



BOTTOMS-UP

Opt-In

Preliminary title: Remote sensing and deep learning for continental-scale species richness predictions

Target journals: Global Ecology and Biogeography

Aims

Deep learning (DL) algorithms are gaining attention in ecology (Christin et al., 2019). Recent applications include the integration of deep neural networks to produce joint species distribution models (Chen et al., 2018; Pichler & Hartig, 2021). The possibility to model the distribution of multiple species from presence data of distant taxonomic groups maintaining individual species information opens novel scenarios for investigating biological diversity.

The wide application of remote sensing data in ecology and the abundance of such data from various sources offer possibilities for integrating data in novel ways. For instance, raw spectral Landsat data were successfully tested as predictors of bird species distribution recently using more traditional approaches (Hopkins et al., 2022). Increased accessibility and processing power through free platforms like Google Earth Engine (Gorelick et al., 2017) can greatly optimize the use of large datasets. Most remarkably, remotely-sensed data can help disentangle patterns of species richness over multiple spatial scales (Rocchini et al., 2011). However, the high number of variables, sites, and species that can be fed in a model may also represent a challenge for reliable inferences. The potentially complex unknown functional form of the species-environment relationship further complicates inference with traditional parametric models. In this context, deep learning algorithms can represent a viable alternative to more traditional approaches to investigate the spatial patterns of biological diversity. DL allows to flexibly combine different sources of data without having to make assumptions about the functional form of the predictors, as is the case with traditional approaches such as linear regression models.

The large data retrievable from remote sensors, the large number of taxa sampled in Bottoms-Up and the development of new joint species distribution models fit with deep learning algorithms provide ground for exploring new hypothesis. They allow to combine advantages of both approaches – the flexibility and predictive superiority of DL and the idea of accounting for unobserved biotic components by using JSDMs. The proposed models have been recently published (Pichler & Hartig, 2021) and preliminary analysis aimed at assessing model performance, ran on the ConFoBi (partner of Bottoms-Up) data were promising (results available on request). The proposed objectives of this opt-in are:

- (1) to assess the relative weight of remotely sensed environmental and climatic data in predicting species occurrence;
- (2) to identify patterns of co-occurring species along environmental gradients and spatial scales.
- (3) to evaluate at which spatial scale species richness of different taxa, functional groups and guilds is best predicted.

The underlying hypothesis are:

- (1) satellite raw spectral data can be used to predict species occurrence and richness of multiple taxa
- (2) species richness of upper-level taxa and co-occurring species can be reliably predicted at different spatial scales

As site-level predictors, we plan to use remotely-sensed data, mainly derived from satellites, in the form of raw spectral band values and satellite-derived indices, such as the NDVI or the NDWI (Campos et al., 2012; Pettorelli et al., 2005). Furthermore, climatic data will also be employed as predictors using the CHELSA database (Karger et al., 2017).

First/lead author: Marco Basile (WSL)

Core authors from BOTTOMS-UP: Sabina Burrascano, ConFoBi (Simone Fontana, Sara Klingenfuss, Carlos Miguel Landivar, Liu Xiang)

Core authors outside BOTTOMS-UP: Maximillian Pichler (Uni Regensburg), Francesco Valerio (Uni Évora)

Please note that if the outline changes substantially (more than 1 aim is revised substantially), or co author(s) are added to the above lists the governing board should re-vote on the project.

Further opt-in authors:

According to the BOTTOMS-UP Bylaws any member of the BOTTOMS-UP Consortium can declare his/her interest to become opt-in author. The first author is required to preliminarily accept one such offer from each dataset that represents at least 2% of the data in the analysis. It is upon the discretion of the first author whether to accept any opt-in offer beyond this requirement. Persons interested in opt-in authorship can be nominated until 01.12.2022 with e-mail to the first author (and cc: to the BOTTOMS-UP Governing Board), explaining which dataset(s) they represent and preferentially also how they could contribute. Note however that such a nomination only means the option to become co-author. In the end, only those persons will be retained as actual co-authors who have made a significant intellectual contribution to the paper during the course of its preparation (in accordance with BOTTOMS-UP Bylaws and compliance to ethics in academy).

Data to be used:

I need presence and abundance data of species. When a site has been sampled multiple times, I need also repeated sampling data. Data needs to be geolocated and at the smallest spatial scale available. I need data on all available taxonomic groups.

Time line:

Deadline for permission of data usage from custodians: 30.06.2022

Extraction of data from BOTTOMS-UP: 31.07.2022

Data preparation and analysis: 30.10.2022

Raw results to be sent to the wider author team: 15.12.2022

Workshop with the wider author team: 30.01.2023

Writing up of the paper (including preparation/optimization of figures): 30.03.2023

Feedback round of co-authors to MS: 30.04.2023

Submission: 30.05.2023

References

- Campos, J. C., Sillero, N., & Brito, J. C. (2012). Normalized difference water indexes have dissimilar performances in detecting seasonal and permanent water in the Sahara-Sahel transition zone. *Journal of Hydrology*, 464–465, 438–446. <https://doi.org/10.1016/j.jhydrol.2012.07.042>
- Chen, D., Xue, Y., & Gomes, C. (2018). *End-to-End Learning for the Deep Multivariate Probit Model*.
- Christin, S., Hervet, É., & Lecomte, N. (2019). Applications for deep learning in ecology. *Methods in Ecology and Evolution*, 10(10), 1632–1644. <https://doi.org/10.1111/2041-210X.13256>
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202, 18–27. <https://doi.org/10.1016/j.rse.2017.06.031>
- Hopkins, L. M., Hallman, T. A., Kilbride, J., Robinson, W. D., & Hutchinson, R. A. (2022). A comparison of remotely sensed environmental predictors for avian distributions. *Landscape Ecology*, 37(4), 997–1016. <https://doi.org/10.1007/s10980-022-01406-y>
- Karger, D. N., Conrad, O., Böhrner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., Zimmermann, N. E., Linder, H. P., & Kessler, M. (2017). Climatologies at high resolution for the earth's land surface areas. *Scientific Data*, 4, 1–20. <https://doi.org/10.1038/sdata.2017.122>
- Pettorelli, N., Vik, J. O., Mysterud, A., Gaillard, J.-M., Tucker, C. J., & Stenseth, N. C. (2005). Using the satellite-derived NDVI to assess ecological responses to environmental change. *Trends in Ecology & Evolution*, 20(9), 503–510. <https://doi.org/10.1016/j.tree.2005.05.011>
- Pichler, M., & Hartig, F. (2021). A new joint species distribution model for faster and more accurate inference of species associations from big community data. *Methods in Ecology and Evolution*, 2021(March), 1–15. <https://doi.org/10.1111/2041-210x.13687>
- Rocchini, D., McGlenn, D., Ricotta, C., Neteler, M., & Wohlgemuth, T. (2011). Landscape complexity and spatial scale influence the relationship between remotely sensed spectral diversity and survey-based plant species richness. *Journal of Vegetation Science*, 22(4), 688–698. <https://doi.org/10.1111/J.1654-1103.2010.01250.X>

Confirmation:

I confirm that I will adhere to the BOTTOMS-UP Bylaws.

Date

20.04.2022

Signature



New Zha