Submitted to Remote Sensing of Environment

Predicting plant biomass and species richness in temperate grasslands across regions, time, and land management with remote sensing and deep learning

Javier Muro[1,2*], Anja Linstädter[3], Paul Magdon[4], Stephan Wöllauer[5], Florian A. Männer[3,6], Lisa-Maricia Schwarz[3,6], Gohar Ghazaryan[7], Johannes Schultz[8], Zbyněk Malenovský[2], Olena Dubovyk[9]

[1] Center for Remote Sensing of Land Surfaces (ZFL), University of Bonn, Germany

[2] Remote Sensing Research Group, University of Bonn, Germany

[3] Biodiversity Research/Systematic Botany, University of Potsdam, Germany

[4] Faculty of Forest Sciences and Forest Ecology, University of Göttingen, Germany

[5] Department of Physical Geography, University of Marburg, Germany

[6] Institute of Crop Science and Resource Conservation (INRES), University of Bonn, Germany

[7] Center for Agricultural Landscape Research (ZALF), Germany

[8] Department of Geography, Ruhr-University Bochum, Germany

[9] Department of Geography, University of Bergen, Norway

***Corresponding author:** jmuro@uni-bonn.de

Abstract:

Spatial predictions of biomass production and biodiversity at regional scale in grasslands are critical to evaluate the effects of management practices across environmental gradients. New generations of remote sensing sensors and machine learning approaches can predict these grassland characteristics with varying accuracy. However, such studies frequently fail to cover a sufficiently broad range of environmental conditions, and their prediction models are often case-specific. To address this gap, we have modelled above-ground biomass and species richness in 150 spatially independent grassland plots of three geographical regions in Germany. These

regions follow a North-South climate gradient and differ in soil types, topography, elevation, climatic conditions, historical contexts, and management intensities. The predictors tested in this study are Sentinel-1 backscatter, Sentinel-2 time series of surface reflectance along with derived vegetation indices and Rao's Q, and a set of topoedaphic variables. We compared the performance of a feed-forward deep neural network (DNN) with a random forest (RF) regression algorithm. The DNN achieved the best estimations of biomass (r^2 = 0.45) when trained with Sentinel-2 surface reflectance only. Moreover, the DNN showed a higher generalizability than RF during spatial cross-validations (i.e., calibrating and validating in different regions, r^2 = 0.38 vs. 0.26). Species richness predictions by both algorithms improved when the full time series of Sentinel-2 surface reflectance values were used (highest r^2 = 0.42 achieved by the DNN), but both performed poorly during spatial cross-validations. Overall, the DNN-based models were more robust than RF models, showed a lower bias and lower systematic error, and required fewer inputs. Explainability analysis indicated that red-edge and near infrared information from May and October was the most relevant to predict species richness. This study presents an important step forward in generating robust spatially explicit predictions of grassland attributes and biodiversity variables across large areas, environmental gradients, and phenological stages.

---

1 Introduction

Grasslands cover one-third of the Earth's terrestrial surface (Latham J et al., 2014) and 70% of its agricultural areas (Reynolds et al., 2005). They are the second largest terrestrial carbon sink (Grace et al., 2006), and host high levels of biodiversity (Petermann and Buzhdygan, 2021). They also provide other critical ecosystem services, including forage for domesticated and wild herbivores (Ferner et al., 2018; Zhao et al., 2020), but also influence the availability of vital resources for other trophic levels, both above and below ground (Hilpold et al., 2018; Le Provost

et al., 2021; Wehner et al., 2021). Unfortunately, the strength and direction of the biodiversity-ecosystem function/biodiversity-ecosystem service (BEF/BES) relationship remains poorly understood (Manning et al., 2019, Pan et al., 2022). Multiple laboratory and small scale field studies have found that biodiversity, as a regulating ecosystem service, tends to show a negative correlation with provisioning services such as fodder production (Allan et al., 2015; Felipe-Lucia et al., 2020; Hilpold et al., 2018), creating a trade-off that has been described as an "efficiency frontier" (Cavender-Bares et al., 2015). However, other small scale experiments have also found that plant biodiversity may have positive effects on multiple ecosystem functions and services, including biomass production (Marquard et al., 2009; Pan et al., 2022; van der Plas, 2019). These contradictory results (with some studies finding trade-offs between biodiversity and productivity; and others observing synergies between them) might also be caused by climatic, edaphic, topographic and historical factors that are known to influence the effects of land management on the BEF/BES relationship (Bernhardt-Römermann et al., 2011; Busch et al., 2018; Le Provost et al., 2021; Müller et al., 2016).

Given the complexity of the BEF/BES relationship, its understanding would greatly benefit from spatially explicit information both on productivity and on biodiversity. In particular, quantitative measurements of plant productivity and biodiversity across environmental gradients play a key role to understand the response of vegetation to the simultaneous effects of disturbances, management practices, and climate change. Multiple studies have used climatic variables from global datasets to model primary production and plant diversity (Cavender-Bares et al., 2020; Randin et al., 2020). However, these models are not applicable at local scales where management decisions influence directly both, the biodiversity and primary productivity, and their related ecosystem functions and services (Busch et al., 2018; Hilpold et al., 2018; Le Provost et al., 2021). Earth observation (EO) offers a unique opportunity to study and express these relationships spatially, at various spatio-temporal scales, and across seasonal phenological

cycles. While multispectral images of optical sensors have been successfully used to model vegetation type (Rapinel et al., 2019), vitality (Reinermann et al., 2019) and phenology (Miao et al., 2017), Synthetic-Aperture Radar (SAR) systems have been used to capture the spatial structure of vegetation and soil moisture (Hill et al., 1999; Muro et al., 2020; Wang et al., 2019).

One way to model vegetation attributes with remote sensing is using statistical data-driven methods, such as machine and deep learning (Ali et al., 2017; Amin et al., 2021; Edirisinghe et al., 2012; Gara et al., 2019; Hill et al., 1999; Klingler et al., 2020; Magiera et al., 2017; Rossi et al., 2019; Schwieder et al., 2020). Model accuracies for biomass predictions reported in the form of coefficient of determination (r^2), tend to range between 0.2 and 0.7. Band ratios such as the enhanced vegetation index (EVI) and normalized difference vegetation index (NDVI) are commonly used as predictors, even though NDVI typically saturates with high amounts of biomass (Aklilu Tesfaye and Gessesse Awoke, 2021). Radiative transfer models have also been used to predict vegetation characteristics with marginally better results when compared to data-driven approaches (Quan et al., 2017; Verrelst et al., 2019). Hybrid retrieval approaches, in which machine learning models are calibrated using synthetic data from radiative transfer models, were shown to be an efficient way of estimating a few quantitative vegetation traits (Ali et al., 2021; Estévez et al., 2022).

The literature on modeling biodiversity indicators of grasslands with satellite imagery (e.g., species richness, Simpson, Shannon or Fisher's alpha indices) is scarcer than for estimating biophysical attributes (Masenyama et al., 2022). Most studies use NDVI as a proxy for productivity, which tends to be related to diversity for unmanaged ecosystems (Wang et al., 2005). Fauvel et al., (2020) predicted Simpson and Shannon indices using Sentinel-1 and -2 bands and derived products, achieving an r^2 of 0.4. Similar approaches were conducted by Lopes et al., (2017), but low accuracies were reported when attempting to predict biodiversity indicators in grassland with SPOT (Satellite Pour l'Observation de la Terre) multispectral time series.

To overcome the difficulty of predicting complex biodiversity indicators with remote sensing, some researchers opted for different strategies. They used, for example, spectral diversity to predict taxonomic or functional diversity (Gholizadeh et al., 2019; Rocchini et al., 2017; Wang et al., 2018). These works are based on the hypothesis that varying plant traits in canopy structure or phenology will cause variations in the optical signals. Nonetheless, Imran et al., (2021) reported inconsistent relationships established between biodiversity indicators and remote sensing metrics across different grasslands in Italy, suggesting that these relations are ecosystem or site-dependent. Gholizadeh et al., (2019) obtained statistically significant relationships between spectral and species diversity for young grassland plots, but no relationship for old grassland plots containing invasive species. Other authors reported successful predictions of floristic gradients using ordination techniques (Feilhauer et al., 2011; Muro et al., 2016) or plant traits (Rossi et al., 2020).

Despite the accurate predictions achieved with machine learning models, we identified three research gaps that currently limit a wider application of remote sensing in predicting grassland characteristics. The first gap is a lack of reference data covering a broad range of environmental conditions. This requirement is often neglected due to challenges in operational implementation of a systematic and standardized sampling design for heterogeneous and large areas. Thus, several published biomass or biodiversity prediction models are built on data with limited variability, (Ali et al., 2021; Fauvel et al., 2020; Gara et al., 2019; Guerini Filho et al., 2020; Schwieder et al., 2020), limiting the transferability of models to other regions. Note that these are not methodological flaws, but rather a consequence of sub-optimal calibration and validation data. In other cases, specific models were established for specific grassland types (Li et al., 2018), which might be feasible for homogenous areas of natural grasslands, but unfeasible for complex mosaics of managed grasslands. Other models based on UAV and terrestrial LiDAR data have

been more successful (Schulze-Brüninghoff et al., 2019) but yet site-specific, considerably more expensive, and of spatially limited aerial coverage.

The second identified gap is neglecting spatial autocorrelation. It occurs when structural dependencies in the data exist, and training and validation samples are split without guaranteeing independency from one another, for instance, randomly (Ploton et al., 2020). As a result, the prediction models suffer from overfitting, which means that their accuracies are overestimated and do not hold when models are applied to new data (Meyer and Pebesma, 2021; Ploton et al., 2020). Spatial autocorrelation is still often not addressed in many published studies (Ali et al., 2021; Dusseux et al., 2022; Gara et al., 2019; Guerini Filho et al., 2020; Li et al., 2018; Meng et al., 2020; Wang et al., 2018; Zeng et al., 2019), and neglecting it in model cross-validation may lead to its drastic overestimation, significant model inaccuracy, and major disagreements between different mapping approaches and between local and global scales (Meyer et al., 2019; Meyer and Pebesma, 2021; Ploton et al., 2020).

The third research gap is the absence of consensus regarding predictors to model biomass and biodiversity indicators. The best-performing spectral band combinations for the same biophysical variables differs across studies (Heiskanen et al., 2013; Mariotto et al., 2013; Rossi et al., 2019; Verrelst et al., 2019). For instance, Guerini Filho et al., (2020) and Villoslada Peciña et al., (2021) reported varying predictors' importance across grassland sites when predicting biomass. Grigera et al., (2007) found that the relationship between NDVI and biomass was significantly different between sown upland pastures and naturalized lowland pastures of South America. Zhou et al., (2014) reported that the relationship between vegetation indices and gross primary productivity of Chinese grasslands varied across latitude, seasons, and drought vs. none-drought years. Dusseux et al., (2022) generated new spectral combinations to predict pasture height. The use of spectral Rao's Q (Rocchini et al., 2017) as predictor of diversity has also been recently questioned under certain settings (Fassnacht et al., 2022; Pacheco-Labrador et al., 2022).

148 The inconsistencies in results across several published studies indicate a potential lack of their
149 generalization ability (Ali et al., 2017; Heiskanen et al., 2013; Mariotto et al., 2013; Rossi et al.,
150 2019; Verrelst et al., 2019). Meyer and Pebesma (2022) offer a further discussion about the
151 limitations of machine learning maps at global and local scales. Recent studies suggest that deep
152 learning (DL), a subfield of machine learning, might generate more accurate and robust models
153 (Kattenborn et al., 2021; Yuan et al., 2020) due to their capabilities to process and organize large
154 amounts of information and to identify higher hierarchical relations between predictors and
155 response variables. An additional advantage is the unnecessity of feature engineering to
156 determine appropriate data transformations (Kattenborn et al., 2021). Feature engineering
157 requires expert knowledge and heuristic decisions, introducing subjective uncertainty and
158 multicollinearity. Applications of quantitative DL regressions with satellite imagery in grasslands
159 are scarce. Since most studies focus on cropland classification and yield (Campos-Taberner et
160 al., 2020; Csillik et al., 2018; Ma et al., 2019; Osco et al., 2020; Zhang et al., 2020), on feature
161 extraction (Hoeser et al., 2020; Hoeser and Kuenzer, 2020; Smith et al., 2019), or on mowing
162 events (Lange et al., 2022; Lobert et al., 2021; Schwieder et al., 2022), a comprehensive analysis
163 of DL capabilities and their interpretability to predict grassland vegetation attributes quantitatively
164 is still missing.

165 To address the three identified research gaps in modelling grassland above-ground biomass and
166 species richness across temperate regions, seasonal phenophases, and management intensities,
167 we formulated the following research hypotheses:

168 - DL regressions, trained and applied on satellite imagery, can efficiently predict biomass
169 and species richness across environmental and management gradients, and
170 - DL algorithms are more accurate and robust in generalization of their predictability than
171 more traditional random forest approaches.

We test our hypotheses by performing (i) a thorough analysis of the performance of random forest (RF) vs. feed-forward deep neural networks (DNN), (ii) a k-fold cross-validation scheme that eliminates spatial and temporal training and testing data dependencies, (iii) a spatial cross-validation to test the generalization capacity of both algorithms, and (iv) an examination of the spatial applicability of the prediction models. Finally, we present the biomass and species richness maps for three regions in Germany and discuss their use, potentials, and limitations.

2 Study area

The Biodiversity Exploratories (hereto called 'exploratories') are a set of three research regions stretching across Germany covering a wide range of environmental conditions, characterized by distinct abiotic and biotic factors (Figure 1). These regions are: The Biosphere Reserve Schorfheide-Chorin (SCH), Hainich National Park and surroundings (HAI), and the Biosphere Reserve Schwäbische Alb (ALB). All three regions can be categorized as temperate grasslands favoring species adapted to this vegetation type. However, their species richness is dependent on the local soil typology and orography, in combination with specific management strategies. For instance, grasslands with very shallow soil or on steep slopes cannot be managed intensively. Each exploratory site consists of 50 experimental grassland plots, representing management practices ranging from intensive to extensive. Some plots are cattle grazed with varying intensities, some pastures are mown at certain seasonal intervals, others are mulched, rolled or left unused. In addition, different levels of fertilization are applied. To quantify management levels in a standardized way, a Land Use Intensity (LUI) index accounting for mowing frequency, grazing intensity and fertilization input was calculated (Blüthgen et al., 2012), and recently upscaled using neural networks of time series of satellite images (Lange et al., 2022). More detailed information on plot selection and site information can be found in Fischer et al., (2010) and Socher et al., (2013).

