EGU22-7529
https://doi.org/10.5194/egusphere-egu22-7529
EGU General Assembly 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Global maps from local data: Towards globally applicable spatial prediction models

Marvin Ludwig[1,2], Álvaro Moreno Martínez[3], Norbert Hölzel[1], Edzer Pebesma[2], and Hanna Meyer[1]
[1]Institute of Landscape Ecology, WWU Münster, Münster, Germany
[2]Institute for Geoinformatics, WWU Münster, Münster, Germany
[3]Image Processing Laboratory, Universitat de València, Valencia, Spain

Global-scale maps are an important tool to provide ecologically relevant environmental variables to researchers and decision makers. Usually, these maps are created by training a machine learning algorithm on field-sampled reference data and the application of the resulting model to associated information from satellite imagery or globally available environmental predictors. However, field samples are often sparse and clustered in geographic space, representing only parts of the global environment. Machine learning models are therefore prone to overfit to the specific environments they are trained on - especially when a large set of predictor variables is utilized. Consequently, model validations have to include an analysis of the models transferability to regions where no training samples are available e.g. by computing the Area of Applicability (AOA, Meyer and Pebesma 2021) of the prediction models.

Here we reproduce three recently published global environmental maps (soil nematode abundances, potential tree cover and specific leaf area) and assess their AOA. We then present a workflow to increase the AOA (i.e. transferability) of the machine learning models. The workflow utilizes spatial variable selection in order to train generalized models which include only predictors that are most suitable for predictions in regions without training samples. We compared the results to the three original studies in terms of prediction performance and AOA. Results indicate that reducing predictors to those relevant for spatial prediction, leads to a significant increase of model transferability without significant decrease of the prediction quality in areas with high sampling density.

Meyer, H. & Pebesma, E. Predicting into unknown space? Estimating the area of applicability of spatial prediction models. *Methods in Ecology and Evolution* 2041–210X.13650 (2021) doi:10.1111/2041-210X.13650.