

Title

Explainable Artificial Intelligence enhances the ecological interpretability of black-box species distribution models

Authors

Masahiro Ryo^{1,2}, Boyan Angelov³, Stefano Mammola^{4,5}, Jamie M. Kass⁶, Blas M. Benito⁷, Florian Hartig⁸

Affiliation

¹ Institute of Biology, Freie Universität Berlin, 14195 Berlin, Germany.

² Berlin-Brandenburg Institute of Advanced Biodiversity Research (BBIB), 14195 Berlin, Germany.

³ Association for Computing Machinery (ACM), 1601 Broadway New York, USA.

⁴ Molecular Ecology Group (MEG), Water Research Institute (IRSA), National Research Council (CNR), 28922 Verbania Pallanza, Italy.

⁵ Laboratory for Integrative Biodiversity Research (LIBRe), Finnish Museum of Natural History (LUOMUS), University of Helsinki, 00100 Helsinki, Finland.

⁶ Biodiversity and Biocomplexity Unit, Okinawa Institute of Science and Technology Graduate University, Kunigami-gun, 904-0495 Okinawa, Japan.

⁷ Department of Ecology & Multidisciplinary Institute for Environment Studies "Ramon Margalef", University of Alicante, San Vicente del Raspeig, Alicante, Spain.

⁸ Faculty of Biology and Pre-Clinical Medicine, Universität Regensburg, 93053 Regensburg, Germany.

Author contribution statement

MR and BA conceived the idea and designed the methodology. BA wrote the R script and analysed the data. SM, JMK, BMB, and FH provided substantial suggestions on species distribution modeling and other theoretical aspects. MR wrote the first draft of the paper and all authors contributed to the writing with suggestions and comments.

Abstract

Species distribution models (SDMs) are widely used in ecology, biogeography, and conservation biology to understand the correlation of species occurrences with the environment and make predictions of how their distributions vary in space and time. During the past two decades, the field has increasingly made use of machine learning approaches for constructing and validating SDMs. Model accuracy has steadily increased as a result, but the interpretability of the fitted models, for example the relative importance of predictor variables for model predictions or their causal effects on focal species, has not always kept pace. Here we draw attention to an emerging subdiscipline of artificial intelligence, explainable AI (xAI), as a toolbox for better interpreting SDMs. xAI aims at deciphering the behavior of complex statistical or machine learning models (e.g., neural networks, random forests, and boosted regression trees), and can produce more transparent and understandable SDM predictions. We describe the rationale behind xAI and provide a list of tools and a reproducible example in R that demonstrates how xAI can help to improve the interpretability of machine learning used for SDMs.

Keywords

xAI; Explainable artificial intelligence; Interpretable machine learning; Species distribution model; Ecological niche modeling; model-agnostic

Main text

Understanding where and why species occur in space and time is central to ecology, biogeography, and conservation biology (Pecl et al. 2017, Araújo et al. 2019). Species distribution models (SDMs) are currently the most widely used approach for this purpose. SDMs correlate species' occurrence records with environmental covariates such as land use and climatic factors, to identify factors that predict species' presence or habitat suitability and to project distributional shifts in response to environmental change (Elith and Leathwick 2009, Booth et al. 2014).

Since the first SDM applications in the early 1980s (Box 1981, Booth et al. 2014, Booth 2018), the field has steadily moved from simple statistical models (e.g., logistic regressions) to more complex statistical methods, often adopting principles or algorithms from the field of machine learning (ML) (Phillips et al. 2006, Elith and Leathwick 2009). Moreover, the community has put substantial efforts into making SDMs more easy-to-use by streamlining the model-building and analytical processes through various software packages, for example, graphical user interfaces (Scachetti-Pereira 2002, Phillips et al. 2006, de Souza Muñoz et al. 2011, Kass et al. 2018) and programming frameworks (*cf.* >10 R packages available for SDM as reviewed in Angelov, 2019). With these developments, SDMs have matured into a widely applied ecological modeling tool that has resulted in more than 6,000 studies using or mentioning SDM in the past two decades (Araújo et al. 2019).