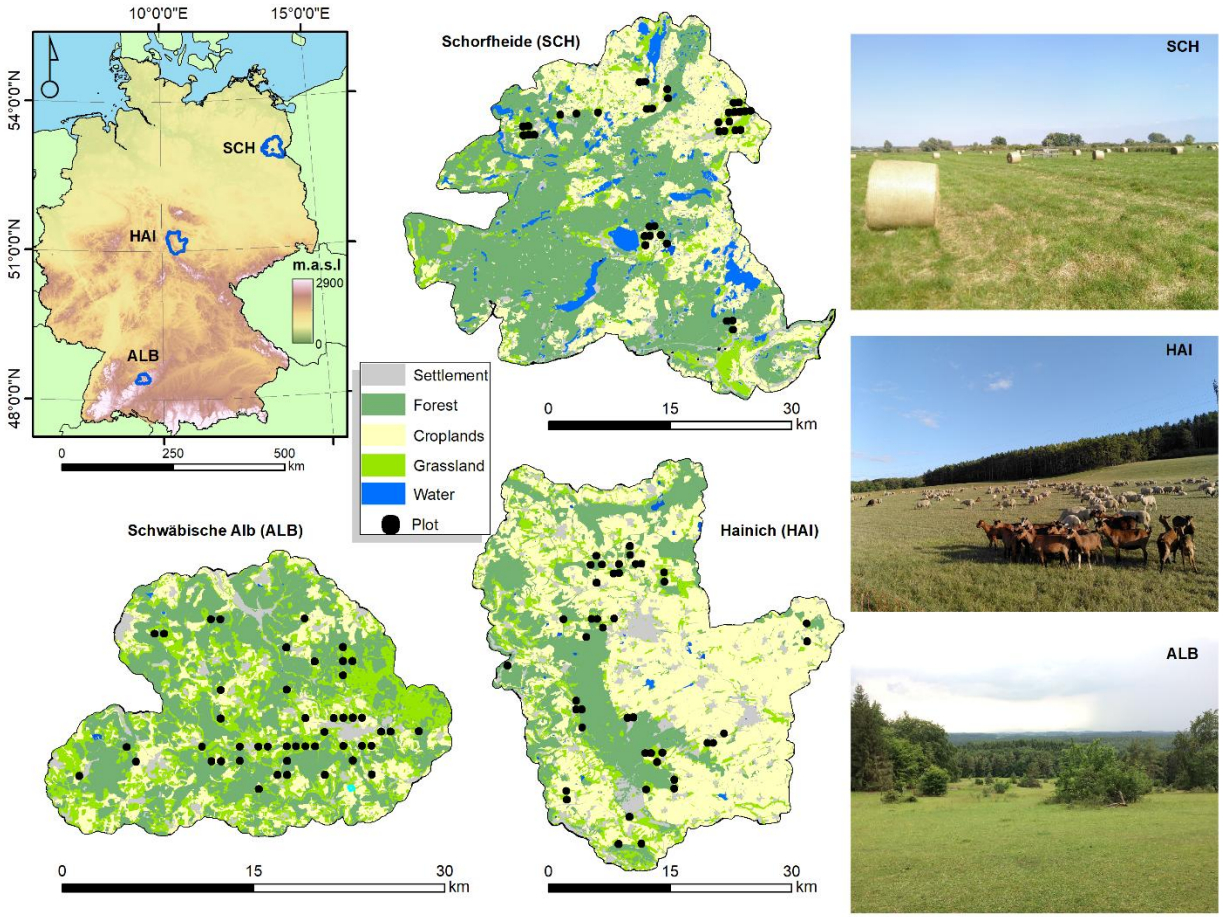


Figure 1: Location of the Biodiversity Exploratories and the experimental plots (approximate). Land use maps modified from: Digital Model Landscape ATKIS (BKG, 2020). The pictures on the right side show 3 typical management practices in the exploratories: i) mowing of extensive grasslands in SCH, ii) high intensity grazing in pastures in HAI, and iii) low intensity management of protected grasslands in ALB.

3 Methods

3.1 Field data

Plant species inventories (Bolliger et al., 2021a) and aboveground biomass (Bolliger et al., 2021b) were collected using a standardized sampling protocol from the 150 plots during the time of peak

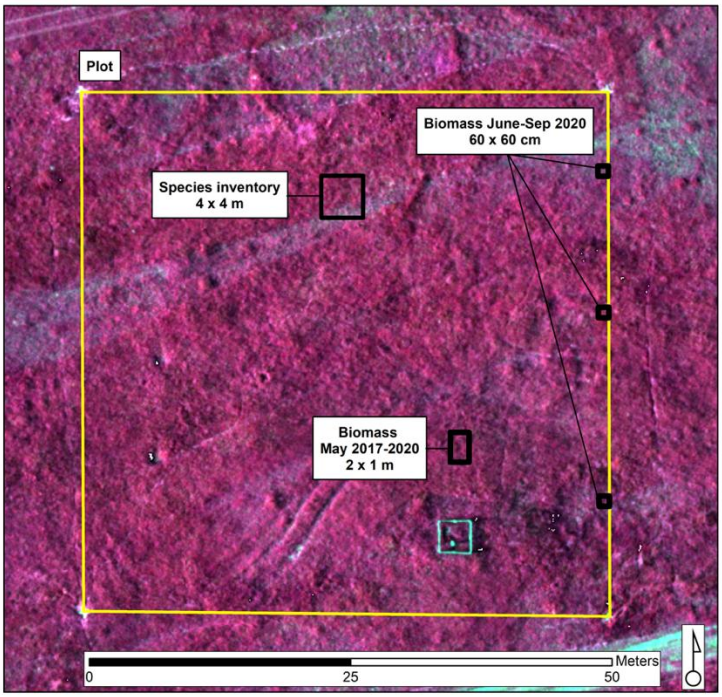
206 standing crop, i.e. in the second half of May from 2017 until 2020, (Figure 2). Aboveground
207 biomass was clipped in a 2 × 1-meter area at 2 cm stubble height and dried at 80°C for 48 hours
208 before weighing. Following the 'peak standing crop' estimation method, dried biomass
209 represented aboveground net primary production (Ruppert and Linstädter, 2014). Species
210 inventories were conducted within a 4 × 4 m area inside each plot. The land use is guaranteed to
211 be constant within the 50 × 50 m plot, and the biomass and plant inventories collected are
212 assumed to be representative of the whole plot (Felipe-Lucia et al., 2020; Fischer et al., 2010; Le
213 Provost et al., 2021). Plots with trees inside or close to their borders were removed from the
214 analysis, resulting in a grand total of 502 samples (Table 1). The soil type, slope and aspect on
215 each plot were extracted from the Biodiversity Exploratories Information System (BExIS)
216 (Ostrowski et al., 2018), and climate information was obtained from the climate station (Hänsel et
217 al., 2021). Average and standard deviation of precipitation and temperature vary considerably
218 across the three exploratories, with Schwäbische Alb having the largest difference in climatic
219 conditions among the three sites (Figure 3). Topographic conditions are also very heterogeneous
220 within and among exploratories, and the three sites cover a wide range of land use intensities.
221 Species richness and biomass productivity are also very heterogeneous, with Hainich site having
222 the highest variability in species richness (Figure 3).

223 Two additional independent field campaigns were conducted to collect biomass data for validation
224 purposes in Schwäbische Alb and Hainich in June and September 2020 (Table 1). During these
225 campaigns, a subset of 40 out of the 150 grassland plots were sampled. The selection of these
226 plots considered the existing range of management intensities and their area covering each
227 exploratory. The biomass was collected from a 0.6 × 0.6 m area within the investigated quadrat
228 (Figure 2) and rescaled to g/m^2.

229

230 Table 1: Number of valid observations per field campaign and exploratory.

	Exploratory			
Campaign	**ALB**	**HAI**	**SCH**	**Total**
May 2017-2020	171	179	152	502
June 2020	8	10	-	18
Sep 2020	11	11	-	22



231

232 Figure 2: Example of a plot and sampling design for species inventories and biomass sampling
233 carried out in May every year, and in June and September 2020. Background image shows false
234 color composite of near infra-red, red, and green bands of a UAV image (spatial resolution ~10
235 cm).

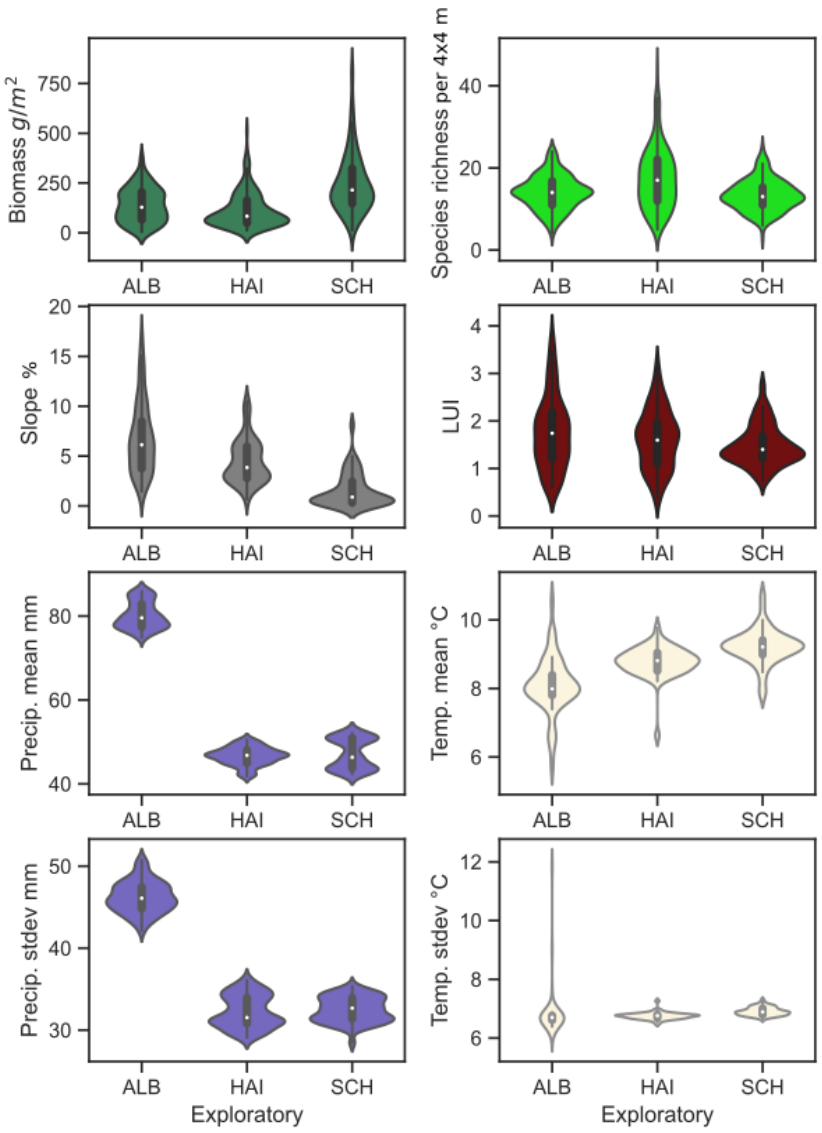


Figure 3: Violin plots showing the differences in biomass productivity, species richness, slope, land-use intensity (LUI) and climatic conditions of the three exploratories between 2017 and 2020.

3.2 Remote sensing data and preprocessing

Sentinel-1 data was obtained from Google Earth Engine (GEE) by computing the median backscatter values of both VV and VH polarizations, and their normalized difference index for May. Results were sampled to 10 m. A description of standardized preprocessing steps of Sentinel-1 can be found in GEE (2018).

Sentinel-2 images were obtained from the Google cloud storage and processed using the FORCE software v 3.6.5 (Frantz, 2019). We downloaded all available Sentinel-2 Level 2A images from eight tiles covering the three exploratory regions in the period 01.01.2017 - 31.12.2020 with less than 50% cloud cover. Analysis-ready data with 10 m spatial resolution was derived by performing atmospheric and topographic corrections using the FORCE Level-2 module (Frantz, 2019). Elevation data was obtained from the Shuttle Radar Topographic Mission (SRTM) 1 Arc-Second (EROS, 2017). From these data, we created a regular synthetic time series with one image every 14 days using the radial basis function interpolation in the Time Series Analysis module of the FORCE processor. We set the maximum temporal distance for the moving average filter to 16 days, the Gaussian kernel with three different standard deviations for (i.e., 8,16,32) and a cut-off value of 0.995 when calculating the weight from the Gaussian distribution. Table 2 shows the synthetic products generated.

Table 2: Sentinel-2 bands information

Band number	Central wavelength (nm)	Band width (nm)	Name
B2	492.4	66	blue
B3	559.8	36	green
B4	664.6	31	red
B5	704.1	15	Red edge 1 (re1)
B6	740.5	15	Red edge 2 (re2)
B7	782.8	20	Red edge 3 (re3)
B8	832.8	106	Near infrared (nir)
B8a	864.7	21	Near infrared b (nirb)
B11	1613.7	91	Short-wave infrared 1 (swir1)
B12	2202.4	175	Short-wave infrared 2 (swir2)

Due to a higher cloud contamination of winter images and their low relevance for vegetation studies, only images from April to the end of October were analyzed.

Ten optical vegetation indices were calculated from the synthetic time series at the time of the field campaigns (Table 3). These indices were extracted from the Index Database (https://www.indexdatabase.de/db/r.php) (Henrich et al., 2012), and are commonly used in the scientific literature for grasslands, and ensure inclusion of all Sentinel-2 bands (Malenovský et al., 2012). The leaf area index (LAI) was computed separately using the Biophysical processor of the Sentinel Application Platform (SNAP) (Weiss et al., 2020) with images acquired as close as possible to the sampling campaigns and downloaded from the Copernicus Open Access Hub.

Table 3: Sentinel-2 derived predictors (Henrich et al., 2012).

Index	Name	Formula	Reference
EVI	Enhanced vegetation index	$2.5 \times \frac{(B8 - B4)}{(B8 + 6 \times B4 - 7.5 * B2) + 1}$	(Huete et al., 2002)
NDVI	Normalized difference vegetation index	$\frac{B8 - B4}{B8 + B4}$	(Tucker, 1979)
MNDVI	Modified NDVI	$\frac{B8a - B12}{B8a + B12}$	(Jurgens, 1997)
GNDVI	Green NDVI	$\frac{B8 - B3}{B8 + B3}$	(Gitelson et al., 1996)
NDII	Normalized difference 819/1600	$\frac{B8 - B11}{B8 + B11}$	(Henrich et al., 2012)
MIRNIR	Normalized difference mir/nir	$\frac{B12 - B8}{B12 + B8}$	(Henrich et al., 2012)
SAVI	Soil adjusted vegetation index	$\frac{B8 - B4}{B8 + B4 + 0.428} \times (1 + 0.428)$	(Huete et al., 2002)
ARVI	Atmospherically resistance vegetation index	$\frac{B8a - B4 - 0.106 \times (B4 - B2)}{B8a + B4 - 0.106 \times (B4 - B2)}$	(Kaufman and Tanre, 1992)
CHLRE	Chlorophyll red-edge	$\left(\frac{B7}{B5}\right)^{-1}$	(Gitelson et al., 2006)
MCARI	Modified chlorophyll absorption reflectance index	$(B5 - B4) - 0.2 \times (B5 - B3) \times \frac{B5}{B4}$	(Daughtry, 2000)
LAI	Leaf Area Index	PROSAIL	(Weiss et al., 2020)
Q	Rao's Q	$Q = \sum \sum d_{ij} \times p_i \times p_j (*)$	(Rocchini et al., 2017)

(*) d_{ij} = pairwise distance between pixels attaining to reflectance values i and j. p_i = relative abundance of pixels attaining to reflect valie i. p_i = relative abundance of pixels attaining to reflectance value j.

The biophysical processor uses eight reflectance bands (B3, B4, B5, B6, B7, B8A, B11, and B12) together with viewing zenith, solar zenith, and relative azimuth angles in the radiative transfer models PROSAIL coupled with a LAI-predicting neural network. Rao's Q (Rocchini et al., 2017) was calculated for the Sentinel-2 synthetic images closest to the field campaigns using a 3 × 3 pixels kernel. Rao's Q accounts for the relative abundance and spectral distances among pixel reflectance values. It is based on the spectral heterogeneity hypothesis, by which areas with a high spectral variability (for the same land cover class) would have a high diversity of plant species. The final list of 178 predictors can be grouped into the 4 following datasets:

- a Sentinel-2 time series of surface reflectance from April to October (16 images/year),
- 10 optical vegetation indices, LAI and Rao's Q for mid-May (time of sampling campaign),
- median VV and VH backscatter from Sentinel-1 and its normalized difference index for May
- a set of topoedaphic variables (soil type, slope, and aspect).

