PREDICTING VEGETATION ATTRIBUTES WITH NEURAL NETWORKS AND SENTINEL-1 & 2

Javier Muro[1*], Anja Linstädter[2], Florian Alfred Männer[3], Lisa-Maricia Schwarz[3], Janik Hoffmann[1], Olena Dubovyk[4]

[1] Center for Remote Sensing of Land Surfaces (ZFL), University of Bonn, Germany
[2] Biodiversity Research/Systematic botany, University of Potsdam, Germany
[3] Institute of Crop Science and Resource Conservation (INRES), University of Bonn, Germany
[4] Geography Department, University of Bergen, Bergen

KEY WORDS: biodiversity; vegetation; land management; grasslands; forest; remote sensing; multitemporal

ABSTRACT:

Evidence suggests that plant traits, plant functional diversity, and species diversity are linked to ecosystem functions to different extents. However, these relationships are sometimes inconsistent because of the presence of environmental gradients (e.g. climate, topography, land use) and scale mismatches between sampling units and landscape processes. Relationships between satellite data and vegetation parameters seem to be also case-specific, which hinders the creation of generalizable models. We have built predictive models of structural parameters and species composition across a broad range of climatic and topoedaphic conditions and management practices across grasslands and forests in Germany. For that, we use Sentinel multitemporal imagery and neural networks. Our models manage to explain 50% of the data variability for structural parameters, show high stability, and can generalize well across environmental gradients and sites. We also found that prediction models of biodiversity parameters show lower predictive capabilities. Spatially continuous models of grassland and forest attributes provide vital information on ecosystem functions at landscape scale. Thus, they can contribute to studying the feedback mechanisms between biodiversity, ecosystem functions, and land management at the scales to which ecological processes occur.

1. INTRODUCTION

New generations of remote sensing sensors and machine learning approaches can predict vegetation characteristics with varying accuracies. However, studies often fail to cover a sufficiently broad range of environmental conditions, and evidence suggests that prediction models are often case specific (Verrelst et al., 2019). Also, spatial dependencies are often not addressed, which translates into overly optimistic models (Meyer and Pebesma, 2021; Ploton et al., 2020). Radiative transfer models are in principle applicable broadly since they don't depend on local training data. However, they might suffer from equiafinity (several input variables can yield the same spectral response).

Deep learning algorithms, such as convolutional neural networks, can identify higher hierarchical patterns in the data and show a superior performance when compared to traditional machine learning algorithms such as decision trees. But they still need to be trained with representative data. While classification labelled data is often available from different sources (e.g. land cover data), continuous data about vegetation parameters is scarcer due to higher costs, which limits the applicability of models to the area they were trained. For instance, many researchers have used multispectral sensors to predict leaf area index (Schwieder et al., 2020), obtaining different relationships depending on the environmental conditions. Figure 1 shows the relationship between leaf area index (LAI) and biomass in grasslands for three sites across Germany in 4 consecutive years. Even though LAI is a good proxy for biomass, their relation is dependent on environmental conditions.

To address these gaps, we have built a deep learning model to predict several structural and biodiversity parameters across heterogenous site conditions in temperate grasslands and forests across Germany. We test its generalization capabilities and accuracies, and discuss their potentials and limitations.

Figure 1: Grassland above ground biomass vs. Leaf Area Index (LAI) derived with PROSAIL from Sentinel-2.

*Corresponding author: jmuro@uni-bonn.de

2. STUDY AREA

The Biodiversity Exploratories (BE) is a research infrastructure project that hosts multiple researches in ecology and land management (Fischer et al., 2010). The BE are formed by a set of 300 plots (150 in grasslands and 150 in forests) spread over three regions that follow a North-South gradient, and are characterized by different soil types, topography, altitude, historical contexts, climatic conditions and management intensities (Figure 2). In grasslands the land management gradient involves extensive and intensive grazing, from none to high fertilization input, and different frequencies of mowing. In forests, the plots cover different development stages of various forest types, such as beech, pine and spruce mixtures, and different frequencies of harvesting.