Whereas the wide availability of complex ML algorithms has encouraged users to build more accurate SDMs, it has not necessarily enhanced the understanding of the fitted models (e.g., deep learning; Christin et al., 2019). The downside of complex ML models is that it is hard to understand how and why they make their predictions, which is why they are often referred to as “black-box” models. In general, there is a trade-off between the accuracy and interpretability of statistical models (Breiman 2001a). Achieving both simultaneously is challenging (Guisan and Thuiller 2005), but most researchers would agree that an ideal SDM is both accurate and easy to interpret (Phillips et al. 2004, Austin 2007, Merow et al. 2014). It is therefore a reasonable question to ask whether ecologists should sacrifice interpretability by using excessively complex algorithms for constructing SDMs in order to procure slight advantages in predictive accuracy (Qiao et al. 2015, Araújo et al. 2019).

The dilemma of gaining accuracy only at the expense of interpretability is not unique to ecology. Fields as diverse as financial risk assessment, medicine, or criminal justice have recently also realized that although ML algorithms have desirable properties for making accurate predictions, it is difficult to understand the rationale underlying these predictions. The lack of interpretability makes these models less reliable or acceptable for scientists and stakeholders alike (Ribeiro et al. 2016, Meske and Bunde 2020). This problem has led to the emerging research area of explainable artificial intelligence (xAI), a subfield of AI also termed interpretable ML (Murdoch et al. 2019), that aims at developing tools for enhancing the interpretability of complex algorithms (Carvalho et al. 2019).

The purposes of this forum article are to provide a brief introduction to the field and several techniques of xAI and to suggest for the first time its potential applicability to SDM research (Fig. 1). This work builds upon previous studies and software that improved accessibility and understanding for novel ML tools in ecology (Cutler et al. 2007, Elith et al. 2008, 2011, Olden et al. 2008, Elith and Graham 2009, Merow et al. 2013, Ryo and Rillig 2017). We acknowledge that some of these methods are already routinely used, and substantial efforts have already been made to improve the interpretation of fitted ML models in SDM research and ecology, independently of the emergence of xAI: e.g., bootstrap approach for key variable detection (Olden and Jackson 2002), novel higher-order interaction discovery (Ryo et al. 2018), and the Maxent “Explain” tool (Phillips 2017). However, these efforts are now being rapidly synthesized and expanded in the scientific domain of xAI, and several tools are readily available. Thus, we call for attention to the tools and principles developed in this field for ecological applications.