3.3 Algorithms

We compared the performance of RF (Breiman, 2001), a popular machine learning algorithm, with a feed-forward DNN, a deep learning network that organizes decision nodes (neurons) in interconnected layers coupled with a backpropagating feedback mechanism. RF was applied using the scikit learn package (Pedregosa et al., 2011), while DNN was implemented using TensorFlow, an infrastructure layer for differentiable programming, and Keras, its interface for deep learning (TensorFlow Developers, 2021).

3.3.1 Random Forest

RF (Breiman, 2001) is a nonlinear non-parametric regression method. It builds random ensembles of individual decision trees (forest) that systematically tests different subsets of variables and observables to make predictions. The value that receives the most votes from all decision trees

will be selected. RF is also resilient to outliers and can handle different data types, but it can fail to predict extreme values (Xu et al., 2016). We parameterized the algorithm by testing different combinations with the GridSearchCV function of scikit learn. The number of trees was set to 500, and the number of splits to 1/3 of the number of predictors, as suggested by Breiman (2001).

3.3.2 Deep Neural Networks

The feedforward DNN (also known as multi-layer perceptron) was implemented using a sequential model where predictors are passed through a first set of neurons forming the input layer. Each neuron has its own statistical weight. By multiplying the input number with the weight, one gets the output of the neuron, which is then transferred to the next layer. While RF has fewer parameters, the DNN has multiple hyper-parameters (Goodfellow et al., 2016). A global optimal configuration of DNN hyperparameters was achieved in the optimization framework Keras-tuner (O'Malley et al., 2019). The user defines the search space, and the model is subsequently run with different combinations of hyperparameters until it finds the combination that minimizes the loss function over the training data. The DNN applied in our study consisted of 5 layers: a normalization layer, an input layer with 64 neurons, three inner layers with 64 neurons each, and an output layer with a single output neuron (Figure 4). We used a rectified linear unit (ReLU) as activation function, which is a computationally efficient and simple function for nonlinear transformations. As the model optimizer we applied an adaptive moment estimation (Adam), which is an extension of the stochastic gradient descent. Finally, as a loss function (i.e., the quantity that the model attempts to minimize while learning) we chose the mean absolute error. To reduce the probability of DNN overfitting, we applied a regularization penalty of L1 to the hidden layers and an early stopping to the number of epochs (number of times the algorithm works through the training dataset).

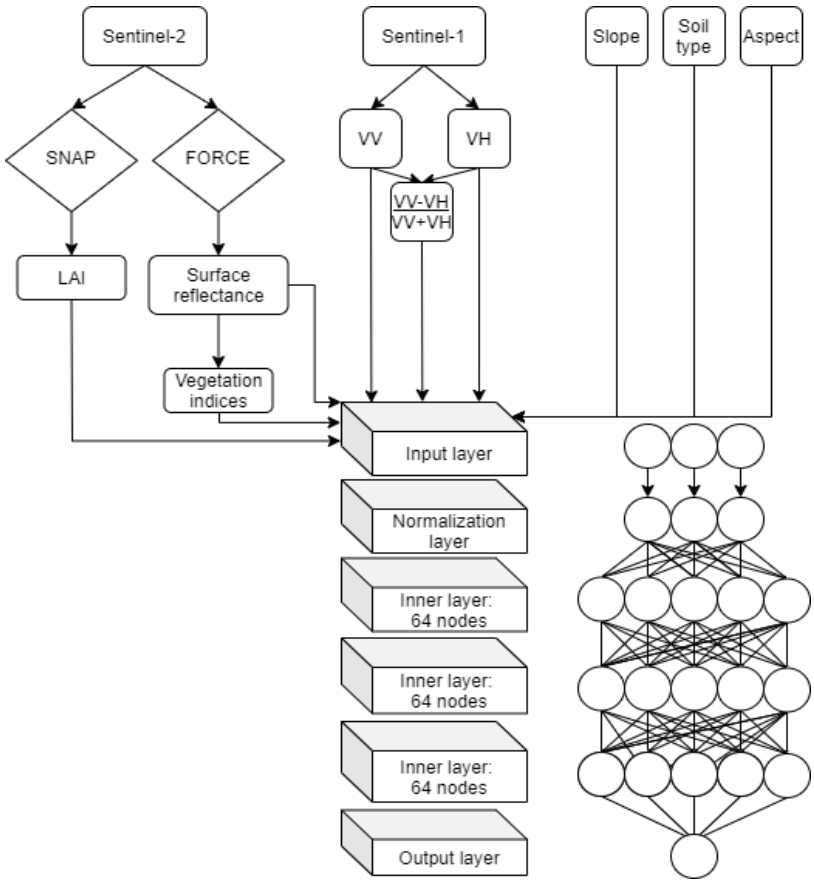


Figure 4. Workflow and architecture of the feed forward neural network (DNN).

3.4 Machine learning regressions

Spatial and temporal dependencies in training and validation data sets might result in an overestimation of models' accuracies (Meyer et al., 2019; Ploton et al., 2020). To account for dependencies and to identify the extent to which each model was capable of generalizing to new data, we performed a series of target-oriented analyses: (i) a group k-fold cross-validation, (ii) a spatial cross-validation across exploratories, (iii) a temporal cross-validation across the phenological cycle, and (iv) an analysis of the area of applicability (Meyer and Pebesma, 2021). The performance of the models is compared through coefficient of determination (r^2), the root mean squared error (RMSE) and the relative root mean squared error (RRMSE), computed in relation to the mean values of the validation dataset. Additionally, we calculated the model-driven

329 systematic component of the RMSE, (sRMSE), and its unsystematic data-driven component
330 (uRMSE) (Willmott, 1981). These components offer additional information to that of RMSE and r^2
331 as they allow a more profound assessment of the models. If sRMSE is higher than the uRMSE,
332 the model is affected by systematic errors and that it will return biased estimations. Conversely,
333 if the uRMSE is higher, the model can be considered optimized (Malenovský et al., 2013).

334 *3.4.1 Predictor importance*

335 Studying predictor importance with leave-one-out approaches can lead to misinterpretations due
336 to inevitable multicollinearity of several predictors (e.g., in the time series data). Thus, predictor
337 importance was evaluated using the SHAP python package, SHapley Additive exPlanations
338 (Lundberg and Lee, 2017). SHAP computes the Shapely additive contribution values from the
339 coalitional game theory. The values are driven by correlations between a predictor and a response
340 variable picked up by the predictive model, but these correlations are not necessarily causal.

341 *3.4.2 K-fold cross-validation*

342 For both algorithms, we performed a group 5-fold cross-validation. Having replicates (i.e.,
343 observations made in the same plot in consecutive years) we guaranteed independence of
344 training and validation observations by ensuring that all replicates were either in the validation
345 dataset or in the training dataset of each fold. The group k-fold cross-validation was iterated 10
346 times to account for stochastic effects. Additionally, we performed a Moran's I test (Cliff, A. D and
347 Ord, J. K, 1981) using the PySAL package (Rey and Anselin, 2007) to confirm the absence of
348 local and global spatial autocorrelation in the biomass and species richness observations.

349 *3.4.3 Spatial and temporal cross-validation across exploratories*

350 We conducted spatial cross-validation tests in which two exploratories were used for training, and
351 the remaining exploratory for validation. Additionally, we tested the biomass models on 40

samples collected independently during June and September 2020 in Hainich and Schwäbische Alb. This was done to test the model predictability during other phenological stages. A temporal cross-validation for species richness was not possible due to differences sampling area between campaigns.

3.4.4 Spatial applicability analysis

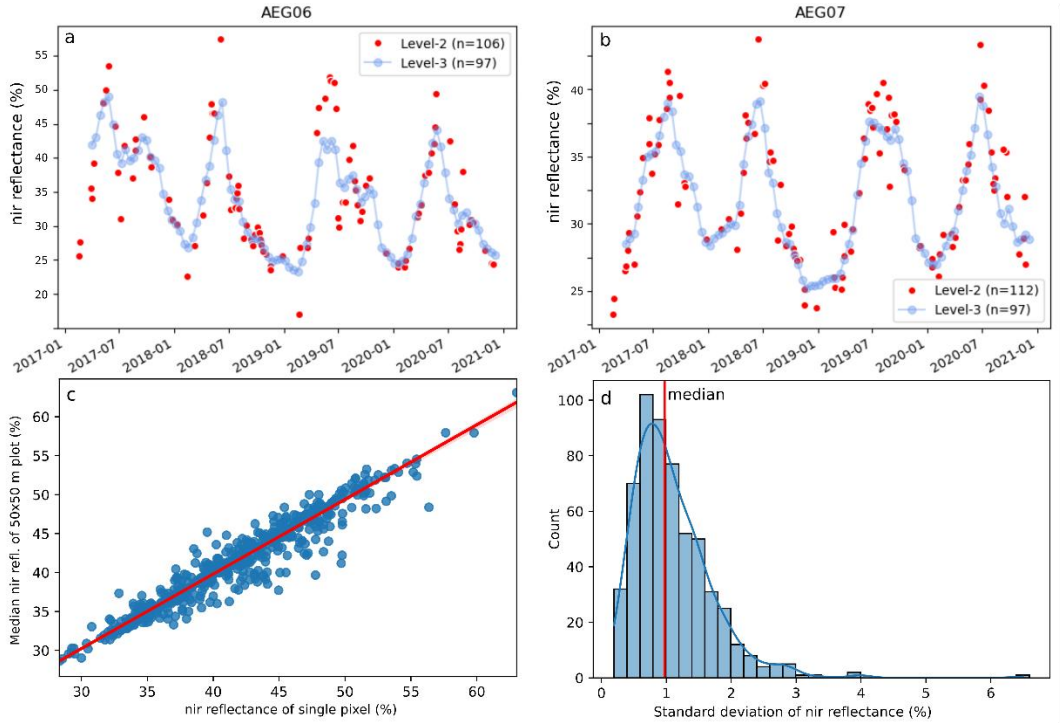
We calculated the area of applicability for the best performing biomass and species richness models using the CAST package (Meyer and Pebesma, 2021). The area of applicability is defined as the area with a distribution of the predictors' value similar to the one used in the model training. Thus, it represents the area where the estimated cross-validation performance holds.

4 Results

4.1 Sentinel-2 image processing and pixel statistics

Figure 5a-b shows the original and interpolated time series of Sentinel-2 nir reflectance (band 8a) for an intensively and an extensively managed plot. FORCE captures well the temporal patterns of each pixel without over smoothing the time series, although the highest reflectance values can occasionally be left out. Figure 5c-d illustrates the variability of nir reflectance in pixels representing the studied plots. The correlogram in panel *c* shows the median pixel values of each 50 × 50 m study plot (used in our predictions) plotted against the single pixel value of the point where plant inventories were conducted. Panel *d* shows the histogram of the standard deviations of the nir values for all the 50 × 50 m plots sampled. Since the nir reflectance is ranging between 28% and 63% and the standard deviation of 93% of the plots is below 2% (i.e., 5.8% of the dynamic range), the within plot spectral variability can be considered low. This evidences that the median reflectance value of 50 × 50 m plots used in our predictive models is sufficiently representative of the plant inventories. Additionally, using the median value prevents the use of

375 poor-quality reflectance values due to cloud/shadow contamination. Figure 1 in the Appendix
376 contains the reflectance statistics for the swir1 and green bands of Sentinel-2.



377

378 Figure 5. Temporal and spatial distributions of Sentinel-2 near infrared (nir) reflectance in the
379 studied plots. The upper panels show the original (level 2) and interpolated (level 3) nir values for
380 an intensive (*a*) and an extensive (*b*) plot across the study period. The lower panels show the nir
381 reflectance statistics for all studied plots. Panel *c* shows the correlation between median nir values
382 of the 50 × 50 m plots versus the nir reflectance of the single pixel where plant inventories were
383 conducted. Panel *d* shows the histogram of the nir standard deviations (n = 572) and their median
384 value, illustrating the low spectral variability of the 50 × 50 plots in the nir. Similar analysis can be
385 found in Figure 1 of the Appendix for the green and swir1 bands of Sentinel-2.

386 **4.2 K-fold cross-validations**

387 *4.2.1 K-fold cross-validation: Biomass*

The cross-validation accuracy assessments of the RF and DNN models for different combinations of predictors are shown in Table 4 and Appendix Figure 2. The highest r^2 and lowest RRMSE were achieved by the DNN model based on the Sentinel-2 surface reflectance bands from May (time of biomass acquisition). The systematic and unsystematic RMSEs tended to be more even for DNN models. The sRMSE component was found to be proportionally higher than the uRMSE for all RF models, indicating their worse performance due to suboptimal design or training.

Table 4. Accuracy metrics for different combination of predictors used in DNN and RF biomass models: coefficient of determination (r^2), root mean squared error relative to the mean (RRMSE), and its systematic (sRMSE) and unsystematic components (uRMSE). Each model was iterated 10 times and the resulting variation is given as the standard deviation of each metric. Predictors: S2-sr are the surface reflectance values for the time of field data acquisition (May), S2-deriv. are the Sentinel-2 derivatives: 10 vegetation indices and LAI (Table 3). S1 are the predictors derived from Sentinel-1 and topoedaphic variables, (TE), contain information on slope, aspect, and soil type. The best result for each metric and combination of predictors is highlighted.

Biomass	**DNN**								**RF**							
Predictors	**r^2**		**RRMSE**		**sRMSE (g/m^2)**		**uRMSE (g/m^2)**		**r^2**		**RRMSE**		**sRMSE (g/m^2)**		**uRMSE (g/m^2)**	
S2-sr	0.45	±0.01	0.47	±0.00	53	±0.4	52	±0.5	0.37	±0.01	0.49	±0.00	61	±0.2	48	±0.1
S2-sr, S1	0.44	±0.01	0.48	±0.05	54	±1.2	52	±1.1	0.39	±0.00	0.49	±0.00	61	±0.2	40	±0.1
S1, TE	0.11	±0.03	0.64	±0.01	74	±1.0	68	±2.4	0.08	±0.00	0.62	±0.00	84	±0.2	49	±0.5
S2-deriv.	0.39	±0.01	0.50	±0.08	60	±0.7	51	±0.7	0.39	±0.00	0.49	±0.00	59	±0.2	50	±0.1
S2-sr, TE.	0.34	±0.02	0.52	±0.06	56	±0.7	59	±1.7	0.38	±0.00	0.49	±0.03	61	±0.1	48	±0.2
Full set	0.34	±0.02	0.53	±0.01	55	±1.0	64	±2.4	0.42	±0.00	0.47	±0.00	58	±0.2	47	±0.2

While both algorithms achieved comparable r^2 values for the same set of predictors, RF tended to slightly overestimate predictions of lower biomass values (Figure 6). Additionally, both types of models underestimated a few plots of abnormally high amounts of biomass.