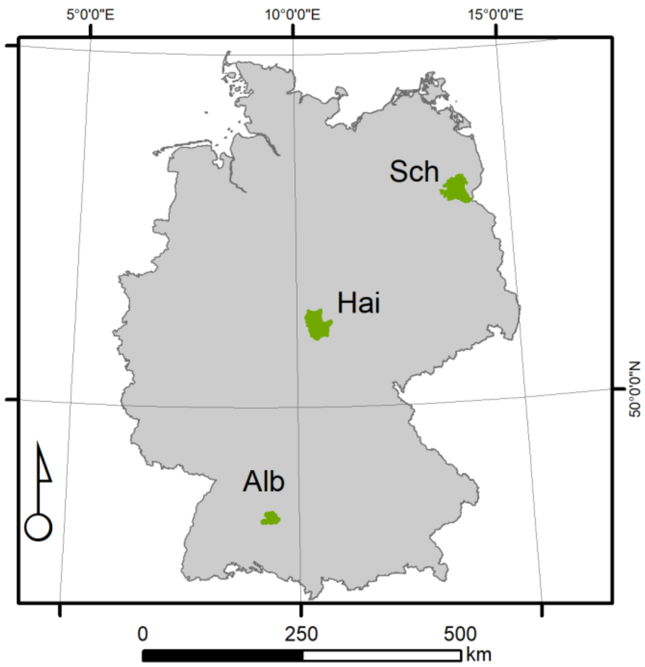

Figure 2: Location of the three BE: Schorfheire, Hainich, and Schwäbische Alb.

3. METHODS

3.1 Satellite data

A synthetic cloudless time series of Sentinel-2 imagery was created using FORCE processing software (Frantz, 2019). For Sentinel-1, we created median multitemporal backscatter composites for summer and winter season in Google Earth Engine (Hoffmann et al., 2022). The study period is ranges from 2017-2020 for grasslands, and 2014-2018 for forests.

3.2 Field data

Forest data was collected through several forest inventories from 2014 to 2018 (Schall and Ammer, 2018), and grassland data was collected every spring from 2017 to 2020 (Bolliger et al., 2021) (Figure 3, Table 1) .
Both grassland and forest datasets can be accessed from the Biodiversity Exploratories Platform (BexIS https://www.bexis.uni-jena.de/).

Two parameters were analysed for each ecosystem; a structural parameter and a biodiversity parameter.

In forests, the structural information is given by the standard deviation of the diameter at breast height (DBH_std). This parameter acts as a proxy of habitat heterogeneity, not only for plants, but also for bird and insect species. The species information is given by the tree species richness within the surveyed plots.