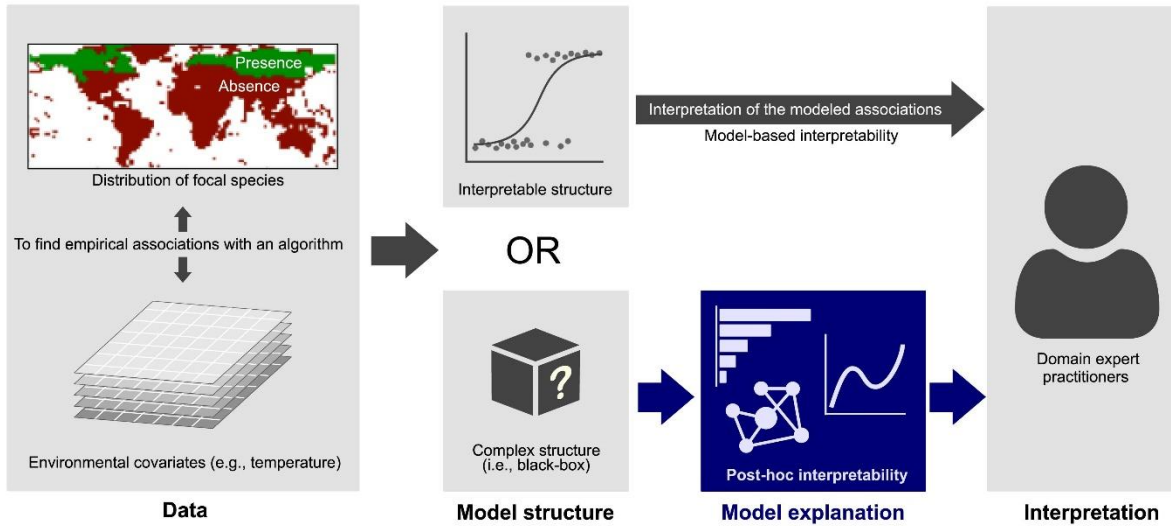


Figure 1. The role of explainable Artificial Intelligence (xAI) in species distribution modeling. Interpretable machine learning methods either target a direct understanding of model architecture (i.e., model-based interpretability) or interpret the model by analyzing the model behavior (i.e., how predictions react to certain inputs; post-hoc interpretability). Many methods of the latter kind are model-agnostic, meaning that they can be used for any model, while the former methods are specific for certain model classes.

To explain how xAI helps ecologists, it will be useful to start with an example. The field of xAI has been developing quickly in recent years, and many new methods have been proposed recently (Table 1, see also Molnar 2019, Murdoch et al. 2019, Biecek and Burzykowski 2020, Boehmke and Greenwell 2020). From those, we selected as an example the Local Interpretable Model-agnostic Explanation (LIME), a post-hoc interpretation method proposed by Ribeiro et al. (2016). “Post-hoc” means that it is implemented after the model has been fit, and model-agnostic means that it is usable for any complex model.

Table 1. Model-agnostic post-hoc methods in explainable Artificial Intelligence (xAI), their approaches, and potential use for species distribution models (SDMs). Model-agnostic means that they can be used for any model. Note that the list may not fully cover all available methods. For the “level” column, “local” means that the method is applicable for understanding how each prediction is made, while “global” means that it is used for understanding the model learned from the dataset.

Level	Method	Approach	Utility for SDM modellers	Ref.
local	Local Interpretable Model-agnostic Explanation	Explains how the model predicts at a given instance with a simple parametric or machine learning model	Inspects which features are important for the prediction at a given location.	(Ribeiro et al. 2016; this work)
local	Shapley values	Explains the relative contribution of each feature to the prediction at a given instance, based on cooperative game theory	Inspects which features are important for the prediction at a given location.	(Lundberg and Lee 2017)
local	Anchors	Builds if-then rules with some features and their thresholds such that the model does not change the prediction at a given instance	Identifies which features are important and how sensitive the prediction is against changes in the selected features at a given location.	(Ribeiro et al. 2018)
local	iBreakDown	Identifies interactions and	Inspects which feature interactions are	(Gosiewska

		measures their contributions	important for the prediction at a given location.	and Biecek 2020)
global	Permutation importance; multi-level block permutation	Measures the relative contribution of each feature to the model accuracy. The latter accounts for the covariance structure.	Quantifies which variables are important, including the mean and confidence intervals, for the accuracy of the model.	(Winkler et al. 2015)
global	Partial dependence plot; individual conditional expectation plot; accumulated local effects plot	Shows the marginal effects of features on the prediction	Visualizes how the associations between the features and response are modeled.	(Hastie et al. 2001, Goldstein et al. 2014, Apley and Zhu 2019)
global	H-statistic	Measures the interaction strength of paired features based on the stability of partial dependence	Identifies key pairwise interactions, either among all possible combinations or given a specific feature.	(Friedman and Popescu 2008)
global	Greenwell's partial dependence-based variable importance	Measures the interaction strength of paired features based on the stability of partial dependence	Identifies key pairwise interactions, either among all possible combinations or given a specific feature.	(Greenwell et al. 2018)