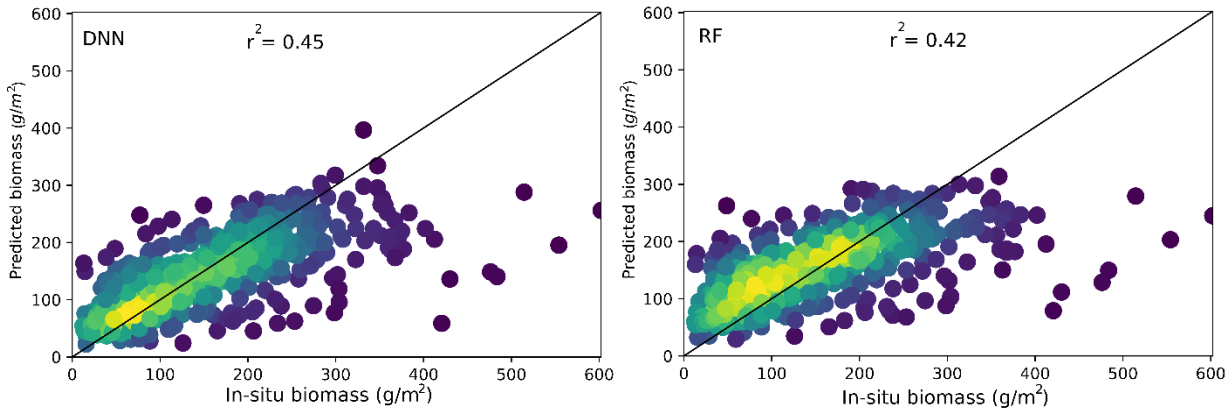


Figure 6. One-to-one density plots of predicted vs. in-situ measured values for the best performing biomass models with DNN (S2_sr) and RF (Full set). Scatterplots show the accumulated results of 5 folds.

4.2.2 K-fold cross-validation: Species richness

The species richness models were tested using the same groups of predictors that were used for the biomass models, with the addition of the time series of surface reflectance values. Although performance of RF improved with the addition of more predictors, it produced estimates with lower accuracies than DNN (Table 5, Appendix Figure 3). RF achieved the highest r^2 and lowest RRMSE when using the time series of Sentinel-2 and the topoedaphic variables. However, its predictions suffered from systematic overestimations of low values and underestimations of high values (Figure 7). This is supported by sRMSE being 1/3 higher than uRMSE. Contrary to this, sRMSE was slightly higher than half of the uRMSE for the best performing DNN model (Table 5).

Table 5. Accuracy metrics for different combination of predictors used in DNN and RF models of species richness: coefficient of determination (r^2), root mean squared error relative to the mean (RRMSE), and its systematic (sRMSE) and unsystematic components (uRMSE). Each model was iterated 10 times, and the resulting variation is expressed as standard deviation of each accuracy metric. Predictors: S2-sr is the Sentinel-2 surface reflectance for the time of field data acquisition

(May). S2-deriv. are the Sentinel-2 derivatives: 10 vegetation indices, LAI and Rao's Q (Table 3). S2_ts is the time series of Sentinel-2 surface reflectance from April to October, and S1 are the Sentinel-1 derived predictors. Topoedaphic variables (TE) contain information on slope, aspect and soil type. The best result for each metric and combination of predictors are highlighted.

Spp. rich.	**DNN**								**RF**							
Predictors	**r^2**		**rRMSE**		**sRMSE**		**uRMSE**		**r^2**		**rRMSE**		**sRMSE**		**uRMSE**	
S2-ts	0.42	±0.01	0.26	±0.00	4.2	±0.2	7.5	±0.2	0.38	±0.02	0.27	±0.00	7.1	±0.01	4.4	±0.02
S2-ts, S1	0.41	±0.01	0.28	±0.01	4.5	±0.1	7.3	±0.2	0.40	±0.02	0.27	±0.07	7.1	±0.02	4.8	±0.02
S2-ts, TE.	0.39	±0.01	0.29	±0.05	4.5	±0.2	7.8	±0.1	0.40	±0.01	0.26	±0.00	6.6	±0.01	4.6	±0.01
S1, S2-sr, S2-deriv, TE	0.28	±0.02	0.31	±0.01	6.1	±0.1	7.2	±0.2	0.34	±0.01	0.29	±0.06	6.9	±0.01	5.2	±0.01
S2-ts, S1, TE	0.41	±0.08	0.28	±0.00	4.3	±0.2	7.5	±0.2	0.40	±0.01	0.26	±0.00	6.6	±0.02	4.5	±0.02
Full set	0.41	±0.01	0.28	±0.00	4.4	±0.1	7.5	±0.2	0.41	±0.02	0.26	±0.05	6.3	±0.02	4.6	±0.02

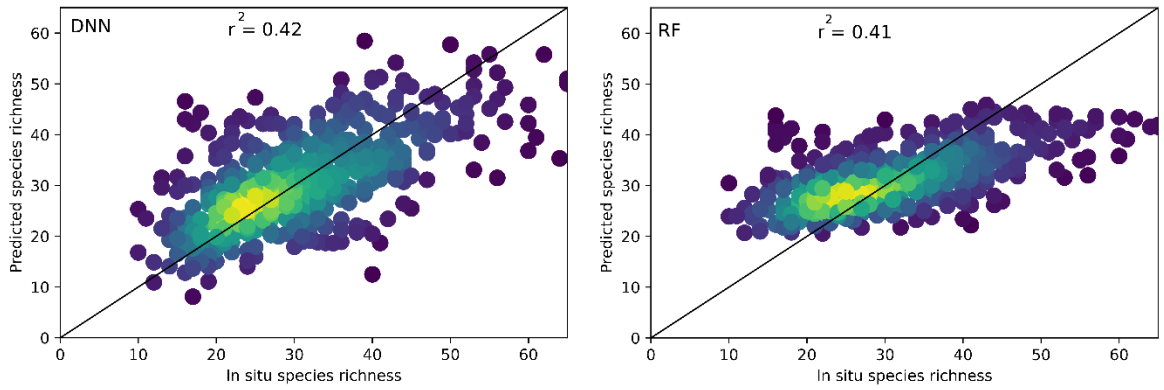


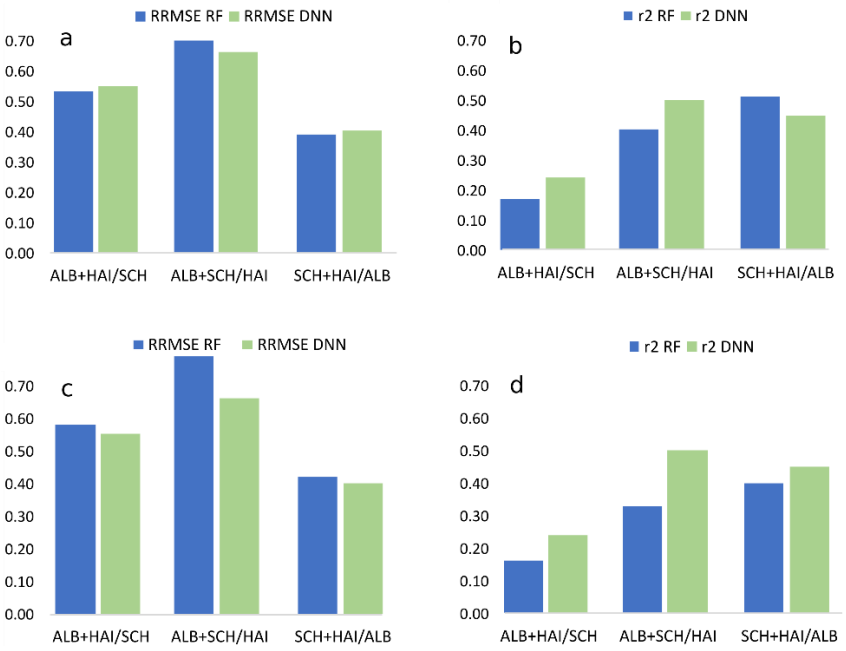
Figure 7. One-to-one density plots of predicted vs. in-situ measured values for the best performing DNN (S2_ts) and RF (Full set) species richness models. Scatterplots show the accumulated results of 5 folds.

4.3 Spatial and temporal cross-validations

4.3.1 Spatial and temporal cross-validation: Biomass

To test the transferability of the models to independent input datasets, we performed a spatial cross-validation where we used two exploratories for training, and the third for validation (Figure

435 8). RF produced overall higher RRMSE and lower or equal r^2 values, even when using the same
436 set of predictors as DNN (Figure 8 c, d). The RF and DNN models were additionally validated
437 against 40 independent data points collected in June and September in Hainich and Schwäbische
438 Alb exploratories. Accuracies achieved for both models were comparable to previous tests (r^2 =
439 0.46), except for five plots with large amounts of dry biomass measured during June (Figure 9).



440

441 Figure 8: RRMSE (a, c) and r^2 (b, d) of the RF and DNN biomass models respectively for the three
442 possible spatial cross-validations. Two exploratories are used for training, and the third one for
443 validation. The total number of samples was 502 (ALB: n = 171, HAI: n = 179, SCH: n = 152).
444 Panes 'a' and 'b' display results for the best predictors for each model (see Table 4), whereas
445 pane 'c' and 'd' show results for the Sentinel-2 surface reflectance values acquired in May (the
446 best predictors for the DNN model). All correlations between predicted and measured data were
447 found to be statistically significant ($p < 0.001$).

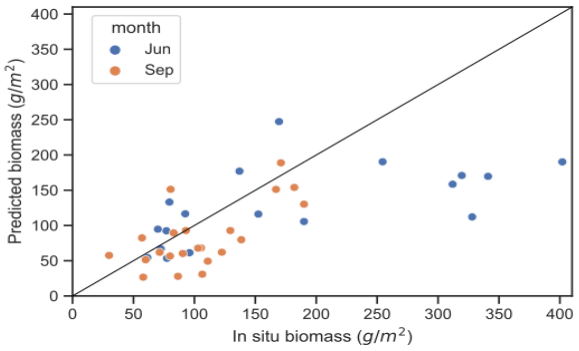


Figure 9: One-to-one relationship between in-situ measured and DNN predicted amount of biomass for independent samples collected during campaigns in June and September (r^2 = 0.46).

4.3.2 Spatial cross validation: Species richness

Regardless of achieved RRMSE results, both types of algorithms returned quite low r^2 values for predicting the species richness with different combinations of training and validation regions, which indicates a lack of generalization capacity of both models, (Figure 10).

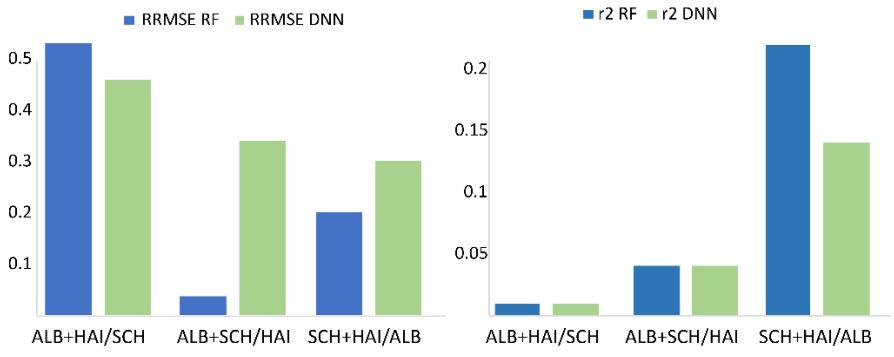


Figure 10: RRMSE (left) and r^2 (right) of RF and DNN species richness models for the three possible spatial cross-validation combinations. Two exploratories are used for training and the third one for validation (i.e., training+training/validation). The time series of Sentinel-2 surface reflectance values were used as predictors in all cases.

4.4 Predictor importance

4.4.1 Predictor importance: Biomass

In both, RF and DNN models, chlorophyll indices MCARI (based on red-edge1, green and red) and CHLRE (based on red-edge1 and 3), and red-edge1 yielded the largest contributions, followed by nir and red-edge2, although with high variability across folds (Figure 11). The frequently used NDVI, EVI and LAI contributed far less to the model. VH showed a higher contribution than VV or their normalized difference index, but still rather low. When analyzing the importance of the Sentinel-2 surface reflectance data for May, the bands with the highest score were red-edge1 and 3, and the lowest scores were for swir2, green, and red bands.

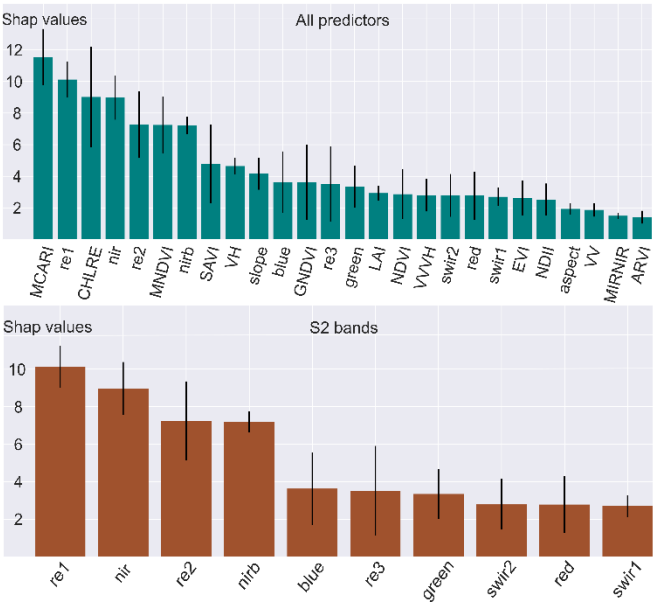


Figure 11. Shap values of the biomass models. Upper chart shows the shap values when all predictors are considered. Lower chart shows the shap values when only the Sentinel-2 bands are considered (the best model). Error bars show standard deviation across folds.

4.4.2 Predictor importance: Species richness

The importance of all Sentinel-2 bands between April and October of each year was analyzed. Figure 12 shows the shap values aggregated temporarily (14 days' time steps) and per band. Interestingly, red-edge3 had the lowest value, while the other two red-edge bands had the highest, yet with a high variability across folds. The importance of the different predictors across time steps was quite consistent, except two peaks appearing at the beginning of May and end of October.

