



COMMENT



<https://doi.org/10.1038/s41467-022-29838-9>

OPEN

Machine learning-based global maps of ecological variables and the challenge of assessing them

Hanna Meyer ^{1✉} & Edzer Pebesma ^{2✉}

The recent wave of published global maps of ecological variables has caused as much excitement as it has received criticism. Here we look into the data and methods mostly used for creating these maps, and discuss whether the quality of predicted values can be assessed, globally and locally.

Fields such as ecology or geosciences have seen a strong increase of studies that apply machine learning methods to produce global maps of environmental variables (prominent examples are, e.g., the global tree restoration potential¹, global soil nematode abundances², or global soil maps³) with the aim of increasing our knowledge about the environment, and of supporting decisions. These maps are often distributed as open data, allowing other researchers to use them as input to compute indicators of all kinds or as input to map yet other variables. Quality measures reported by the authors are impressive but often contradict with experts' opinions (e.g., see comments to Bastin et al.¹ or discussions in Wyborn and Evens⁴). Ploton et al.⁵ attribute this contradiction to the use of validation strategies that ignore spatial autocorrelation in the data, and argue in favor of using spatial cross-validation methods. Wadoux et al.⁶ argue that spatial cross-validation is not the right way to evaluate map accuracy. Meyer and Pebesma⁷ argue that the practice of using sparse and non-representative reference data makes model assessment impossible for areas with conditions that are very different from the training data. Here, we try to unravel some of these arguments by focusing on the data, the methods used, and the limits to our ability to assess spatial predictions.

Global reference data used in machine learning applications

In common global predictive mapping tasks (described in, e.g., Van den Hoogen et al.⁸), models are trained using reference data from field sampling. These data are then spatially matched with predictor variables with global coverage. A machine learning model (often Random Forest) is then fitted (trained) and applied to the predictors to obtain a global map with predicted values of the target variable.

Most machine learning methods as well as common validation strategies assume that the reference data are independent and identically distributed, which is in the spatial mapping context for instance guaranteed when they were obtained as a simple random sample from the target area. It is, however, hard to imagine that a global, spatially random sample will ever be collected when it involves taking in situ samples (e.g., collecting soil parameters, or counting soil nematodes). None of the global studies mentioned above is based on data collected as a probability sample; most of them are based on creating a database by merging all data available from

¹Institute of Landscape Ecology, Westfälische Wilhelms-Universität Münster, Heisenbergstraße 2, Münster 48149, Germany. ²Institute for Geoinformatics, Westfälische Wilhelms-Universität Münster, Heisenbergstraße 2, Münster 48149, Germany. ✉email: hanna.meyer@uni-muenster.de; edzer.pebesma@uni-muenster.de

different sources. As a consequence, these data are strongly concentrated, e.g., in Europe and Northern America, and within these regions, they are extremely clustered around areas that received intense research. We are aware that large gaps in geographic space do not always imply large gaps in feature space, but it is the former that most concerns accuracy of the maps of focus here, as we will discuss.

For three publicly available datasets that were used for global mapping, Fig. 1A–C compares the distributions of the spatial distances of reference data to their nearest neighbor (pink) with the distribution of distances from all points of the global land surface to the nearest reference data point (prediction locations, blue). The difference between the two distributions reflects the degree of spatial clustering in the reference data: Fig. 1D shows