The aim of LIME is to explain how the fitted complex model creates a prediction for a given instance (i.e., a grid cell or other local neighborhood). To this end, for each instance, LIME fits a “local surrogate” model (a simple model; e.g. a logistic regression or decision tree) that approximates the behavior of the complex model for a limited area of the n -dimensional space defined by the predictor variables. Searching for the local surrogate model is formulated as $\operatorname{argmin} L(f, g, \pi_x) + \Omega(g)$. The term $L(f, g, \pi_x)$ calculates the difference in accuracy between the complex model f (e.g. random forests) and a simple model g (e.g. linear model) at the target prediction x and the surrounding neighborhood of proximity π in the n -dimensional space. The term $\Omega(g)$ is the complexity of the simple model represented as the number of parameters. The LIME algorithm minimizes $L + \Omega$ to replace the complex model by the simpler one, while attempting to avoid losing accuracy. A key assumption of LIME is that the necessary degree of model complexity depends on the data domain for which predictions should be made.

Hence, LIME helps us remove ‘unnecessary’ complexity from a global model to better understand how it arrives at local predictions. Although a complex algorithm may be necessary to accurately model species distributions at coarse spatial scales (e.g., the full species range), a simpler algorithm is often sufficiently accurate at finer scales where conservation and management activities actually take place. In fact, many parameters that would apply to the larger scale are not as important at more local scales, where most of the parameters can often be assumed to be constants (but see Potter et al. 2013).

The simplified model can also be used for model analysis and validation, as we demonstrate in an example where we provide site-level assessment and interpretation for an SDM for the African elephant (Box 1, Fig. 2). Most complex algorithms were primarily designed to improve predictions, and design principles such as boosting, bagging, or deep layers in neural networks usually complicate the interpretation of the fitted model. For example, suppose one fits a random forest model to a focal species with a range of different predictor variables and the model predicts the presence or high suitability for the species at a particular site. One may want to know why the model made such a prediction. For example, is it due to optimal climatic conditions, resource availability, or other reasons? LIME can help to analyze how the importance of the predictor variables changes with scale and/or subregion (Ryo et al. 2018) and which variables are most relevant for a particular location or scenario.

More broadly, xAI methods can help to analyze and approximate the global and local behavior of the model and identify the reasons for why particular predictions are made (although such reasons are not necessarily causal). It is widely appreciated that statistical models can use non-causal predictor variables to make predictions (i.e., the model predicts the right outcome for the wrong reason (Fourcade et al. 2018)). This is not necessarily a problem, because non-causal factors can act as proxies for unobserved and unobservable causal factors to improve predictions. However, the use of such non-causal model structures is problematic when predicting under conditions where the correlation structures of predictor variables change (Dormann et al. 2013). It is therefore important to determine the extent to which the fitted model reflects the true causal structure, and thus the mechanisms actually driving these relationships.

xAI cannot directly answer these questions, but it can help ecologists to examine the question of causality. For example, an xAI analysis may show that model predictions depend on predictor variables that are determined *a priori* as unlikely relevant for the focal species, or that the relevance of predictor variables changes in geographical or environmental space in a way that is ecologically counterintuitive. These results may lead the researcher to reconsider the extent to which the fitted model reflects true mechanistic relationships, as well as the extent to which it can be used for extrapolation or to inform direct management interventions. In such a way, xAI can be combined with ecological and biogeographical knowledge to create a richer and more accurate interpretation of fitted ML models.

In conclusion, we hope that this article will encourage applications of xAI tools in the SDM research domain to strengthen mutual understanding between modelers and practitioners. Expert knowledge from both groups can be used to assess how local predictions are made based on the output of xAI, and this should inform model selection and conservation or management action. To conclude, we think that demystifying the decisions that complex models make is a necessary step towards producing models that can explain real-world ecological data (Mammola et al. 2019, Araújo et al. 2019).

Box 1: Explaining the distribution of the African elephant with xAI

We demonstrate here an application of the LIME approach for SDMs with R (R Core Team 2019), using as an example the distribution of the African elephant (*Loxodonta africana*). The R script to reproduce the analysis with detailed settings is available on Zenodo (<https://doi.org/10.5281/zenodo.3904245>). Note that our intention is purely demonstrational: we seek neither to advance the ecological knowledge of this species nor to adhere to all the best modeling practices (e.g., we did not consider spatial autocorrelation or model tuning).