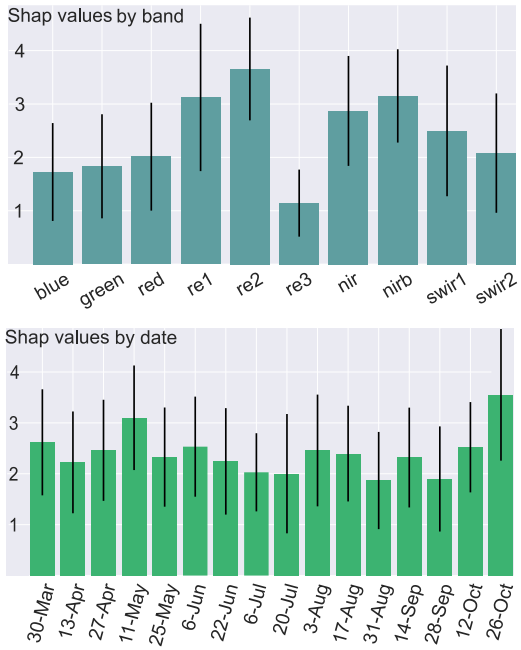


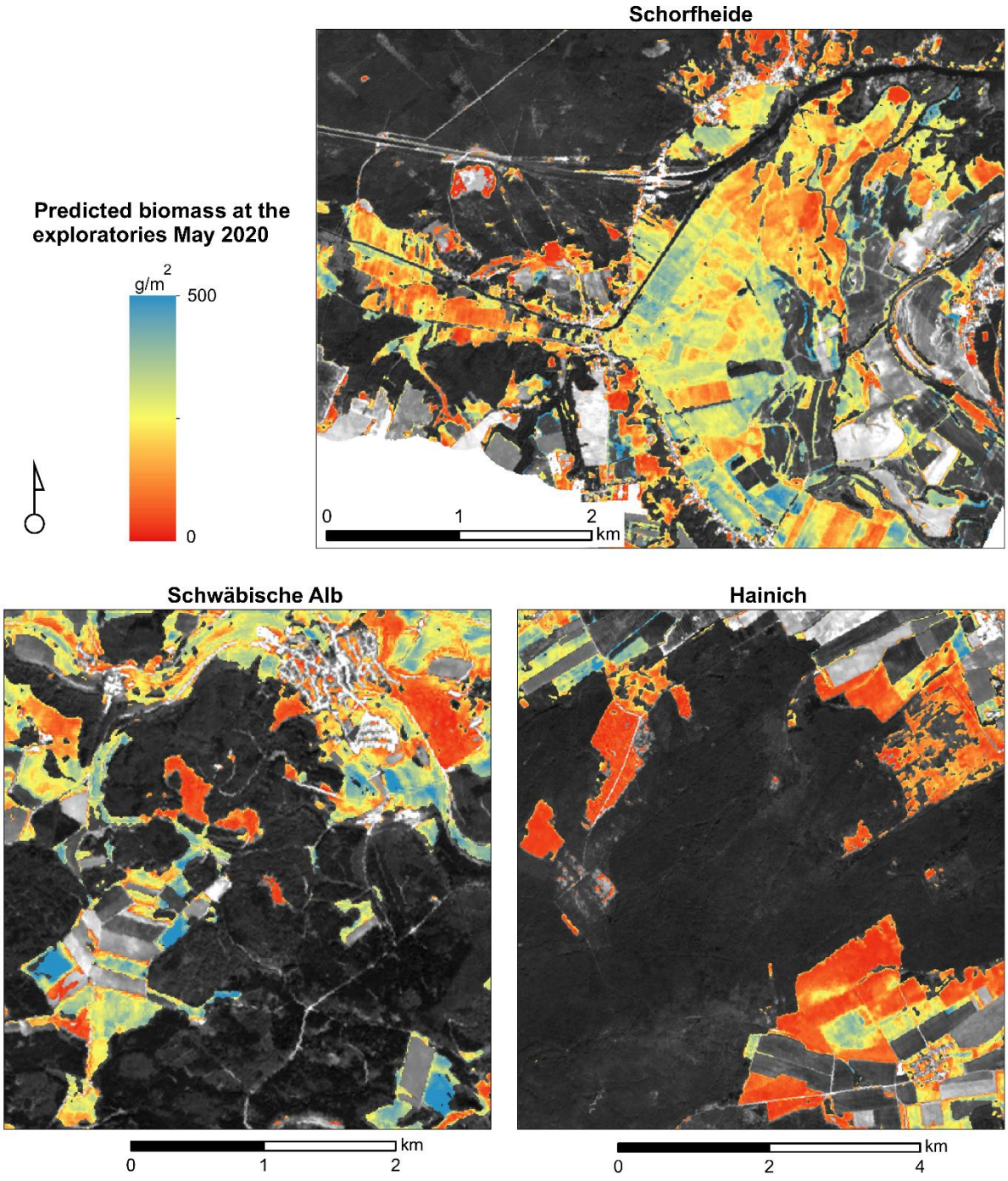
Figure 12. Shap values for the species richness model. Values are aggregated by band (upper chart) and by 14-days' time step (lower chart). Error bars show standard deviation across folds.

4.5 Spatial applicability analysis

The best models for biomass estimation (i.e., DNN based on Sentinel-2 surface reflectance for May) and species richness predictions (i.e., DNN based on Sentinel-2 time series of surface reflectance) were used to study their spatial applicability. The results indicated that the cross-validation holds for the vast majority of areas classified as grasslands by the Digital Model

Landscape (BKG, 2020) (see Appendix Figures 4 and 5). Figure 13 and Figure 14 show subsets of the biomass and species richness predictions where the models were applicable. Predictions were very similar across the four investigated years and biomass and species richness estimates were spatially negatively correlated (Pearson correlation coefficient of -0.35).

For different levels of LUI at the Alb exploratory, we also produced temporal estimations of standing biomass throughout the study period (Figure 15). Plots of low and medium LUI (LUI $\leq$ 2) displayed overall higher levels of biomass across each year, whereas plots with LUI > 2 displayed lower mean biomass levels per year protruded by more frequent local peaks. The plots responsible for peak values across the season tend to be the same, suggesting that these predictions are product of management practices, even though some are located outside of common seasonal peaks (May/June).

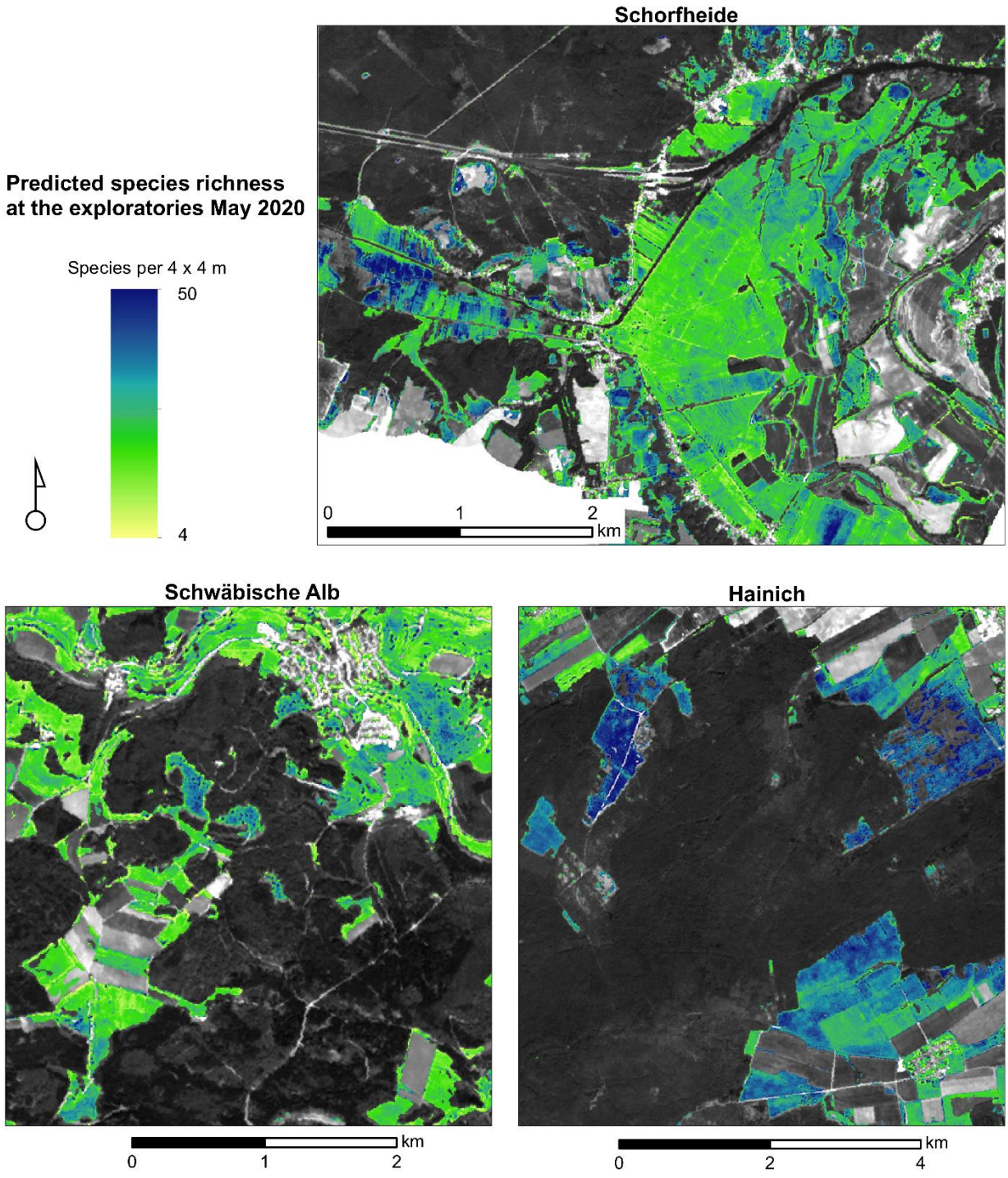


499

500 Figure 13: Predictions of biomass for subsets of the three exploratories in May 2020. Backdrop:

501 greyscale of Sentinel-2 blue band reflectance.

502



503

504 Figure 14: Predictions of species richness per 4 × 4 m ground spatial sampling units for subsets

505 of the three exploratories for May 2020. Backdrop: greyscale of Sentinel-2 blue band reflectance.

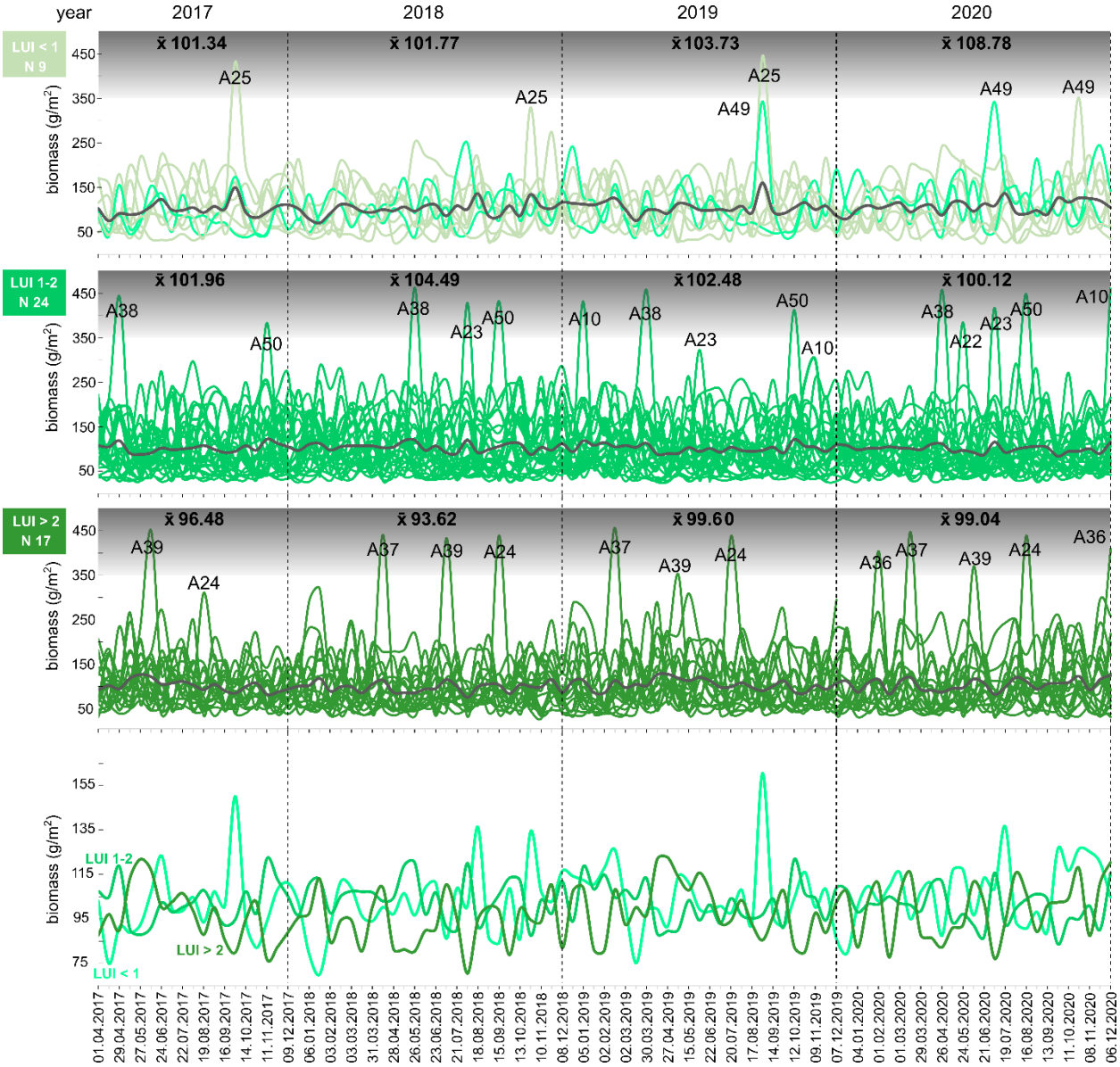


Figure 15. Temporal predictions of standing biomass for all plots at the Schwäbische Alb exploratory, divided into three LUI classes: LUI: LUI<1, 1<LUI≤2, and LUI>2). The average amounts of biomass per year are given by x̄ (black lines). The last chart shows the average values for the three LUI classes. The shaded area indicates the upper 5% of biomass values recorded in the field. Plot numbers with abnormally high amounts of predicted biomass are indicated in the chart.

5 Discussion

The following three limitations in remote-sensing based modelling of grassland characteristics are common across the literature: an insufficient (too narrow) variability of the reference data, spatio-temporal dependencies in the reference data, (Meyer et al., 2019; Ploton et al., 2020) and an inconsistency of remote sensing predictors across studies (Verrelst et al., 2019). This study addresses all these issues explicitly, producing thoroughly validated and robust models with unbiased accuracy estimates. Using field data from 150 plots in the exploratories (Bolliger et al., 2021b, 2021a) according to a systematic and unified experimental setup, we have ensured that our prediction models cover a wide array of topoedaphic, climatic, and management conditions found in Germany (Fischer et al., 2010) and are free from spatiotemporal dependencies. We achieved accuracies that are comparable to other studies (Aneece et al., 2017; Bae et al., 2019; Fauvel et al., 2020; Imran et al., 2021; Lopes et al., 2017; Peng et al., 2019; Wang et al., 2018), while limiting the number of model predictors and correlation amongst them. Considering the fact that maps of, for instance, soil type are not generally available at sufficient accuracy, restricting the predictors to only Sentinel-2 surface reflectance complies with the principle of parsimony, and facilitates more operational production of comparable spatial estimates.

5.1 Biomass

Band-based spectral indices such as EVI and NDVI are often used to calculate aboveground biomass and crop yield (Edirisinghe et al., 2012; Grüner et al., 2020; Guerini Filho et al., 2020; Zhou et al., 2014), but in our analysis they were outperformed by the chlorophyll based indices (MCARI and CHLRE) (Table 4). CHLRE has been found to have a strong relationship with primary productivity and be sensitive to variations of canopy chlorophyll content (Muramatsu, 2019). Despite both algorithms performed well with the band ratios, their selection and computation require expert knowledge and heuristic decisions. This can potentially introduce subjective

538 uncertainty, hindering the interpretation of predictor's importance analysis. The best results were
539 achieved with the DNN model using the 10 surface reflectance bands of Sentinel-2. Although the
540 best RF model achieved similar accuracies and error distribution (Figure 6), it required a larger
541 number of predictors and produced estimates with a high sRMSE due to systematically
542 overestimated low biomass amounts (Table 4 & Figure 6).

543 Sentinel-1 based predictors performed poorly, even when combined with topographic information
544 that was expected to compensate for spatial geometric distortions (Table 4). It has been argued
545 that the integration of optical and SAR datasets improves prediction capacities through
546 complementing each other, and reducing negative influences of soil background and weather
547 conditions. Naidoo et al., (2019) reported that a combination of Sentinel-1 and -2 information
548 yielded better predictions of biomass in wetlands. In their case, however, incorporation of SAR
549 data might have provided information capturing differences in wetlands' water dynamics (Muro et
550 al., 2020, 2019) rather than changes in biomass cumulation. Wang et al., (2019) also reported
551 successful predictions of above-ground biomass in pastures from Sentinel-1 data, but their study
552 covered only a limited range of environmental conditions (i.e., flat areas with a single soil type).
553 In our study, the fusion of optical and SAR data did not represent a major improvement with
554 respect to the use of optical data.