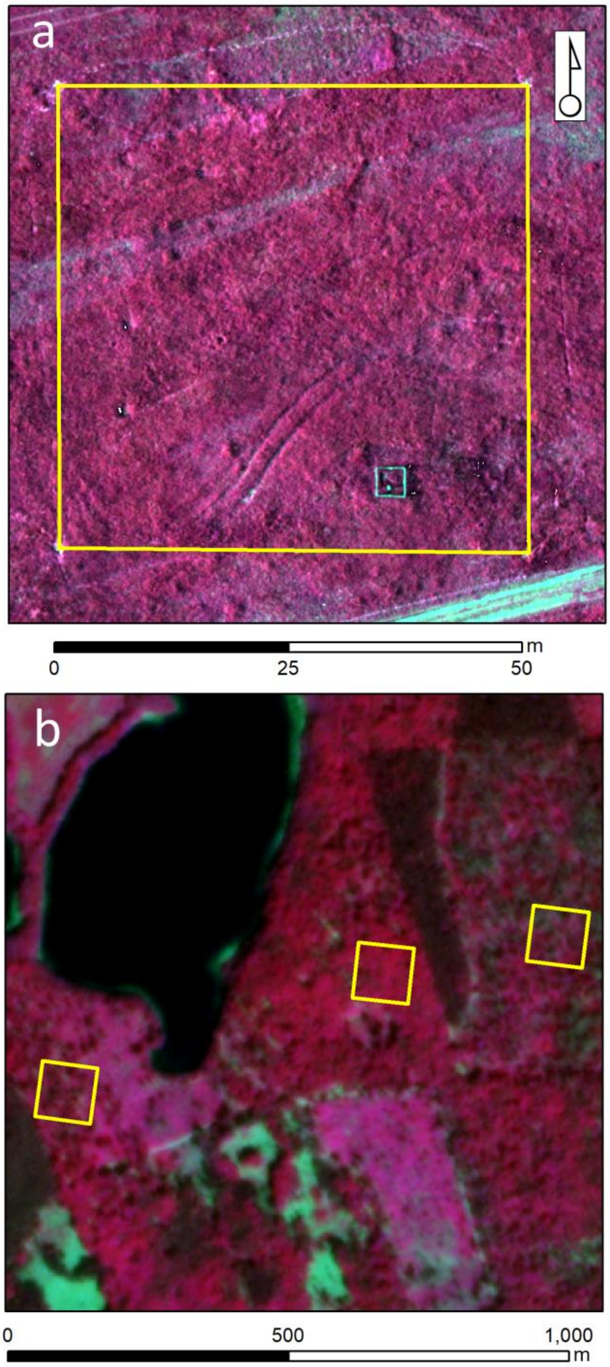

Figure 3: a) UAV false color image of a grassland plot, 50x50 m. b) Planet Dove false color image of three forest plots of different types and mixtures of species and age class (100x100 m).

Table 1: Number of observations available in each exploratory for grasslands and forest plots.

	Grasslands	Forests
Schorfheide-Chorin	152	50
Hainich-Dün	179	50
Swaibian Alb	171	50

In grasslands, the structural information is given by the plant height, a proxy for biomass productivity and habitat type. The biodiversity information analysed was the Shannon index, which measures the diversity of species in a community.

3.3 Analysis

A feed-forward neural network was built in Python using Keras and Tensorflow (Figure 4) (TensorFlow Developers, 2021). Hyperparameters were determined using Keras-Tuner (O'Malley et al., 2019). A k-fold target-oriented validation was set up to ensure independence between training and validation data.

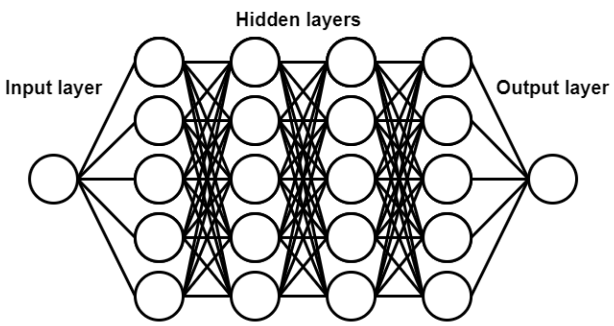

Figure 4: Simplified representation of the feed-forward neural network used.

The importance of the different predictors was obtained using shap python package (Lundberg and Lee, 2017) in the case of grasslands. Shap measures the importance of each predictor in relation to the model, rather than the true importance of that predictor. In the case of forests, the importance of each predictor was evaluated using ablation analyses.

The absence of spatial autocorrelation was corroborated with Morans'I.

4. RESULTS

4.1 Grasslands

4.1.1 Structural parameters

Vegetation height models achieved coefficients of determination (r^2) of 0.43 and Relative Root Mean Squared Errors (RRMSE) of 0.36 (Figure 5). The predictions were quite robust and showed a low bias, especially considering the broad range of grassland types included (wet pastures, meadows, highly fertilized plots, extensive mountain grasslands on slopes…). A few observations of high height values tended to be underestimated. These corresponded to organic rich and swampy grasslands with nettles and other rather uncommon vegetation compositions in temperate grasslands and pastures.