We applied the random forests algorithm (Breiman 2001b) for modeling the distribution of *L. africana* using occurrence data downloaded from GBIF (Navarro and Jackson in press, Musila et al. 2019, naturgucker_de 2020, Questagame 2020, Ueda 2020), 10,000 randomly sampled background points, and standard bioclimatic variables from WorldClim v2 (Fick and Hijmans 2017). For data acquisition and processing, we used the `sdmbench` package (Angelov 2018), for model training the `m1r` package (Bischi et al. 2016), and for model explanation the `lime` package (Pedersen and Benesty 2019; but note that the `breamDown` package is an alternative, Biecek and Grudziak, 2020). The data was split into training and testing data, 70% and 30% respectively.

Conventionally, model assessment relies heavily on visual inspections of the mapped model predictions (in this case, species' habitat suitability; upper-left panel in Fig. 2), accuracy metrics, variable importance rankings, and variable associations (lower-left panel). In this example, we interpret that the model is accurate when evaluated on testing data (Area Under the ROC Curve = 0.98) and that the most important variables are the precipitation of the wettest quarter and the temperature of the coldest and driest quarter. This interpretation is important for biogeographical understanding, but it does not help us assess how reliable the model is or what the locally important variables are at the local scale, where actual management and/or conservation occurs.

Local surrogates can help alleviate this issue. With LIME we show site-level model validation at three randomly chosen sites (right panel). At site A, the model predicts high habitat suitability (0.95), and we ask why it makes such a prediction. With LIME, we can confirm that the prediction is supported by all top five environmental conditions at the site. At site B, the model also predicts equally high suitability (0.97), but the reasons for the prediction differ from those for site A. As these sites are so distant from one another (approx. 2,500 km), it is reasonable that these sites may be similarly suitable for different reasons. At site C, the model predicts low suitability (0.34) because of a combination of both positive and negative environmental factors. Hence, at site C, a careful investigation may be warranted to confirm the presence or absence of the species.

The habitat suitability at site A (0.95) is slightly lower than that at site B (0.97), although at site A all predictor variables support the prediction while at site B one variable is against the prediction (i.e., temperature seasonality). This is potentially because of (i) the effects of the other variables that are ranked lower than five and/or (ii) the local surrogate model did not perfectly explain the global model. We do not intend to solve these issues in this exercise, but they can be taken as potential caveats of LIME.

As demonstrated, individual LIME explanations for local sites can help us better explore spatial variations in variable importance, which in turn, can contribute to more reasonable conservation and management decisions with higher interpretability for the model at the local scale.