555 DNN achieved better results in the spatial cross-validation due to better generalization capabilities
556 (Figure 8) (Kattenborn et al., 2021; Yuan et al., 2020). For both algorithms, r^2 coefficients were
557 especially low when Schorfheide was used as the validation region. Environmental conditions and
558 land management in Schorfheide are quite different from the other two regions, specifically, local
559 soils have a high organic content (Busch et al., 2018), grass growth is occasionally boosted by
560 mulching, and pastures are rolled over rather than mown. Since the sensor cannot detect the
561 biomass tucked under the dense upper layers, all predictive models underestimated seven
562 observations with abnormally high biomass (Figure 6). The same phenomenon is also observed

563 during the temporal validation (Figure 9), where predictions in five plots with exceptional large
564 amounts of accumulated senescent biomass (300-400 g/m^2) were underestimated. Note that the
565 upper limit of the biomass collected at the exploratories every year is around 350 g/m^2 (Figure 3).
566 Apart from these underestimations, the temporal validation suggests that the models are
567 applicable across phenophases (r^2 = 0.46, Figure 9).

568 Biomass predictions across the season at Schwäbische Alb (Figure 15) revealed that low LUI
569 plots had, in general, higher amounts of biomass within each year, and especially during
570 September and October. High LUI plots displayed less biomass per year but more frequent
571 biomass peaks. The occurrence of such biomass peaks is consistent across plots and appeared
572 around the same seasonal phenophases of investigated years. This suggests that they are not
573 prediction errors but rather results of management practices. Predictions of biomass amounts
574 across the season have important implications for the ecology of studied grassland ecosystems,
575 as they show the availability of forage and habitat for other organisms, and their dependencies
576 on land management practices.

577 **5.2 Species richness**

578 Our achieved r^2 coefficients (Figure 7) are comparable to other similar studies that used
579 multispectral imagery to predict either species richness or other biodiversity indicators, such as
580 Simpson or Shannon indices (e.g., r^2 = 0.45 in Fauvel et al. (2020) and r^2 = 0.2 in Lopes et al.
581 (2017)). As a reference point, other studies using imaging spectroscopy (i.e. hyperspectral data),
582 achieved r^2 coefficients between 0.4-0.5 (Aneece et al., 2017; Imran et al., 2021; Peng et al.,
583 2019; Wang et al., 2018).

584 DNN and RF showed similar RMSE and r^2 coefficients. However, similarly to the biomass case,
585 RF sRMSE was higher, predictive models required more inputs (Table 5), and they were prone
586 to underestimate high values and overestimate low values (Figure 7 and Appendix Figure 3). DNN

587 predictions showed more stability as the errors were evenly distributed throughout their dynamic
588 range of values. Underestimations at species rich areas and overestimations at species poor
589 areas renders impractical the application of machine learning models for biodiversity conservation
590 and land-sharing/land-sparing strategies. Thus, it is essential to paid attention to the distribution
591 of the errors besides reporting on r^2 and RMSE.

592 Besides the information provided by the predictions, interpretability analyses allow us to find the
593 rationale behind the predictions and provide meaningful connections to ecological theories and
594 mechanisms. For instance, Wolanin et al., (2020) inferred what combinations of climatic patterns
595 were associated with higher yield, and Campos-Taberner et al., (2020) found that imagery from
596 August was most relevant to identify different types of croplands using multitemporal satellite
597 imagery. Our temporal analysis of the shap values showed two peaks in predictor importance
598 (Figure 12); May (time of field data acquisition) and October. The importance of other of dates
599 varied across folds. This suggests that the model learns from patterns occurring at specific
600 phenophases, such as flowering and senescence, or land use practices during in May and
601 October. It is has been shown that land use intensity in the exploratories strongly influences
602 biodiversity (Felipe-Lucia et al., 2020; Socher et al., 2013). Nevertheless, more research is
603 required to investigate to what extent management practices during these two particular months
604 could have a more relevant effect in biodiversity.

605 The spectral variation hypothesis has been suggested as a suitable biodiversity indicator since it
606 does not require training data. It has been used in several studies (Rocchini et al., 2021, 2019,
607 2017) but recently also questioned by other researches (Fassnacht et al., 2022). Results of this
608 study and our previous research in forests (Hoffmann et al., 2022) do not support validity of the
609 spectral variation hypothesis. The inclusion of spectral Rao's Q as predictor did not improve the
610 species richness models (Table 5). Additionally, we found no linear correlation ($r^2 < 0.01$) between
611 spectral Rao'S Q and species richness nor with of Rao's Q species diversity index (Rao, 1982).

The low performance of spectral Rao's Q as predictor of species richness might be due to spatial scale issues (Khare et al., 2019), especially taking into account the relative spectral homogeneity of the studied plots across the Sentinel-2 bands (Figure 5d and Appendix Figure 1). Higher resolution sensors might yield better results (Pacheco-Labrador et al., 2022). Nonetheless, our results support the call of Fassnacht et al., (2022) for further increasing the research examining the key drivers of the spectral variation hypothesis.

The transferability of species richness models trained with data from two exploratories and tested on the third one was limited (spatial cross-validation, Figure 10). This is product of a classical problem of domain adaptation in machine learning, where the feature space of the test and training data differ significantly. One possible solution was presented by Meyer and Pebesma (2021), who used the feature space of the training data and the weights assigned by the model to each predictor to estimate the area where the accuracies of the model would hold, a.k.a., the area of applicability. Our results of this approach revealed that the species richness model is valid for a vast majority of grasslands around the exploratories (Appendix Figure 5). This highlights the importance of models being trained with heterogenous datasets that cover a wide range of environmental and management conditions. Finally, species richness predictions should consider the mismatch between remote sensing image pixel size and field sampling area (Ferner et al., 2021). One possibility is rescaling the number of species from 4 × 4 m quadrats to the 10 m pixel-size via a non-linear transformation using species-area models. This may introduce extra uncertainties without providing additional relevant information. However, if species-area relationships per community are known, the output models can be used to calculate the effective number of species per management unit. By keeping the scale constant (g/m^2 or species per 4 × 4 m), as in our case, the observation scale, modelling scale, and operational scale remain consistent (Wu and Li, 2009). Although the field data was not collected with remote sensing

636 applications in mind (as it often happens in ecology experiments), the validation results indicate
637 the suitability of our approach.

638 **5.3 Potentials and limitations**

639 Plant species diversity of managed grasslands varies in both space and time, as it depends on i)
640 relatively stable abiotic environmental conditions (such as climate, topography and soil texture),
641 ii) more variable abiotic conditions (such as weather and precipitation patterns) and iii) land-use
642 activities. Our dataset comprised spatial gradients of all these environmental conditions, similarly
643 to previous gradient studies (Ahlborn et al., 2020; Ferner et al., 2018; Guuroh et al., 2018; Lange
644 et al., 2022; Schwieder et al., 2022). As with other datasets used in remote sensing studies (Bae
645 et al., 2019; Hoffmann et al., 2022), our field information was acquired without considering remote
646 sensing requirements, specifically, not matching field and pixel sampling sizes. The results in
647 Figure 5c-d and Appendix Figure 1 show that the median reflectance of the 25 pixels of Sentinel-
648 2 covering each plot is spectrally representative of the single pixel containing the 4 × 4 m inventory
649 plot. Calibrating and validating the machine learning models with the plot median reflectance
650 reduces the risk of including in our analysis poorly radiometrically, atmospherically or
651 geographically preprocessed pixels.

652 Numerous comparative studies report on the accuracies achieved by a broad variety of machine
653 and deep learning models (RF, support vector machines, Gaussian processes, k-nearest
654 neighbors) (Campos-Taberner et al., 2020; Fauvel et al., 2020; Osco et al., 2020; Wolanin et al.,
655 2020). However, the vast majority of them report only on r^2 coefficients and RMSE. Calculating
656 the RMSE systematic and unsystematic components provides a better comparison between
657 models by determining whether the error stems from various uncertainties in the data (e.g. time-
658 lag between satellite and field observations, fast changing atmospheric conditions, poor pixel

interpolation, instrumental errors, sampled biomass and species not being representative in highly heterogenous plots) or from the model itself (e.g. model overtraining).

The use of spatial cross-validation in cases where data is not spatially clustered has been questioned by Wadoux et al., (2021). Despite the fact that neither species richness nor biomass were spatially autocorrelated (Appendix Table 1), we tested our prediction models in target-oriented schemes, where one whole region (in our case, an exploratory) was left out. The obtained low accuracies are result of a poor domain adaptation between exploratories, and show the importance of calibrating predictive models with data representing in full the existing spatial heterogeneity. Although species pools were similar in size across regions, strong regional differences in plant diversity were found at the plot level (Socher et al., 2013), which resulted in very different types of spectral-temporal profiles for each exploratory. DNN, as well as RF, must learn from the spectral-temporal patterns of the three exploratories to make reliable predictions. Thus, our spatial cross-validation should not be seen as an accuracy measure of our models' predictions, but rather as a stress-test of the generalization capacity of both models. The final accuracies are given by the 5-fold cross validation, where models are calibrated for the conditions of the three exploratories.

Following up studies focused on modelling the provision of ecosystem services with land-use practices under varying environmental conditions (Busch et al., 2018; Le Provost et al., 2021), we demonstrated the benefits to model their biodiversity and productivity across both space and time to understand their interactions at landscape level. For instance, Figure 13 and Figure 14 show how biomass and species richness vary across the landscape, and Figure 15 reveals differences in standing biomass for different LUI within several growing seasons. Remote sensing based estimations of standing biomass within the growing season might yield better estimations of aboveground net primary production than the 'peak standing crop' estimation method (Ruppert and Linstädter, 2014), and subsequently, contribute further to the better causal understanding of

684 the BES/BEF relationships. Although heavily fertilized plots produce higher amounts of total
685 biomass per season, they also display lower yearly mean amounts of standing biomass (Figure
686 15), which has consequences in food, dispersal and habitat availability for multiple organisms
687 such as insects. According to the outcomes of our study, the permanently monitored grasslands
688 of the exploratories could be used as training sites for operational mapping of biomass and
689 species richness across gradients, sites, and phenological cycles in Central Europe. Cross-region
690 biomass monitoring could serve as a quantitative alert system to match forage demand with its
691 availability, especially in cases of a high climatic variability and extreme events like droughts
692 (Grigera et al., 2007). Similarly, regular spatial information on biodiversity would allow, for
693 instance, to determining spill-over rates of dispersing organisms from nearby rich habitats that is
694 required to maintain species populations at regional scale (Le Provost et al., 2021). Other studies
695 have been successful at modelling different aspects of functional diversity (Pacheco-Labrador et
696 al., 2022; Rossi et al., 2020), rather than taxonomic diversity, since it provides information on the
697 functional value of each community. Combining information at landscape scale on biomass
698 productivity, and taxonomic and functional diversity can be used to propose more insightful
699 mechanisms to manage ecosystems (Moore, 2013).

700 **6. Conclusion**

701 In this study, we applied deep neural networks and random forest algorithms for estimations of
702 above-ground plant biomass and plant species richness in German grasslands spread across a
703 broad range of management and climatic gradients. Neural networks required fewer predictors,
704 had a lower prediction bias and systematic errors, and had a more robust transferability to new
705 data than random forest. Successful biomass predictions carried out across various regions and
706 several phenological cycles pave the way for a near-real time operational grassland productivity
707 monitoring at regional and even national scales. Prediction models of species richness benefited
708 from the use of multispectral time series, but showed good performances only when trained with

observations of all three exploratories. Despite of this, the models produced spatially explicit datasets of species richness that revealed local hotspots of biodiversity, and can be used to evaluate the temporal influence of regional processes in ecological models. Such models can ultimately guide grassland management towards balancing trade-offs between forage provision and biodiversity at landscape scale. This study highlights the importance of rigorous validations of remote sensing driven prediction models with respect to their actual robustness and spatiotemporal transferability, minimizing the consequent risk of grassland mismanagement.

Acknowledgments

We thank the managers of the three exploratories Kirsten Reichel-Jung, Iris Steitz, and Sandra Weithmann, Juliane Vogt, 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, 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 of Schwäbische Alb, and the UNESCO Biosphere Reserve Schorfheide-Chorin as well as all landowners for the excellent collaboration. We thank as well Ralph Bolliger for providing and maintaining the biomass and species inventory datasets. We are also very grateful for the recommendations and insights of the anonymous reviewers.

Data availability statement:

All data necessary to replicate the analyses is available in the repository of the Biodiversity Exploratories BExIS (https://www.bexis.uni-jena.de/). Accession numbers: biomass: 27426, 26151, 24166, 23486. Sentinel-2 time series: 30969. Species inventories: 27386. Species

richness maps: 31214. Python scripts are available from Github https://github.com/Havi-muro/SeBAS_project.git

Funding

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.

Bibliography

Ahlborn, J., von Wehrden, H., Lang, B., Römermann, C., Oyunbileg, M., Oyuntsetseg, B., Wesche, K., 2020. Climate – grazing interactions in Mongolian rangelands: Effects of grazing change along a large-scale environmental gradient. J. Arid Environ. 173, 104043. https://doi.org/10.1016/j.jaridenv.2019.104043

Aklilu Tesfaye, A., Gessesse Awoke, B., 2021. Evaluation of the saturation property of vegetation indices derived from sentinel-2 in mixed crop-forest ecosystem. Spat. Inf. Res. 29, 109–121. https://doi.org/10.1007/s41324-020-00339-5

Ali, A.M., Darvishzadeh, R., Skidmore, A., Gara, T.W., Heurich, M., 2021. Machine learning methods' performance in radiative transfer model inversion to retrieve plant traits from Sentinel-2 data of a mixed mountain forest. Int. J. Digit. Earth 14, 106–120. https://doi.org/10.1080/17538947.2020.1794064

Ali, I., Cawkwell, F., Dwyer, E., Green, S., 2017. Modeling Managed Grassland Biomass Estimation by Using Multitemporal Remote Sensing Data—A Machine Learning Approach. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 10, 3254–3264. https://doi.org/10.1109/JSTARS.2016.2561618

Allan, E., Manning, P., Alt, F., Binkenstein, J., Blaser, S., Blüthgen, N., Böhm, S., Grassein, F., Hölzel, N., Klaus, V.H., Kleinebecker, T., Morris, E.K., Oelmann, Y., Prati, D., Renner, S.C., Rillig, M.C., Schaefer, M., Schloter, M., Schmitt, B., Schöning, I., Schrumpf, M., Solly, E., Sorkau, E., Steckel, J., Steffen-Dewenter, I., Stempfhuber, B., Tschapka, M., Weiner, C.N., Weisser, W.W., Werner, M., Westphal, C., Wilcke, W., Fischer, M., 2015. Land use intensification alters ecosystem multifunctionality via loss of biodiversity and changes to functional composition. Ecol. Lett. 18, 834–843. https://doi.org/10.1111/ele.12469

Amin, E., Verrelst, J., Rivera-Caicedo, J.P., Pipia, L., Ruiz-Verdú, A., Moreno, J., 2021. Prototyping Sentinel-2 green LAI and brown LAI products for cropland monitoring. Remote Sens. Environ. 255, 112168. https://doi.org/10.1016/j.rse.2020.112168