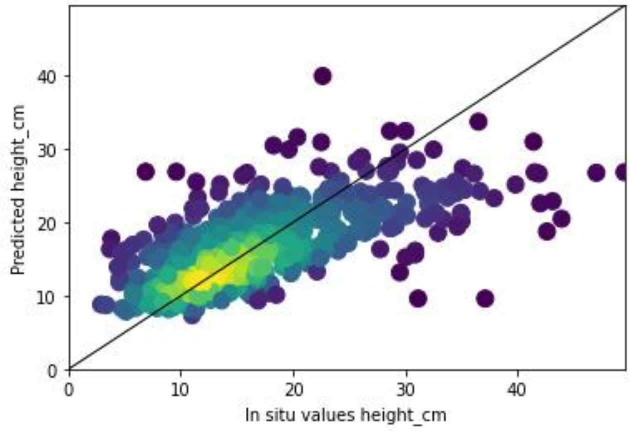

Figure 5: Scatter plot of predicted vs. in situ vegetation height in grasslands.

4.1.2 Biodiversity parameters

The models performed poorlier at predicting species diversity. Shannon index models returned r^2 = 0.23 and RRMSE = 0.20 (Figure 6). The results showed some prediction capabilities, but with high bias at high and low values. For conservation applications, this bias is suboptimal since we are particularly interested in mapping the highly diverse areas for conservation, and the lowly diverse areas for forage production.

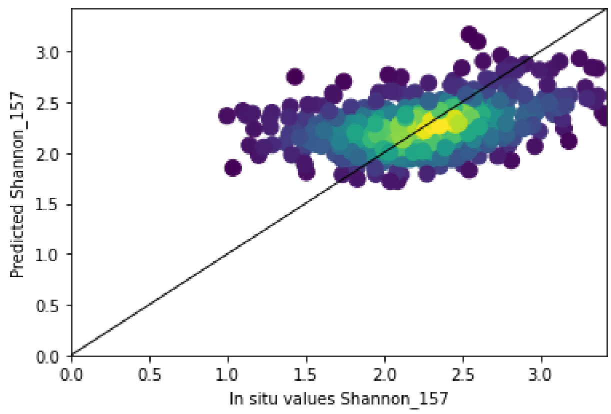

Figure 6: Scatter plot of predicted vs. in situ Shannon index in grasslands.

4.1.3 Interpretability analysis

The Shap analysis indicated that red-edge1 and 2, near-infrared 1 and 2, blue and swir1 bands were most important for predicting plant height.

For predicting diversity indices, near-infrared 1 and 2, red-edge 2 and swir2 were the most important predictors, although shap values varied across the time series (Figure 7).

https://doi.org/10.5194/isprs-archives-XLIII-B3-2022-945-2022

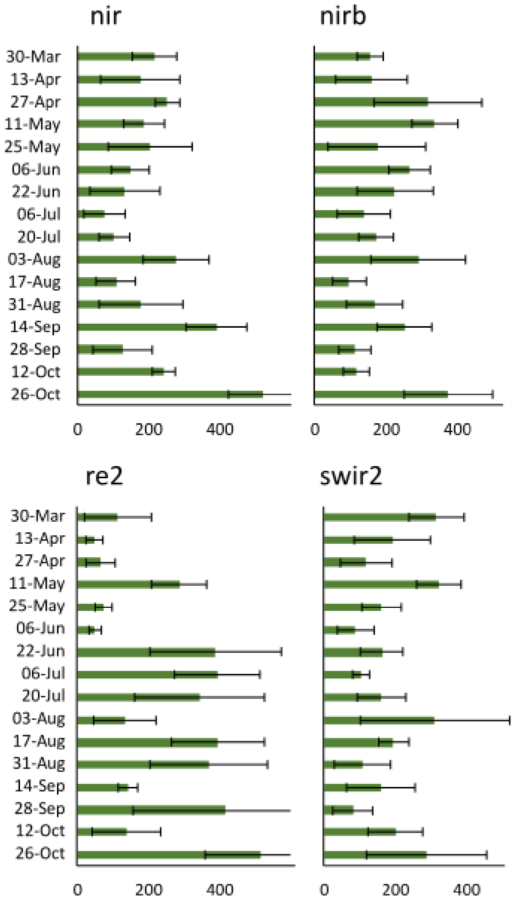

Figure 7: Shap analysis of the importance of each predictor relative to the model. Horizontal axis is unitless.