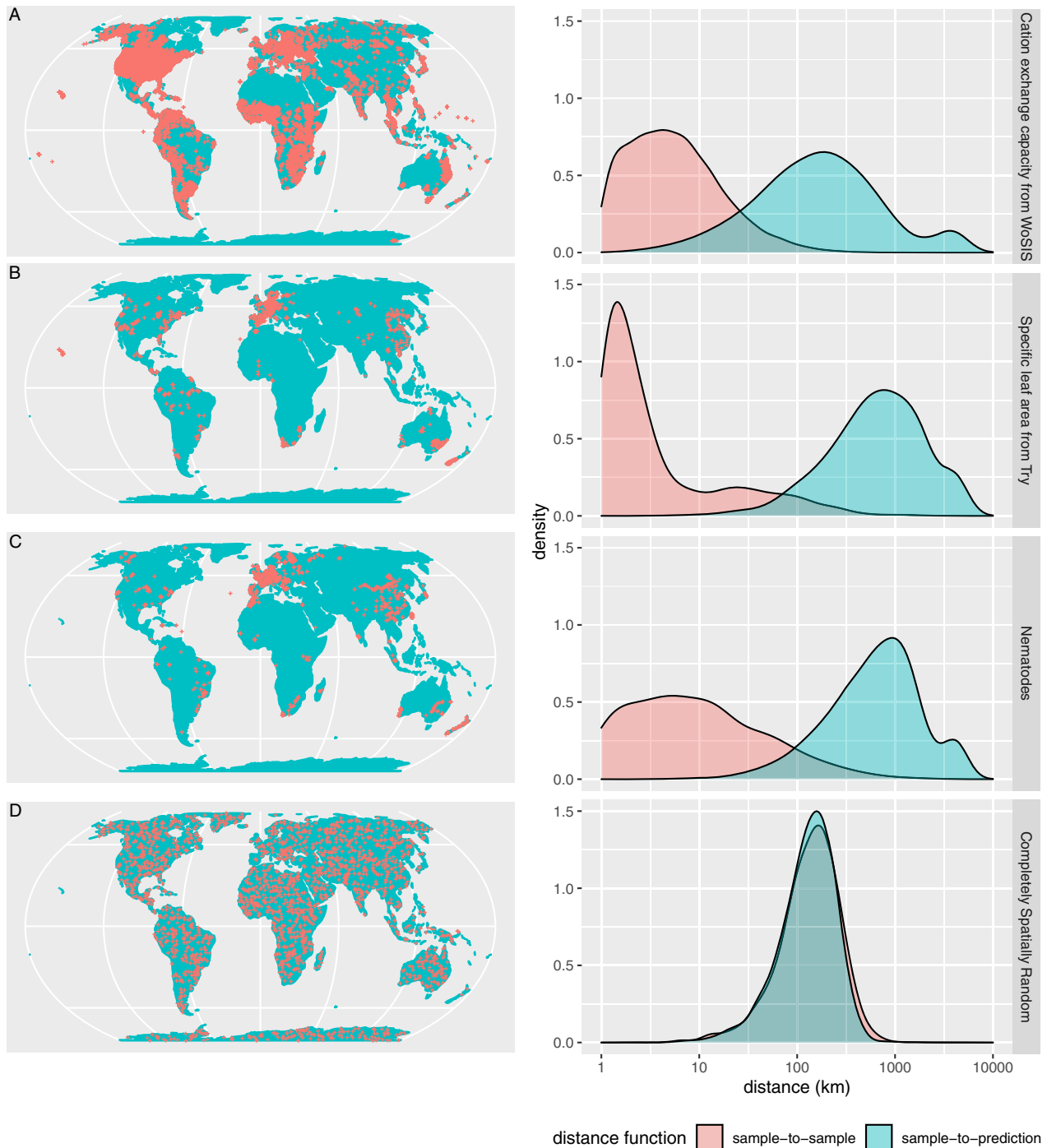


Fig. 1 Spatial distance distributions in global mapping studies. Spatial distribution (left; equal Earth projection) and distribution of nearest neighbor distances (right; sample-to-sample distance in pink, prediction-location-to-sample distance in blue) for three different publicly available datasets: cation exchange capacity in the soil from the WoSIS database²³ as used for global soil mapping³ (**A**), specific leaf area from the TRY database²⁴ as used for the global mapping in Moreno-Martinez et al. (2018)²⁵ (**B**), and the nematodes dataset compiled by Van den Hoogen et al. (2019)² (**C**). For comparison, the fourth dataset is a simulated completely spatially random sample of the same size as the nematode dataset (**D**). Distance distributions were calculated and visualized using the R package “CAST”²⁶.

the distributions for a simulated spatially random sample of the same size as Fig. 1C. The clustered pattern has certain consequences and raises challenges for accuracy assessment that we will discuss in the following.

Map quality: global or local assessment?

The quality of global maps can be assessed in different ways. One way is global assessment where a single statistic is chosen to summarize the quality of the entire map: the map accuracy. For a categorical variable, this can be the probability that for a randomly chosen location on the map, the map value corresponds to the true value. For a continuous variable, it can be the RMSE, describing for a randomly chosen location on the map the expected difference between the mapped value and the true value. When a probability sample, such as a completely spatially random sample, is available for the area for which a global assessment is needed, then map accuracy can be estimated model-free (also called design-based, e.g., by using the unweighted sample mean in case of a completely spatially random sample). This circumvents modeling of spatial correlation because observations are independent by design^{6,9}. This approach is called model-free because no model needs to be assumed about the distribution or correlation of the data: the only source of randomness is the random selection of sample units from a target population. If a probability sample is not available this approach cannot be used, and automatically the accuracy assessment approach becomes model-based¹⁰, which involves modeling a spatial process by assuming distributions and taking spatial correlations into account, and choosing estimation methods accordingly.