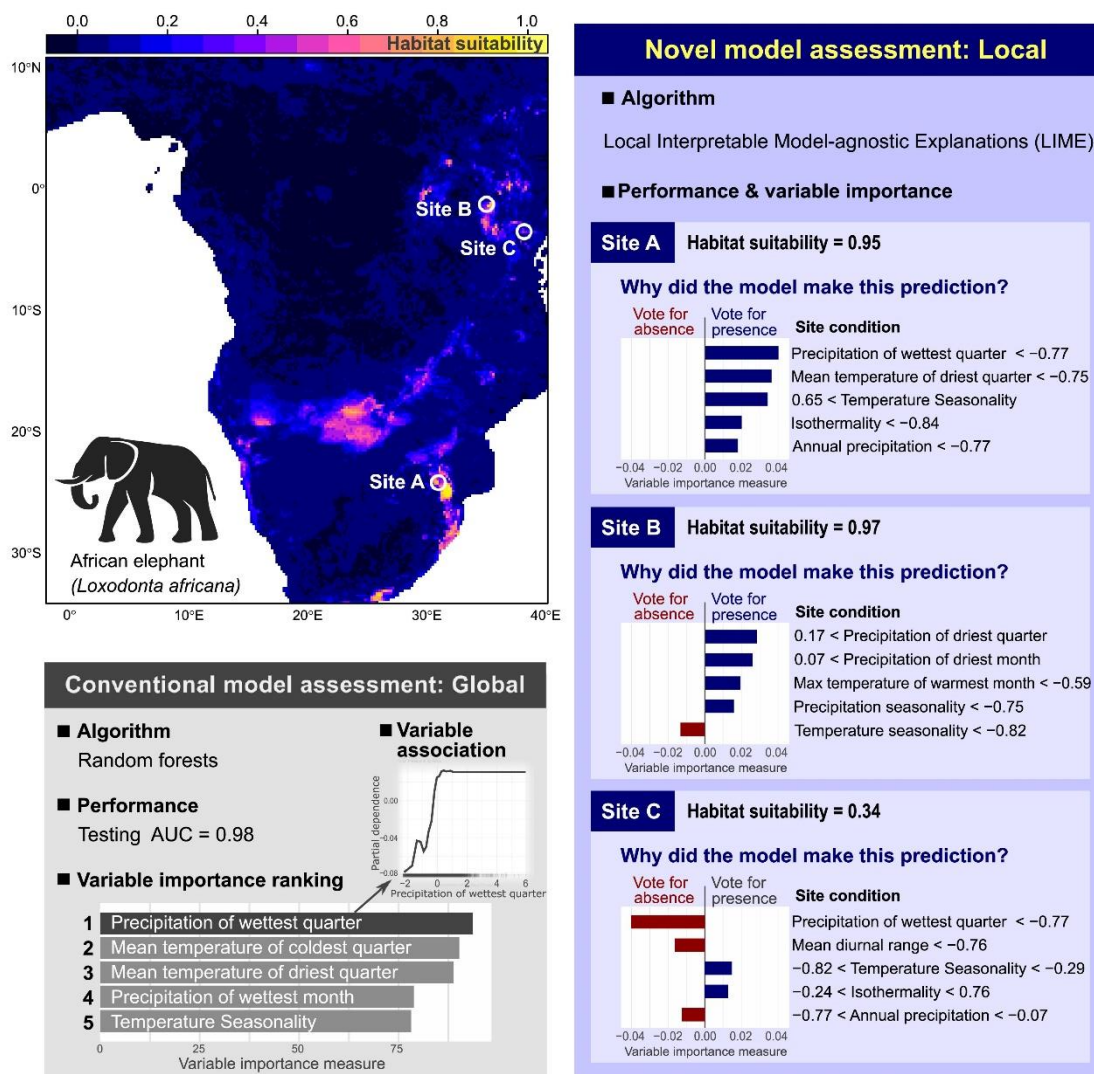


Figure 2. Interpreting the species distribution model of the African elephant (*Loxodonta africana*) based on model assessment at scales relevant to both biogeographic processes and conservation and/or management (global and local). Model interpretation at the local scale applies Local Interpretable Model-agnostic Explanations (LIME), an explainable artificial intelligence (xAI) technique (see Table 1 for other techniques).

References

- Angelov, B. 2018. sdmbench: R package for benchmarking species distribution models. – J. Open Source Softw. 3: 847.
- Angelov, B. 2019. Review of species distribution modeling open-source software. – doi: 10.13140/RG.2.2.33804.67203.
- Apley, D. W. and Zhu, J. 2019. Visualizing the Effects of Predictor Variables in Black Box Supervised Learning Models. – arXiv:1612.08468v2.
- Araújo, M. B. et al. 2019. Standards for distribution models in biodiversity assessments. – Sci. Adv. 5: eaat4858.
- Austin, M. 2007. Species distribution models and ecological theory: A critical assessment and some possible new approaches. – Ecol. Model. 200: 1–19.
- Biecek, P. and Grudziak, A. 2018. breakDown: Model Agnostic Explainers for Individual Predictions. – The R Journal 10: doi: 10.32614/RJ-2018-072.