Images from October tended to show the highest importance in all relevant bands. Peaks of importance were also observed in spring (April-May, depending on the band, Figure 7). This is likely due to spectral and timing differences in the greening and browning phases, reflecting the effects of species composition and land management practices. Land management practices are the main driver of biodiversity in grasslands (Felipe-Lucia et al., 2020). In other words, grassland plots with different biodiversity characteristics will look more different around April and October, when the greening and senescence phases starts.

4.2 Forests

4.2.1 Structural parameters

Sentinel-1 metrics showed a good predictive capabilities of forest structural parameters, (DBH_std), with an $r^2 = 0.51$ and a RRMSE of 0.33. The results showed no systematic bias (Figure 8). Isolated poorer predictions might be caused by forest changes undergone within the inventory and study period (2014-2018), creating mismatches between the image and the field data (trees falling or harvested, bark beetle outbreaks, etc…).

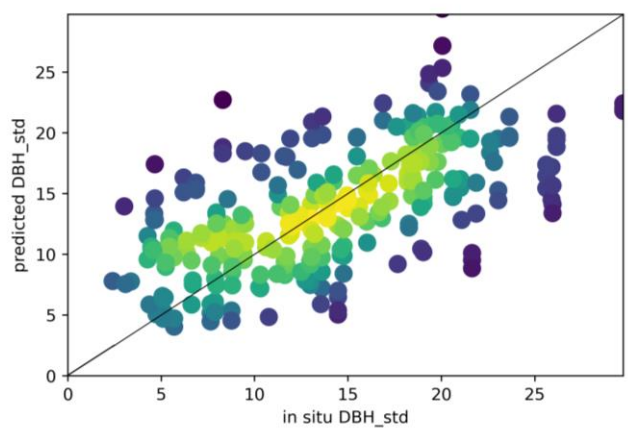

Figure 8: Scatter plot of predicted vs. in situ standard deviation of diameter at breast height (DBH_std) in forests

4.2.2 Biodiversity parameters

Like in the case of grasslands, tree species was not predicted accurately, with large underestimations at high species counts. The highest r^2 was 0.2 and RRMSE was 0.61.

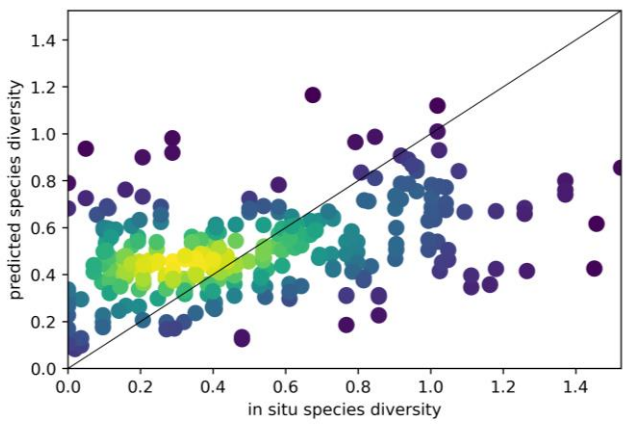

Figure 9: Scatter plot of predicted vs. in situ species diversity in forests, given by Shannon index.

4.2.3 Interpretability analysis

Ablation and leave one out analysis suggested that Sentinel-1 features (especially VV and VH backscatter from winter) were the most effective predictors of structural parameters, in agreement with the results of other researchers (Bae et al., 2019). Since the error rates and bias were so high in the Shannon index predictions, the interpretability of the predictors used could not be evaluated with enough reliability.