Aneece, I.P., Epstein, H., Lerdau, M., 2017. Correlating species and spectral diversities using hyperspectral remote sensing in early-successional fields. Ecol. Evol. 7, 3475–3488. https://doi.org/10.1002/ece3.2876

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

772 mapping of forest biodiversity from space. Nat. Commun. 10, 4757.
773 https://doi.org/10.1038/s41467-019-12737-x
774 Bernhardt-Römermann, M., Römermann, C., Sperlich, S., Schmidt, W., 2011. Explaining
775 grassland biomass - the contribution of climate, species and functional diversity depends
776 on fertilization and mowing frequency: Ecosystem properties and grassland biomass. J.
777 Appl. Ecol. 48, 1088–1097. https://doi.org/10.1111/j.1365-2664.2011.01968.x
778 BKG, 2020. Basis Landscape Model (ATKIS Basis-DLM) © GeoBasis-DE.
779 Blüthgen, N., Dormann, C.F., Prati, D., Klaus, V.H., Kleinebecker, T., Hölzel, N., Alt, F., Boch,
780 S., Gockel, S., Hemp, A., Müller, J., Nieschulze, J., Renner, S.C., Schöning, I.,
781 Schumacher, U., Socher, S.A., Wells, K., Birkhofer, K., Buscot, F., Oelmann, Y.,
782 Rothenwöhrer, C., Scherber, C., Tscharntke, T., Weiner, C.N., Fischer, M., Kalko,
783 E.K.V., Linsenmair, K.E., Schulze, E.-D., Weisser, W.W., 2012. A quantitative index of
784 land-use intensity in grasslands: Integrating mowing, grazing and fertilization. Basic
785 Appl. Ecol. 13, 207–220. https://doi.org/10.1016/j.baae.2012.04.001
786 Bolliger, R., Prati, D., Fisher, M., 2021a. Vegetation Records for Grassland EPs, 2008 - 2020.
787 Biodiversity Exploratories Information System (Bexis). Dataset ID = 27386.
788 Bolliger, R., Prati, D., Schäfer, D., Fischer, M., 2021b. Vegetation Records for 150 Grassland
789 EPs in 2017-2020, Header Data without Species Identities (incl. biomass). Biodiversity
790 Exploratories Information System (Bexis). Dataset IDs=27426, 26151, 24166, 23486.
791 Breiman, L., 2001. Random forest. Mach. Learn. 45, 5–32.
792 https://doi.org/10.1023/A:1010933404324
793 Busch, V., Klaus, V.H., Penone, C., Schäfer, D., Boch, S., Prati, D., Müller, J., Socher, S.A.,
794 Niinemets, Ü., Peñuelas, J., Hölzel, N., Fischer, M., Kleinebecker, T., 2018. Nutrient
795 stoichiometry and land use rather than species richness determine plant functional
796 diversity. Ecol. Evol. 8, 601–616. https://doi.org/10.1002/ece3.3609
797 Campos-Taberner, M., García-Haro, F.J., Martínez, B., Izquierdo-Verdiguier, E., Atzberger, C.,
798 Camps-Valls, G., Gilabert, M.A., 2020. Understanding deep learning in land use
799 classification based on Sentinel-2 time series. Sci. Rep. 10, 17188.
800 https://doi.org/10.1038/s41598-020-74215-5
801 Cavender-Bares, J., Gamon, J.A., Townsend, P.A. (Eds.), 2020. Remote sensing of plant
802 biodiversity. Springer, Cham.
803 Cavender-Bares, J., Polasky, S., King, E., Balvanera, P., 2015. A sustainability framework for
804 assessing trade-offs in ecosystem services. Ecol. Soc. 20, art17.
805 https://doi.org/10.5751/ES-06917-200117
806 Cliff, A. D, Ord, J. K, 1981. Spatial processes: models & applications. Taylor Francis.
807 Csillik, O., Cherbini, J., Johnson, R., Lyons, A., Kelly, M., 2018. Identification of Citrus Trees
808 from Unmanned Aerial Vehicle Imagery Using Convolutional Neural Networks. Drones 2,
809 39. https://doi.org/10.3390/drones2040039
810 Daughtry, C., 2000. Estimating Corn Leaf Chlorophyll Concentration from Leaf and Canopy
811 Reflectance. Remote Sens. Environ. 74, 229–239. https://doi.org/10.1016/S0034-
812 4257(00)00113-9
813 Dusseux, P., Guyet, T., Pattier, P., Barbier, V., Nicolas, H., 2022. Monitoring of grassland
814 productivity using Sentinel-2 remote sensing data. Int. J. Appl. Earth Obs.
815 Geoinformation 111, 102843. https://doi.org/10.1016/j.jag.2022.102843
816 Edirisinghe, A., Clark, D., Waugh, D., 2012. Spatio-temporal modelling of biomass of intensively
817 grazed perennial dairy pastures using multispectral remote sensing. Int. J. Appl. Earth
818 Obs. Geoinformation 16, 5–16. https://doi.org/10.1016/j.jag.2011.11.006
819 EROS, 2017. Shuttle Radar Topography Mission (SRTM) 1 Arc-Second Global.
820 https://doi.org/10.5066/F7PR7TFT
821 Estévez, J., Salinero-Delgado, M., Berger, K., Pipia, L., Rivera-Caicedo, J.P., Wocher, M.,
822 Reyes-Muñoz, P., Tagliabue, G., Boschetti, M., Verrelst, J., 2022. Gaussian processes

retrieval of crop traits in Google Earth Engine based on Sentinel-2 top-of-atmosphere data. Remote Sens. Environ. 273, 112958. https://doi.org/10.1016/j.rse.2022.112958

Fassnacht, F.E., Müllerová, J., Conti, L., Malavasi, M., Schmidtlein, S., 2022. About the link between biodiversity and spectral variation. Appl. Veg. Sci. 25. https://doi.org/10.1111/avsc.12643

Fauvel, M., Lopes, M., Dubo, T., Rivers-Moore, J., Frison, P.-L., Gross, N., Ouin, A., 2020. Prediction of plant diversity in grasslands using Sentinel-1 and -2 satellite image time series. Remote Sens. Environ. 237, 111536. https://doi.org/10.1016/j.rse.2019.111536

Feilhauer, H., Faude, U., Schmidtlein, S., 2011. Combining Isomap ordination and imaging spectroscopy to map continuous floristic gradients in a heterogeneous landscape. Remote Sens. Environ. 115, 2513–2524. https://doi.org/10.1016/j.rse.2011.05.011

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. 117, 28140–28149. https://doi.org/10.1073/pnas.2016210117

Ferner, J., Linstädter, A., Rogass, C., Südekum, K.-H., Schmidtlein, S., 2021. Towards Forage Resource Monitoring in subtropical Savanna Grasslands: going multispectral or hyperspectral? Eur. J. Remote Sens. 54, 364–384. https://doi.org/10.1080/22797254.2021.1934556

Ferner, J., Schmidtlein, S., Guuroh, R.T., Lopatin, J., Linstädter, A., 2018. Disentangling effects of climate and land-use change on West African drylands' forage supply. Glob. Environ. Change 53, 24–38. https://doi.org/10.1016/j.gloenvcha.2018.08.007

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 Appl. Ecol. 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 Sens. 11, 1124. https://doi.org/10.3390/rs11091124

Gara, T.W., Darvishzadeh, R., Skidmore, A.K., Wang, T., Heurich, M., 2019. Accurate modelling of canopy traits from seasonal Sentinel-2 imagery based on the vertical distribution of leaf traits. ISPRS J. Photogramm. Remote Sens. 157, 108–123. https://doi.org/10.1016/j.isprsjprs.2019.09.005

GEE, 2018. Sentinel-1 algorithms in Google Earth Engine [WWW Document]. URL https://developers.google.com/earth-engine/sentinel1

Gholizadeh, H., Gamon, J.A., Townsend, P.A., Zygielbaum, A.I., Helzer, C.J., Hmimina, G.Y., Yu, R., Moore, R.M., Schweiger, A.K., Cavender-Bares, J., 2019. Detecting prairie biodiversity with airborne remote sensing. Remote Sens. Environ. 221, 38–49. https://doi.org/10.1016/j.rse.2018.10.037

Gitelson, A.A., Kaufman, Y.J., Merzlyak, M.N., 1996. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. Remote Sens. Environ. 58, 289–298. https://doi.org/10.1016/S0034-4257(96)00072-7

Gitelson, A.A., Keydan, G.P., Merzlyak, M.N., 2006. Three-band model for noninvasive estimation of chlorophyll, carotenoids, and anthocyanin contents in higher plant leaves. Geophys. Res. Lett. 33, 2006GL026457. https://doi.org/10.1029/2006GL026457

873 Goodfellow, Bengio, Y, Courville, A, 2016. Chapter 6: Deep Feedforwad Networks, in: Deep
874 Learning. Massachusetts: MIT Press, Cambridge, pp. 168–224.
875 Grace, J., Jose, J.S., Meir, P., Miranda, H.S., Montes, R.A., 2006. Productivity and carbon
876 fluxes of tropical savannas. J. Biogeogr. 33, 387–400. https://doi.org/10.1111/j.1365-
877 2699.2005.01448.x
878 Grigera, G., Oesterheld, M., Pacín, F., 2007. Monitoring forage production for farmers' decision
879 making. Agric. Syst. 94, 637–648. https://doi.org/10.1016/j.agsy.2007.01.001
880 Grüner, E., Wachendorf, M., Astor, T., 2020. The potential of UAV-borne spectral and textural
881 information for predicting aboveground biomass and N fixation in legume-grass mixtures.
882 PLOS ONE 15, e0234703. https://doi.org/10.1371/journal.pone.0234703
883 Guerini Filho, M., Kuplich, T.M., Quadros, F.L.F.D., 2020. Estimating natural grassland biomass
884 by vegetation indices using Sentinel 2 remote sensing data. Int. J. Remote Sens. 41,
885 2861–2876. https://doi.org/10.1080/01431161.2019.1697004
886 Guuroh, R.T., Ruppert, J.C., Ferner, J., Čanak, K., Schmidtlein, S., Linstädter, A., 2018. Drivers
887 of forage provision and erosion control in West African savannas—A macroecological
888 perspective. Agric. Ecosyst. Environ. 251, 257–267.
889 https://doi.org/10.1016/j.agee.2017.09.017
890 Hänsel, F., Forteva, S., Wöllauer, S., Nauss, T., 2021. Public climate data of the Biodiversity
891 Exploratories, Instrumentation and Remote Sensing project. Biodiversity Exploratories
892 Information System (Bexis). Dataset IDs= 24766.
893 Heiskanen, J., Rautiainen, M., Stenberg, P., Mõttus, M., Vesanto, V.-H., 2013. Sensitivity of
894 narrowband vegetation indices to boreal forest LAI, reflectance seasonality and species
895 composition. ISPRS J. Photogramm. Remote Sens. 78, 1–14.
896 https://doi.org/10.1016/j.isprsjprs.2013.01.001
897 Henrich, Jung, A., Götze, C., Sasndow, C., Thürkow, D., Gläßer, C., 2012. Entwicklung einer
898 Datenbank für Fernerkundungsindizes.
899 Hill, M.J., Donald, G.E., Vickery, P.J., 1999. Relating Radar Backscatter to Biophysical
900 Properties of Temperate Perennial Grassland. Remote Sens. Environ. 67, 15–31.
901 https://doi.org/10.1016/S0034-4257(98)00063-7
902 Hilpold, A., Seeber, J., Fontana, V., Niedrist, G., Rief, A., Steinwandter, M., Tasser, E.,
903 Tappeiner, U., 2018. Decline of rare and specialist species across multiple taxonomic
904 groups after grassland intensification and abandonment. Biodivers. Conserv. 27, 3729–
905 3744. https://doi.org/10.1007/s10531-018-1623-x
906 Hoeser, T., Bachofer, F., Kuenzer, C., 2020. Object Detection and Image Segmentation with
907 Deep Learning on Earth Observation Data: A Review—Part II: Applications. Remote
908 Sens. 12, 3053. https://doi.org/10.3390/rs12183053
909 Hoeser, T., Kuenzer, C., 2020. Object Detection and Image Segmentation with Deep Learning
910 on Earth Observation Data: A Review-Part I: Evolution and Recent Trends. Remote
911 Sens. 12, 1667. https://doi.org/10.3390/rs12101667
912 Hoffmann, J., Muro, J., Dubovyk, O., 2022. Predicting Species and Structural Diversity of
913 Temperate Forests with Satellite Remote Sensing and Deep Learning. Remote Sens.
914 14, 1631. https://doi.org/10.3390/rs14071631
915 Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., Ferreira, L.G., 2002. Overview of the
916 radiometric and biophysical performance of the MODIS vegetation indices. Remote
917 Sens. Environ. 83, 195–213. https://doi.org/10.1016/S0034-4257(02)00096-2
918 Imran, H.A., Gianelle, D., Scotton, M., Rocchini, D., Dalponte, M., Macolino, S., Sakowska, K.,
919 Pornaro, C., Vescovo, L., 2021. Potential and Limitations of Grasslands α-Diversity
920 Prediction Using Fine-Scale Hyperspectral Imagery. Remote Sens. 13, 2649.
921 https://doi.org/10.3390/rs13142649

Jurgens, C., 1997. The modified normalized difference vegetation index (mNDVI) a new index to determine frost damages in agriculture based on Landsat TM data. Int. J. Remote Sens. 18, 3583–3594. https://doi.org/10.1080/014311697216810

Kattenborn, T., Leitloff, J., Schiefer, F., Hinz, S., 2021. Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. ISPRS J. Photogramm. Remote Sens. 173, 24–49. https://doi.org/10.1016/j.isprsjprs.2020.12.010

Kaufman, Y.J., Tanre, D., 1992. Atmospherically resistant vegetation index (ARVI) for EOS-MODIS. IEEE Trans. Geosci. Remote Sens. 30, 261–270. https://doi.org/10.1109/36.134076

Khare, S., Latifi, H., Rossi, S., 2019. Forest beta-diversity analysis by remote sensing: How scale and sensors affect the Rao's Q index. Ecol. Indic. 106, 105520. https://doi.org/10.1016/j.ecolind.2019.105520

Klingler, A., Schaumberger, A., Vuolo, F., Kalmár, L.B., Pötsch, E.M., 2020. Comparison of Direct and Indirect Determination of Leaf Area Index in Permanent Grassland. PFG – J. Photogramm. Remote Sens. Geoinformation Sci. 88, 369–378. https://doi.org/10.1007/s41064-020-00119-8

Lange, M., Feilhauer, H., Kühn, I., Doktor, D., 2022. Mapping land-use intensity of grasslands in Germany with machine learning and Sentinel-2 time series. Remote Sens. Environ. 277, 112888. https://doi.org/10.1016/j.rse.2022.112888

Latham J, Cumani R, Bloise M, 2014. Global Land Cover SHARE (GLC-SHARE) database Beta-Release Version 1.0.