- Biecek, P. and Burzykowski, T. 2020. Explanatory Model Analysis - Explore, Explain and Examine Predictive Models. – Available at <https://pbiecek.github.io/ema/>
- Bischl, B. et al. 2016. mlr: Machine Learning in R. – J. Mach. Learn. Res. 17: 1–5.
- Boehmke, B. and Greenwell, B. 2020. Hands-On Machine Learning with R. – Available at <https://bradleyboehmke.github.io/HOML/>.
- Booth, T. H. 2018. Why understanding the pioneering and continuing contributions of BIOCLIM to species distribution modelling is important. – Austral Ecol. 43: 852–860.
- Booth, T. H. et al. 2014. bioclim: the first species distribution modelling package, its early applications and relevance to most current MaxEnt studies. – Divers. Distrib. 20: 1–9.
- Box, E. O. 1981. Predicting physiognomic vegetation types with climate variables. – Vegetatio 45: 127–139.
- Breiman, L. 2001a. Statistical Modeling: The Two Cultures. – Stat. Sci. 16: 199–215.
- Breiman, L. 2001b. Random Forests. – Mach. Lang. 45: 5–32.
- Carvalho, D. V. et al. 2019. Machine Learning Interpretability: A Survey on Methods and Metrics. – Electronics 8: 832.
- Christin, S. et al. 2019. Applications for deep learning in ecology. – Methods Ecol. Evol. 10: 1632–1644.
- Cutler, D. R. et al. 2007. Random Forests for Classification in Ecology. – Ecology 88: 2783–2792.
- de Souza Muñoz, M. E. et al. 2011. openModeller: a generic approach to species' potential distribution modelling. – Geoinformatica 15: 111–135.
- Dormann, C. F. et al. 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. – Ecography 36: 27–46.
- Elith, J. and Leathwick, J. R. 2009. Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. – Annu. Rev. Ecol. Evol. Syst. 40: 677–697.
- Elith, J. and Graham, C. H. 2009. Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. – Ecography 32: 66–77.
- Elith, J. et al. 2008. A working guide to boosted regression trees. – J. Anim. Ecol. 77: 802–813.
- Elith, J. et al. 2011. A statistical explanation of MaxEnt for ecologists. – Divers. Distrib. 17: 43–57.
- Fick, S. E. and Hijmans, R. J. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. – Int. J. Climatol. 37: 4302–4315.
- Fourcade, Y. et al. 2018. Paintings predict the distribution of species, or the challenge of selecting environmental predictors and evaluation statistics. – Glob. Ecol. Biogeogr. 27: 245–256.
- Friedman, J. H. and Popescu, B. E. 2008. Predictive learning via rule ensembles. – Ann. Appl. Stat. 2: 916–954.
- Goldstein, A. et al. 2014. Peeking Inside the Black Box: Visualizing Statistical Learning with Plots of Individual Conditional Expectation. – arXiv:1309.6392v2.
- Gosiewska, A. and Biecek, P. 2020. Do Not Trust Additive Explanations. – arXiv:1903.11420v3.
- Greenwell, B. M. et al. 2018. A Simple and Effective Model-Based Variable Importance Measure. – arXiv:1805.04755v1.
- Guisan, A. and Thuiller, W. 2005. Predicting species distribution: offering more than simple habitat models. – Ecol. Lett. 8: 993–1009.
- Hastie, T. et al. 2001. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. – Springer Science & Business Media.
- Kass, J. M. et al. 2018. Wallace: A flexible platform for reproducible modeling of species niches and distributions built for community expansion. – Methods Ecol. Evol. 9: 1151–1156.
- Lundberg, S. and Lee, S-I. 2017. A Unified Approach to Interpreting Model Predictions. – arXiv:1705.07874v2.
- Mammola, S. et al. 2019. Associations between habitat quality, body size and reproductive fitness in the alpine endemic spider *Vesubia jugorum*. – Glob. Ecol. Biogeogr. 28: 1325–1335.
- Merow, C. et al. 2013. A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. – Ecography 36: 1058–1069.
- Merow, C. et al. 2014. What do we gain from simplicity versus complexity in species distribution models? – Ecography 37: 1267–1281.
- Meske, C. and Bunde, E. 2020. Using Explainable Artificial Intelligence to Increase Trust in Computer Vision. – arXiv:2002.01543v1.