5. DISCUSSION

Spatially explicit models of vegetation attributes can improve our understanding of the effects of land management on biodiversity and ecosystem functions at regional scales. Although several studies have modelled species distributions using climatic and geomorphologic data, these are often at a resolution incompatible with land management practices, or for

https://doi.org/10.5194/isprs-archives-XLIII-B3-2022-945-2022 |

monitoring ecosystems under the rapidly changing conditions of the Anthropocene (Randin et al., 2020).

Creating remotely sensed based models of species composition or other biodiversity metrics in seminatural environments is challenging. Some authors have been successful at mapping floristic compositions of rather pristine natural forests at landscape scale (Pérez Chaves et al., 2018), but most studies are successful only at local scale, making models not replicable to other areas.

Monitoring structural parameters such as biomass, LAI, vegetation height, stand density, trunk diameter heterogeneity, or canopy gaps can be a much more reliable predictor of vegetation conditions and changes at regional or continental scales, and be applicable for land management and monitoring.

For instance, plant height in grasslands can be monitored at farm scale (Figure 10). Field data can be conveniently collected using a raising plate meter, and it is a good proxy for standing biomass, and in consequence, for primary productivity if data for the whole season is available.

Leaf area index can be used to study the patterns that govern the global leaf economics spectrum, and field data can also be collected fairly easily, for forests as well as for grasslands. For other parameters, such as stand density, an r^2 of 0.47 and RRMSE of 0.33 were achieved when fusing Sentinel-1 and Sentinel-2 predictors (see Hoffmann et al., (2022) for further details).

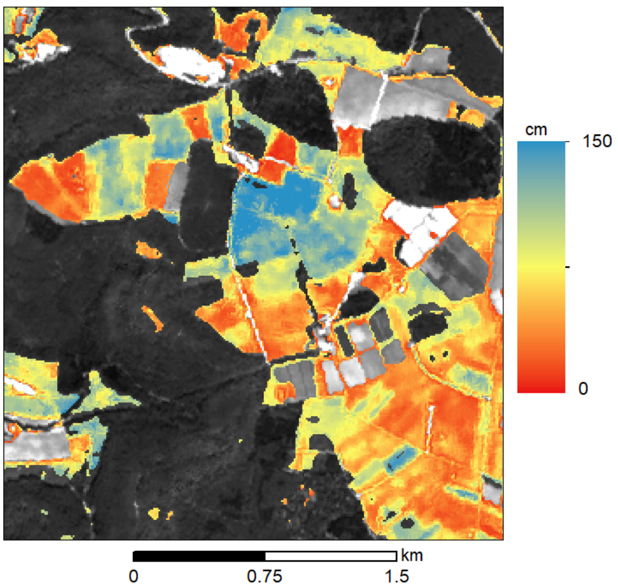

Figure 10: Plant height in grasslands during mid May 2020, modelled with Sentinel-2.

Other forest structural parameters such as gap frequency of DBH_std cannot be mapped at the finest spatial resolution of the satellite, since they vary with area in a non-linear way. In our case, the forest plots were 100 x 100 m, and the DBH_std map has to be given at that resolution, lowering the resolution by one order of magnitude. Despite of that, and despite the salt pepper effects produced by the radar imagery that were used as predictors, the boundaries between different forest types are visible in our results (Figure 11). Conifer plantations showed patterns of low structural diversity, while surrounding deciduous forests exhibited a larger structural diversity.