Using naive random n -fold or leave-one-out cross-validation methods (or a simple random train-test split) to assess global model quality (usually equated with map accuracy) makes sense when the data are independent and identically distributed. When this is not the case, dependencies between nearby samples, e.g., in a spatial cluster, are ignored and result in biased, overly optimistic model assessment, as shown in, e.g., Ploton et al.⁵. Alternative cross-validation approaches such as spatial cross-validation^{5,11} that control for such dependencies are the only way to overcome this bias. Different spatial cross-validation strategies have been developed in the past few years, all aiming at creating independence between cross-validation folds^{5,11–13}. Cross-validation creates prediction situations artificially by leaving out data points and predicting their value from the remaining points. If the aim is to assess the accuracy of a global map, the prediction situations created need to resemble those encountered while predicting the global map from the reference data (see Fig. 1 and discussions in Milà et al.¹⁴). This occurs naturally when reference data were obtained by (completely spatially random) probability sampling, but in other cases, this has to be forced for instance by controlling spatial distances (spatial cross-validation). Such forcing, however, is only possible when the distances in space that need to be resembled are available in the reference data. In the extreme case where all reference data come from a single cluster, this is impossible. When all reference data come from a small number of clusters, larger distances are available between clusters but do not provide substantial independent information about variation associated with these distances. Lack of information about larger distances means that we cannot assess the quality of predictions associated with such distances and cannot properly estimate global quality measures. Alternative approaches such as experiments with synthetic data¹⁵ or a validation using independent data at a higher level of integration¹⁶ would then be options to support confidence in the predictions.

Another way of accuracy assessment is local assessment: for every location, a quality measure is reported, again as probability

or prediction error. Such a local assessment predicts how close the map value is to newly observed values at particular locations. If the measurement error is quantified explicitly, a smoother, measurement-error-free value may be predicted¹⁰. If the model accounts for change of support^{10,17}, prediction errors may refer to average values over larger areas such as 1×1 , 5×5 , or 10×10 km grid cells. Examples of local assessment in the context of global ecological mapping are modeled prediction errors using Quantile Regression Forests¹⁸ or mapped variance of predictions made by ensembles^{1,2}. Neither of these examples quantifies spatial correlation or measurement error, or addresses change of support, as it is known from other modeling frameworks¹⁹. By omitting to model the spatial process, the local accuracy estimates as presented in the global studies that motivated this comment are disputable.

The difference between global and local assessment is striking, in particular for global maps. A global, single number averages out all variability in prediction errors, and obscures any differences, e.g., between continents or climate zones. It is of little value for interpreting the quality of the map for particular regions.

Limits to accuracy assessment

Maps, and in particular global maps, create a strong feeling of satisfaction, suggesting we now know it all. They are however also used, enlarged, torn apart, read in detail, and may form the basis for local decisions of all kinds, or even form the inputs for follow-up models. If a global map does not come with clear instructions about its value, like a prescription for subsequent use, it is easy to abuse it. Wyborn and Evans⁴ rightly ask about “what changes are global maps, and their creators, trying to bring about in the world?”, and suggest a re-engagement with empirical studies of local and regional contexts while seeking co-construction with those having local knowledge. The fact that creating global maps of anything nowadays is so easy does not mean these maps are always useful.

Technically, a trained Random Forest (or other) model can be applied globally as long as global predictors are available. Predictions far beyond reference data, however, often lead to extrapolation situations in the predictor space and models produce typically meaningless predictions when provided with predictor values that do not resemble the training data. The same applies to local accuracy estimates when based on the variance of predictions⁷. A good coverage of training data in the predictor space is hence required to produce globally applicable predictions. Since distances in geographic space often go along with distances in the feature space, it can be assumed that this is not given for many prediction models that are based on sparse and clustered reference data. In Meyer and Pebesma⁷, we suggest a procedure to limit spatial predictions to the area of applicability of the model: global maps would need to gray out areas where predictor values are too different from values in the training data—the areas for which we cannot assess the quality of predictions. Similar approaches have been suggested and discussed, e.g., by Jung et al.¹⁶. Limiting predictions to the area of applicability of the model is not only relevant to avoid wrong conclusions about prediction patterns but also to avoid propagation of large errors: many global maps of environmental variables used the global soil maps produced by Hengl et al.³ as input predictors^{1,2,20}. The global soil maps by Hengl et al.³ in turn used other modeled maps as an input (e.g., WorldClim²¹). If the latter maps had labeled locations with predictions for which quality cannot be assessed, or for which quality was really low, the follow-up study could have benefited from it. Without that information, both WorldClim and the soil layers were taken as if they contained true values.

We argue that showing predicted values on global maps without reliable indication of global and local prediction errors or

the limits of the area of applicability, and distributing these for reuse, is not congruent with basic scientific integrity. Reusing such global maps while ignoring prediction errors amplifies this problem, hence more transparency and clear indication about the limitations of predictions is required. Global maps are being distributed digitally and could be used for purposes of decision making, e.g., in the context of nature conservation²². We call for global maps of ecological variables to be published only when they are accompanied by properly derived local and global accuracy measures.

