bridging Maxent and R will lower the threshold of methodology explorations for Maxent community. In the era of big data, single client-based close-source graphical-interface scientific software will be replaced by working environments featured by open-source and programming, or potentially be replaced by server-based/cloud-based web application (Qiao *et al.*, 2012; Feng *et al.*, in press).

**References**

Elith, J., Graham, C.H., Anderson, R.P., et al. (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129-151.

Feng, X., Castro, M., Linde, E. & Papeş, M. (in press) Armadillo Mapper: a case study of an online application to update potential distribution estimates. *Tropical Conservation Science*,

Hijmans, R.J., Phillips, S., Leathwick, J. & Elith, J. (2013) *dismo: species distribution modeling*. 1 November 2014.

Hirzel, A.H., Hausser, J., Chessel, D. & Perrin, N. (2002) Ecological-niche factor analysis: how to compute habitat-suitability maps without absence data? *Ecology*, **83**, 2027-2036.

Joppa, L.N., McInerny, G., Harper, R., Salido, L., Takeda, K., O'Hara, K., Gavaghan, D. & Emmott, S. (2013) Troubling trends in scientific software use. *Science*, **340**, 814-815.

Kass, J.M., Vilela, B., Aeillo-Lammens, M.E., Muscarella, R. & Anderson, R.P. (2017) *Wallace: a modular platform for reproducible ecological modeling*.

Moreno-Amat, E., Mateo, R.G., Nieto-Lugilde, D., Morueta-Holme, N., Svenning, J.-C. & García-Amorena, I. (2015) Impact of model complexity on cross-temporal transferability in MaxEnt species distribution models: an assessment using paleobotanical data. *Ecological Modelling*, **312**, 308-317.

Muscarella, R., Galante, P.J., Soley-Guardia, M., Boria, R.A., Kass, J.M., Uriarte, M. & 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.

Naimi, B. & Araújo, M.B. (2016) sdm: a reproducible and extensible R platform for species distribution modelling. *Ecography*, **39**, 368-375.

Peterson, A.T., Soberón, J., Pearson, R.G., Anderson, R.P., Martínez-Meyer, E., Nakamura, M. & Araújo, M.B. (2011) *Ecological Niches and Geographic Distributions*. Princeton University Press, Princeton, NJ.

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

Phillips, S.J., Dudík, M. & Schapire, R.E. (2004) A maximum entropy approach to species distribution modeling. In: *Proceedings of the Twenty-First International Conference on Machine Learning*, pp. 655–662. ACM press, Banff, Alberta, Canada.

Phillips, S.J., Anderson, R.P., Dudík, M., Schapire, R.E. & Blair, M.E. (2017) Opening the black box: an open-source release of Maxent. *Ecography*, **40**, 887-893.

Qiao, H., Lin, C., Ji, L. & Jiang, Z. (2012) mMWeb - an online platform for employing multiple ecological niche modeling algorithms. *PLoS ONE*, **7**, e43327.

Qiao, H., Lin, C., Jiang, Z. & Ji, L. (2015) Marble algorithm: a solution to estimating ecological niches from presence-only records. *Scientific Reports*, **5**, 14232-14232.

Qiao, H., Feng, X., Escobar, L.E., Peterson, A.T., Soberon, J., Zhu, G. & Papeş, M. (under review) A comprehensive evaluation of transferability of ecological niche modeling algorithms. In:

Stockwell, D. (1999) The GARP modelling system: problems and solutions to automated spatial prediction. *International Journal of Geographical Information Science*, **13**, 143-158.

Thuiller, W., Georges, D., Engler, R. & Breiner, F. (2016) *biomod2: Ensemble Platform for Species Distribution Modeling*.