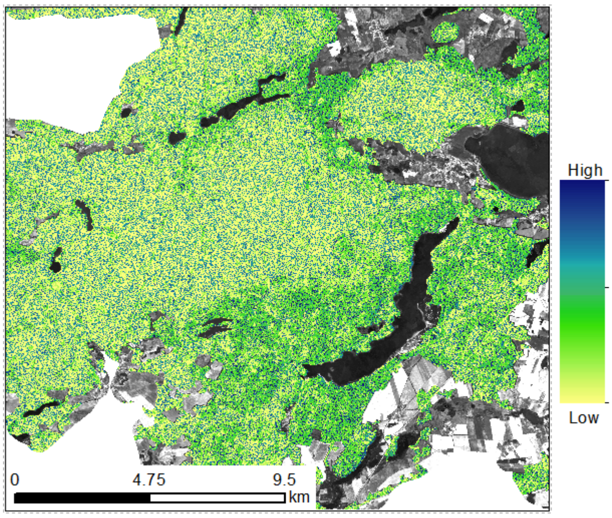

Figure 11: Standard deviation of diameter at breast height (DBH std) in forests, modelled with Sentinel-1 multitemporal features.

6. CONCLUSSIONS

Deep learning models can identify higher hierarchical patterns from time series of satellite imagery, and create scalable maps of vegetation structural parameters, essential for monitoring and studying the consequences of land management practices or environmental disturbances (e.g. drought, pest outbreaks) in ecosystem functions and services. Biodiversity indices, on the other hand, might require of different approaches to be mapped such as classifications with convolutional neural networks (rather than regressions, as we have done). The availability of dense and multi-temporal in-situ observations on various vegetation parameters at fine resolution is an important prerequisite of such analysis.

ACKNOWLEDGEMENTS

We thank managers of the three Exploratories Kirsten Reichel-Jung, Iris Steitz, Sandra Weithmann, Katrin Lorenzen, Juliane Vogt, Martin Gorke, Miriam Teuscher and all former managers for their work in maintaining the plot and project infrastructure; Victoria Grießmeier for giving support through the central office, Michael Owonibi and Andreas Ostrowski for managing the central database, and Markus Fischer, Eduard Linsenmair, Dominik Hessenmöller, Daniel Prati, Ingo Schöning, François Buscot, Ernst-Detlef Schulze, Wolfgang W. Weisser and the late Elisabeth Kalko for their role in setting up the Biodiversity Exploratories project. We thank the administration of the Hainich National Park, the UNESCO Biosphere Reserve Swabian Alb, and the UNESCO Biosphere Reserve Schorfheide-Chorin as well as all landowners for the excellent collaboration. We also thank Peter Schall for providing the field datasets of the second forest inventory in the BExis platform.

This research is part of the project Sensing Biodiversity Across Scales (SeBAS), grant code DU 1596/1-1, and has been funded by the German Research Foundation (DFG) under the priority program 1374.

REFERENCES

Bae, S., Levick, S.R., Heidrich, L., Magdon, P., Leutner, B.F., Wöllauer, S., Serebryanyk, A., Nauss, T., Krzystek, P., Gossner, M.M., Schall, P., Heibl, C., Bässler, C., Doerfler, I., Schulze, E.-D., Krah, F.-S., Culmsee, H., Jung, K., Heurich, M., Fischer, M., Seibold, S., Thorn, S., Gerlach, T., Hothorn, T., Weisser, W.W., Müller, J., 2019. Radar vision in the mapping of forest biodiversity from space. Nat Commun 10, 4757. https://doi.org/10.1038/s41467-019-12737-x

Bolliger, R., Prati, D., Schäfer, D., Fischer, M., 2021. Vegetation Records for 150 Grassland EPs in 2017-2020, Header Data without Species Identities (incl. biomass). Biodiversity Exploratories Information System (Bexis). Dataset IDs=27426, 26151, 24166, 23486.

Felipe-Lucia, M.R., Soliveres, S., Penone, C., Fischer, M., Ammer, C., Boch, S., Boeddinghaus, R.S., Bonkowski, M., Buscot, F., Fiore-Donno, A.M., Frank, K., Goldmann, K., Gossner, M.M., Hölzel, N., Jochum, M., Kandeler, E., Klaus, V.H., Kleinebecker, T., Leimer, S., Manning, P., Oelmann, Y., Saiz, H., Schall, P., Schloter, M., Schöning, I., Schrumpf, M., Solly, E.F., Stempfhuber, B., Weisser, W.W., Wilcke, W., Wubet, T., Allan, E., 2020. Land-use intensity alters networks between biodiversity, ecosystem functions, and services. Proc Natl Acad Sci USA 117, 28140–28149. https://doi.org/10.1073/pnas.2016210117