Received: 8 October 2021; Accepted: 1 March 2022;

Published online: 22 April 2022

References

- Bastin, J.-F. et al. The global tree restoration potential. *Science* **365**, 76–79 (2019).
- Van den Hoogen, J. et al. Soil nematode abundance and functional group composition at a global scale. *Nature* **572**, 194–198 (2019).
- Hengl, T. et al. Soilgrids250m: global gridded soil information based on machine learning. *PLoS One* **12**, e0169748 (2017).
- Wyborn, C. & Evans, M. C. Conservation needs to break free from global priority mapping. *Nat. Ecol. Evol.* **5**, 1322–1324 (2021).
- Ploton, P. et al. Spatial validation reveals poor predictive performance of large-scale ecological mapping models. *Nat. Commun.* **11**, 4540 (2020).
- Wadoux, A. M.-C., Heuvelink, G. B., de Bruin, S. & Brus, D. J. Spatial cross-validation is not the right way to evaluate map accuracy. *Ecol. Modell.* **457**, 109692 (2021).
- Meyer, H. & Pebesma, E. Predicting into unknown space? Estimating the area of applicability of spatial prediction models. *Methods Ecol. Evol.* **12**, 1620–1633 (2021).
- Van den Hoogen, J. et al. A geospatial mapping pipeline for ecologists. Preprint at *bioRxiv* (2021).
- Stehman, S. V. Basic probability sampling designs for thematic map accuracy assessment. *Int. J. Remote Sens.* **20**, 2423–2441 (1999).
- Cressie, N. *Statistics for Spatial Data* rev edn (John Wiley & Sons, 1993).
- Roberts, D. R. et al. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography* **40**, 913–929 (2017).
- Valavi, R., Elith, J., Lahoz-Monfort, J. J. & Guillerá-Arroita, G. blockcv: an R package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. *Methods Ecol. Evol.* **10**, 225–232 (2019).
- Wenger, S. J. & Olden, J. D. Assessing transferability of ecological models: an underappreciated aspect of statistical validation. *Methods Ecol. Evol.* **3**, 260–267 (2012).
- Milà, C., Mateu, J., Pebesma, E. & Meyer, H. Nearest neighbour distance matching Leave-One-Out Cross-Validation for map validation. *Methods in Ecology and Evolution*. **00**, 1–13 (2022).
- Jung, M., Reichstein, M. & Bondeau, A. Towards global empirical upscaling of fluxnet eddy covariance observations: validation of a model tree ensemble approach using a biosphere model. *Biogeosciences* **6**, 2001–2013 (2009).
- Jung, M. et al. Scaling carbon fluxes from eddy covariance sites to globe: synthesis and evaluation of the fluxcom approach. *Biogeosciences* **17**, 1343–1365 (2020).
- Chiles, J.-P. & Delfiner, P. *Geostatistics: Modeling Spatial Uncertainty* 2nd edn (John Wiley & Sons, 2012).
- Hengl, T., Nussbaum, M., Wright, M., Heuvelink, G. & Gräler, B. Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. *PeerJ* **6**, e5518 (2018).
- Wikle, C. K. Hierarchical models in environmental science. *Int. Stat. Rev.* **71**, 181–199 (2003).
- Ma, H. et al. The global distribution and environmental drivers of aboveground versus belowground plant biomass. *Nat. Ecol. Evol.* **5**, 1110–1122 (2021).
- Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. & Jarvis, A. Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* **25**, 1965–1978 (2005).
- Schmidt-Traub, G. National climate and biodiversity strategies are hamstrung by a lack of maps. *Nat. Ecol. Evol.* **5**, 1325–1327 (2021).
- Batjes, N. H., Ribeiro, E. & van Oostrum, A. Standardised soil profile data to support global mapping and modelling (wosis snapshot 2019). *Earth Syst. Sci. Data* **12**, 299–320 (2020).
- Kattge, J. et al. TRY plant trait database – enhanced coverage and open access. *Glob. Change Biol.* **26**, 119–188 (2020).
- Moreno-Martinez, A. et al. A methodology to derive global maps of leaf traits using remote sensing and climate data. *Remote Sens. Environ.* **218**, 69–88 (2018).
- Meyer, H. & Ludwig, M. CAST: ‘caret’ applications for spatial-temporal models. R package version 0.6.0. <https://CRAN.R-project.org/package=CAST> (2022).

Author contributions

H.M. and E.P. contributed equally to this work.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Hanna Meyer or Edzer Pebesma.

Peer review information *Nature Communications* thanks Markus Reichstein for their contribution to the peer review of this work.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022