- Molnar, C. 2019. Interpretable Machine Learning. – Available at <https://christophm.github.io/interpretable-ml-book/>
- Murdoch, W. J. et al. 2019. Definitions, methods, and applications in interpretable machine learning. – *Proc. Natl. Acad. Sci.* 116: 22071–22080.
- Musila, S. et al. 2019. Occurrence records of mammal species in Tana River Basin, Kenya v1.1. – *Natl. Mus. Kenya*. – doi: 10.15468/0msz3d. Accessed via GBIF.org on 2020-05-30.
- Naturgucker de 2020. naturgucker occurrence dataset. – doi:10.15468/uc1apo. Accessed via GBIF.org on 2020-05-30.
- Navarro, R. and Jackson, C. Kenya Virtual Museum Records v1.11. – doi:10.15468/jolylt. Accessed via GBIF.org on 2020-05-30.
- Olden, J. D. and Jackson, D. A. 2002. Illuminating the “black box”: a randomization approach for understanding variable contributions in artificial neural networks. – *Ecol. Model.* 154: 135–150.
- Olden, J. D. et al. 2008. Machine learning methods without tears: a primer for ecologists. – *Q. Rev. Biol.* 83: 171–193.
- Pecl, G. T. et al. 2017. Biodiversity redistribution under climate change: Impacts on ecosystems and human well-being. – *Science*. 355:eaai9214.
- Pedersen, T. L. and Benesty, M. 2019. lime: Local Interpretable Model-Agnostic Explanations. – Available at <https://CRAN.R-project.org/package=lime>.
- Phillips, S. J. 2017. A Brief Tutorial on Maxent. – Available at https://biodiversityinformatics.amnh.org/open_source/maxent/ Accessed on 2020-06-22.
- Phillips, S. J. et al. 2004. A maximum entropy approach to species distribution modeling. – *Proc. Twenty-First Int. Conf. Mach. Learn.* 655–662.
- Phillips, S. J. et al. 2006. Maximum entropy modeling of species geographic distributions. – *Ecol. Model.* 190: 231–259.
- Potter, K. A. et al. 2013. Microclimatic challenges in global change biology. – *Glob. Change Biol.* 19: 2932–2939.
- Qiao, H. et al. 2015. No silver bullets in correlative ecological niche modelling: insights from testing among many potential algorithms for niche estimation. – *Methods Ecol. Evol.* 6: 1126–1136.
- Questagame 2020. Questagame weekly feed, occurrence dataset. – doi:10.15468/slqqt8. Accessed via GBIF.org on 2020-05-30.
- R Core Team 2019. R: A Language and Environment for Statistical Computing. Vienna, Austria. Retrieved from <https://www.R-project.org/>.
- Ribeiro, M. T. et al. 2016. “Why Should I Trust You?”: Explaining the Predictions of Any Classifier. – *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*: 1135–1144.
- Ribeiro, M. T. et al. 2018. Anchors: High-Precision Model-Agnostic Explanations. – *AAAI Conference on Artificial Intelligence* Available at <https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/16982>.
- Ryo, M. et al. 2018. Nonlinear higher order abiotic interactions explain riverine biodiversity. – *J. Biogeogr.* 45: 628–639.
- Ryo, M. and Rillig, M. C. 2017. Statistically reinforced machine learning for nonlinear patterns and variable interactions. – *Ecography*. 8: e01976.
- Scachetti-Pereira, R. 2002. DesktopGarp: a software package for biodiversity and ecologic research. – Available at <http://www.nhm.ku.edu/desktopgarp/>.
- Ueda, K. 2020. iNaturalist Research-grade Observations. iNaturalist.org. Occurrence dataset. – doi:10.15468/ab3s5x. Accessed via GBIF.org on 2020-05-30.
- Winkler, A. M. et al. 2015. Multi-level block permutation. – *NeuroImage* 123: 253–268.