Fischer, M., Bossdorf, O., Gockel, S., Hänsel, F., Hemp, A., Hessenmöller, D., Korte, G., Nieschulze, J., Pfeiffer, S., Prati, D., Renner, S., Schöning, I., Schumacher, U., Wells, K., Buscot, F., Kalko, E.K.V., Linsenmair, K.E., Schulze, E.-D., Weisser, W.W., 2010. Implementing large-scale and long-term functional biodiversity research: The Biodiversity Exploratories. Basic and Applied Ecology 11, 473–485. https://doi.org/10.1016/j.baae.2010.07.009

Frantz, D., 2019. FORCE—Landsat + Sentinel-2 Analysis Ready Data and Beyond. Remote Sensing 11, 1124. https://doi.org/10.3390/rs11091124

Hoffmann, J., Muro, J., Dubovyk, O., 2022. Predicting Species and Structural Diversity of Temperate Forests with Satellite Remote Sensing and Deep Learning. Remote Sensing 14, 1631. https://doi.org/10.3390/rs14071631

Lundberg, S., Lee, S.-I., 2017. A Unified Approach to Interpreting Model Predictions. Presented at the 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.

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

O'Malley, T., Bursztein, E., Long, J., Chollet, F., Jin, H., Invernizzi, L., others, 2019. KerasTuner.
Pérez Chaves, P., Ruokolainen, K., Tuomisto, H., 2018. Using remote sensing to model tree species distribution in Peruvian lowland Amazonia. Biotropica 50, 758–767. https://doi.org/10.1111/btp.12597

Ploton, P., Mortier, F., Réjou-Méchain, M., Barbier, N., Picard, N., Rossi, V., Dormann, C., Cornu, G., Viennois, G., Bayol, N., Lyapustin, A., Gourlet-Fleury, S., Pélissier, R., 2020. Spatial validation reveals poor predictive performance of large-scale ecological mapping models. Nat Commun 11, 4540. https://doi.org/10.1038/s41467-020-18321-y

Randin, C.F., Ashcroft, M.B., Bolliger, J., Cavender-Bares, J., Coops, N.C., Dullinger, S., Dirnböck, T., Eckert, S., Ellis, E., Fernández, N., Giuliani, G., Guisan, A., Jetz, W., Joost, S., Karger, D., Lembrechts, J., Lenoir, J., Luoto, M., Morin, X., Price, B., Rocchini, D., Schaepman, M., Schmid, B., Verburg, P., Wilson, A., Woodcock, P., Yoccoz, N., Payne, D., 2020. Monitoring biodiversity in the Anthropocene using remote sensing in species distribution models. Remote Sensing of Environment 239, 111626. https://doi.org/10.1016/j.rse.2019.111626

Schall, P., Ammer, C., 2018. Stand Composition Based on 2nd Forest Inventory, All Forest EPs, 2014–2018. Biodiversity Exploratories Information System (BexIS). Dataset ID = 22907.

Schwieder, M., Buddeberg, M., Kowalski, K., Pfoch, K., Bartsch, J., Bach, H., Pickert, J., Hostert, P., 2020. Estimating Grassland Parameters from Sentinel-2: A Model Comparison Study. PFG. https://doi.org/10.1007/s41064-020-00120-1

TensorFlow Developers, 2021. TensorFlow. Zenodo. https://doi.org/10.5281/ZENODO.4724125

Verrelst, J., Malenovský, Z., Van der Tol, C., Camps-Valls, G., Gastellu-Etchegorry, J.-P., Lewis, P., North, P., Moreno, J., 2019. Quantifying Vegetation Biophysical Variables from Imaging Spectroscopy Data: A Review on Retrieval Methods. Surv Geophys 40, 589–629. https://doi.org/10.1007/s10712-018-9478-y