Le Provost, G., Thiele, J., Westphal, C., Penone, C., Allan, E., Neyret, M., van der Plas, F., Ayasse, M., Bardgett, R.D., Birkhofer, K., Boch, S., Bonkowski, M., Buscot, F., Feldhaar, H., Gaulton, R., Goldmann, K., Gossner, M.M., Klaus, V.H., Kleinebecker, T., Krauss, J., Renner, S., Scherreiks, P., Sikorski, J., Baulechner, D., Blüthgen, N., Bolliger, R., Börschig, C., Busch, V., Chisté, M., Fiore-Donno, A.M., Fischer, M., Arndt, H., Hoelzel, N., John, K., Jung, K., Lange, M., Marzini, C., Overmann, J., Pašalić, E., Perović, D.J., Prati, D., Schäfer, D., Schöning, I., Schrumpf, M., Sonnemann, I., Steffan-Dewenter, I., Tschapka, M., Türke, M., Vogt, J., Wehner, K., Weiner, C., Weisser, W., Wells, K., Werner, M., Wolters, V., Wubet, T., Wurst, S., Zaitsev, A.S., Manning, P., 2021. Contrasting responses of above- and belowground diversity to multiple components of land-use intensity. Nat. Commun. 12, 3918. https://doi.org/10.1038/s41467-021-23931-1

Li, F., Chen, J.Q., Zeng, Y., Wu, B.F., Zhang, X.Q., 2018. Renewed Estimates of Grassland Aboveground Biomass Showing Drought Impacts. J. Geophys. Res. Biogeosciences 123, 138–148. https://doi.org/10.1002/2017JG004255

Lobert, F., Holtgrave, A.-K., Schwieder, M., Pause, M., Vogt, J., Gocht, A., Erasmi, S., 2021. Mowing event detection in permanent grasslands: Systematic evaluation of input features from Sentinel-1, Sentinel-2, and Landsat 8 time series. Remote Sens. Environ. 267, 112751. https://doi.org/10.1016/j.rse.2021.112751

Lopes, M., Fauvel, M., Ouin, A., Girard, S., 2017. Potential of Sentinel-2 and SPOT5 (Take5) time series for the estimation of grasslands biodiversity indices, in: 2017 9th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp). Presented at the 2017 9th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp), IEEE, Brugge, Belgium, pp. 1–4. https://doi.org/10.1109/Multi-Temp.2017.8035206

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.

Ma, J., Li, Y., Chen, Y., Du, K., Zheng, F., Zhang, L., Sun, Z., 2019. Estimating above ground biomass of winter wheat at early growth stages using digital images and deep

convolutional neural network. Eur. J. Agron. 103, 117–129. https://doi.org/10.1016/j.eja.2018.12.004

Magiera, A., Feilhauer, H., Waldhardt, R., Wiesmair, M., Otte, A., 2017. Modelling biomass of mountainous grasslands by including a species composition map. Ecol. Indic. 78, 8–18. https://doi.org/10.1016/j.ecolind.2017.02.039

Malenovský, Z., Homolová, L., Zurita-Milla, R., Lukeš, P., Kaplan, V., Hanuš, J., Gastellu-Etchegorry, J.-P., Schaepman, M.E., 2013. Retrieval of spruce leaf chlorophyll content from airborne image data using continuum removal and radiative transfer. Remote Sens. Environ. 131, 85–102. https://doi.org/10.1016/j.rse.2012.12.015

Malenovský, Z., Rott, H., Cihlar, J., Schaepman, M.E., García-Santos, G., Fernandes, R., Berger, M., 2012. Sentinels for science: Potential of Sentinel-1, -2, and -3 missions for scientific observations of ocean, cryosphere, and land. Remote Sens. Environ. 120, 91–101. https://doi.org/10.1016/j.rse.2011.09.026

Mariotto, I., Thenkabail, P.S., Huete, A., Slonecker, E.T., Platonov, A., 2013. Hyperspectral versus multispectral crop-productivity modeling and type discrimination for the HyspIRI mission. Remote Sens. Environ. 139, 291–305. https://doi.org/10.1016/j.rse.2013.08.002

Marquard, E., Weigelt, A., Roscher, C., Gubsch, M., Lipowsky, A., Schmid, B., 2009. Positive biodiversity-productivity relationship due to increased plant density. J. Ecol. 97, 696–704. https://doi.org/10.1111/j.1365-2745.2009.01521.x

Masenyama, A., Mutanga, O., Dube, T., Bangira, T., Sibanda, M., Mabhaudhi, T., 2022. A systematic review on the use of remote sensing technologies in quantifying grasslands ecosystem services. GIScience Remote Sens. 59, 1000–1025. https://doi.org/10.1080/15481603.2022.2088652

Meng, B., Liang, T., Yi, S., Yin, J., Cui, X., Ge, J., Hou, M., Lv, Y., Sun, Y., 2020. Modeling Alpine Grassland Above Ground Biomass Based on Remote Sensing Data and Machine Learning Algorithm: A Case Study in East of the Tibetan Plateau, China. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 13, 2986–2995. https://doi.org/10.1109/JSTARS.2020.2999348

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

Meyer, H., Reudenbach, C., Hengl, T., Katurji, M., Nauss, T., 2018. Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation. Environ. Model. Softw. 101, 1–9. https://doi.org/10.1016/j.envsoft.2017.12.001

Meyer, H., Reudenbach, C., Wöllauer, S., Nauss, T., 2019. Importance of spatial predictor variable selection in machine learning applications – Moving from data reproduction to spatial prediction. Ecol. Model. 411, 108815. https://doi.org/10.1016/j.ecolmodel.2019.108815

Miao, L., Müller, D., Cui, X., Ma, M., 2017. Changes in vegetation phenology on the Mongolian Plateau and their climatic determinants. PLOS ONE 12, e0190313. https://doi.org/10.1371/journal.pone.0190313

Moore, J.C., 2013. Diversity, Taxonomic versus Functional, in: Encyclopedia of Biodiversity. Elsevier, pp. 648–656. https://doi.org/10.1016/B978-0-12-384719-5.00036-8

Müller, I.B., Buhk, C., Lange, D., Entling, M.H., Schirmel, J., 2016. Contrasting effects of irrigation and fertilization on plant diversity in hay meadows. Basic Appl. Ecol. 17, 576–585. https://doi.org/10.1016/j.baae.2016.04.008

Muramatsu, K., 2019. Use of chlorophyll index-green and the red-edge chlorophyll index to derive an algorithm for estimating gross primary production capacity, in: Neale, C.M., Maltese, A. (Eds.), Remote Sensing for Agriculture, Ecosystems, and Hydrology XXI.

Presented at the Remote Sensing for Agriculture, Ecosystems, and Hydrology XXI, SPIE, Strasbourg, France, p. 6. https://doi.org/10.1117/12.2532468

Muro, J., doninck, J.V., Tuomisto, H., Higgins, M.A., Moulatlet, G.M., Ruokolainen, K., 2016. Floristic composition and across-track reflectance gradient in Landsat images over Amazonian forests. ISPRS J. Photogramm. Remote Sens. 119, 361–372. https://doi.org/10.1016/j.isprsjprs.2016.06.016

Muro, J., Strauch, A., Fitoka, E., Tompoulidou, M., Thonfeld, F., 2019. Mapping Wetland Dynamics With SAR-Based Change Detection in the Cloud. IEEE Geosci. Remote Sens. Lett. 1–4. https://doi.org/10.1109/LGRS.2019.2903596

Muro, J., Varea, A., Strauch, A., Guelmami, A., Fitoka, E., Thonfeld, F., Diekkrüger, B., Waske, B., 2020. Multitemporal optical and radar metrics for wetland mapping at national level in Albania. Heliyon 6, e04496. https://doi.org/10.1016/j.heliyon.2020.e04496

Naidoo, L., van Deventer, H., Ramoelo, A., Mathieu, R., Nondlazi, B., Gangat, R., 2019. Estimating above ground biomass as an indicator of carbon storage in vegetated wetlands of the grassland biome of South Africa. Int. J. Appl. Earth Obs. Geoinformation 78, 118–129. https://doi.org/10.1016/j.jag.2019.01.021

O'Malley, T., Bursztein, E., Long, J., Chollet, F., Jin, H., Invernizzi, L., others, 2019. KerasTuner.

Osco, L.P., Ramos, A.P.M., Faita Pinheiro, M.M., Moriya, É.A.S., Imai, N.N., Estrabis, N., Ianczyk, F., Araújo, F.F. de, Liesenberg, V., Jorge, L.A. de C., Li, J., Ma, L., Gonçalves, W.N., Marcato Junior, J., Eduardo Creste, J., 2020. A Machine Learning Framework to Predict Nutrient Content in Valencia-Orange Leaf Hyperspectral Measurements. Remote Sens. 12, 906. https://doi.org/10.3390/rs12060906

Ostrowski, A., Nieschulze, J., Schulze, E.-D., Fisher, M., Ayasse, M., Weisser, W., König-Ries, B., 2018. Basic Information of all Experimental Plots (EPs). Biodiversity Exploratories Information System (Bexis). Dataset ID= 20826.

Pacheco-Labrador, J., Migliavacca, M., Ma, X., Mahecha, M., Carvalhais, N., Weber, U., Benavides, R., Bouriaud, O., Barnoaiea, I., Coomes, D.A., Bohn, F.J., Kraemer, G., Heiden, U., Huth, A., Wirth, C., 2022. Challenging the link between functional and spectral diversity with radiative transfer modeling and data. Remote Sens. Environ. 280, 113170. https://doi.org/10.1016/j.rse.2022.113170

Pan, Q., Symstad, A.J., Bai, Y., Huang, J., Wu, J., Naeem, S., Chen, D., Tian, D., Wang, Q., Han, X., 2022. Biodiversity–productivity relationships in a natural grassland community vary under diversity loss scenarios. J. Ecol. 110, 210–220. https://doi.org/10.1111/1365-2745.13797

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. 12, 2825–2830.

Peng, Y., Fan, M., Bai, L., Sang, W., Feng, J., Zhao, Z., Tao, Z., 2019. Identification of the Best Hyperspectral Indices in Estimating Plant Species Richness in Sandy Grasslands. Remote Sens. 11, 588. https://doi.org/10.3390/rs11050588

Petermann, J.S., Buzhdygan, O.Y., 2021. Grassland biodiversity. Curr. Biol. 31, R1195–R1201. https://doi.org/10.1016/j.cub.2021.06.060

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

Quan, X., He, B., Yebra, M., Yin, C., Liao, Z., Zhang, X., Li, X., 2017. A radiative transfer model-based method for the estimation of grassland aboveground biomass. Int. J. Appl. Earth Obs. Geoinformation 54, 159–168. https://doi.org/10.1016/j.jag.2016.10.002

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 Sens. Environ. 239, 111626. https://doi.org/10.1016/j.rse.2019.111626

Rao, C.R., 1982. Diversity and dissimilarity coefficients: A unified approach. Theor. Popul. Biol. 21, 24–43. https://doi.org/10.1016/0040-5809(82)90004-1

Rapinel, S., Mony, C., Lecoq, L., Clément, B., Thomas, A., Hubert-Moy, L., 2019. Evaluation of Sentinel-2 time-series for mapping floodplain grassland plant communities. Remote Sens. Environ. 223, 115–129. https://doi.org/10.1016/j.rse.2019.01.018

Reinermann, S., Gessner, U., Asam, S., Kuenzer, C., Dech, S., 2019. The Effect of Droughts on Vegetation Condition in Germany: An Analysis Based on Two Decades of Satellite Earth Observation Time Series and Crop Yield Statistics. Remote Sens. 11, 1783. https://doi.org/10.3390/rs11151783

Rey, S.J., Anselin, L., 2007. PySAL: A Python Library of Spatial Analytical Methods. Rev. Reg. Stud. 37, 5–27.

Reynolds, S.G., Frame, J., Food and Agriculture Organization of the United Nations (Eds.), 2005. Grasslands: developments, opportunities, perspectives. Food and Agricultural Organization of the United Nations; Science Publishers, Inc, Rome : Enfield, NH.

Rocchini, D., Marcantonio, M., Da Re, D., Bacaro, G., Feoli, E., Foody, G.M., Furrer, R., Harrigan, R.J., Kleijn, D., Iannacito, M., Lenoir, J., Lin, M., Malavasi, M., Marchetto, E., Meyer, R.S., Moudry, V., Schneider, F.D., Šímová, P., Thornhill, A.H., Thouverai, E., Vicario, S., Wayne, R.K., Ricotta, C., 2021. From zero to infinity: Minimum to maximum diversity of the planet by spatio-parametric Rao's quadratic entropy. Glob. Ecol. Biogeogr. 30, 1153–1162. https://doi.org/10.1111/geb.13270

Rocchini, D., Marcantonio, M., Da Re, D., Chirici, G., Galluzzi, M., Lenoir, J., Ricotta, C., Torresani, M., Ziv, G., 2019. Time-lapsing biodiversity: An open source method for measuring diversity changes by remote sensing. Remote Sens. Environ. 231, 111192. https://doi.org/10.1016/j.rse.2019.05.011

Rocchini, D., Marcantonio, M., Ricotta, C., 2017. Measuring Rao's Q diversity index from remote sensing: An open source solution. Ecol. Indic. 72, 234–238. https://doi.org/10.1016/j.ecolind.2016.07.039

Rossi, C., Kneubühler, M., Schütz, M., Schaepman, M.E., Haller, R.M., Risch, A.C., 2020. From local to regional: Functional diversity in differently managed alpine grasslands. Remote Sens. Environ. 236, 111415. https://doi.org/10.1016/j.rse.2019.111415

Rossi, M., Niedrist, G., Asam, S., Tonon, G., Tomelleri, E., Zebisch, M., 2019. A Comparison of the Signal from Diverse Optical Sensors for Monitoring Alpine Grassland Dynamics. Remote Sens. 11, 296. https://doi.org/10.3390/rs11030296

Ruppert, J.C., Linstädter, A., 2014. Convergence between ANPP estimation methods in grasslands — A practical solution to the comparability dilemma. Ecol. Indic. 36, 524–531. https://doi.org/10.1016/j.ecolind.2013.09.008

Schulze-Brüninghoff, D., Hensgen, F., Wachendorf, M., Astor, T., 2019. Methods for LiDAR-based estimation of extensive grassland biomass. Comput. Electron. Agric. 156, 693–699. https://doi.org/10.1016/j.compag.2018.11.041

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 – J. Photogramm. Remote Sens. Geoinformation Sci. https://doi.org/10.1007/s41064-020-00120-1

Schwieder, M., Wesemeyer, M., Frantz, D., Pfoch, K., Erasmi, S., Pickert, J., Nendel, C., Hostert, P., 2022. Mapping grassland mowing events across Germany based on combined Sentinel-2 and Landsat 8 time series. Remote Sens. Environ. 269, 112795. https://doi.org/10.1016/j.rse.2021.112795

Smith, L.N., Byrne, A., Hansen, M.F., Zhang, W., Smith, M.L., 2019. Weed classification in grasslands using convolutional neural networks, in: Zelinski, M.E., Taha, T.M., Howe, J., Awwal, A.A., Iftekharuddin, K.M. (Eds.), Applications of Machine Learning. Presented at the Applications of Machine Learning, SPIE, San Diego, United States, p. 42. https://doi.org/10.1117/12.2530092

Socher, S.A., Prati, D., Boch, S., Müller, J., Baumbach, H., Gockel, S., Hemp, A., Schöning, I., Wells, K., Buscot, F., Kalko, E.K.V., Linsenmair, K.E., Schulze, E.-D., Weisser, W.W., Fischer, M., 2013. Interacting effects of fertilization, mowing and grazing on plant species diversity of 1500 grasslands in Germany differ between regions. Basic Appl. Ecol. 14, 126–136. https://doi.org/10.1016/j.baae.2012.12.003

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

Tucker, C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. Remote Sens. Environ. 8, 127–150. https://doi.org/10.1016/0034-4257(79)90013-0

van der Plas, F., 2019. Biodiversity and ecosystem functioning in naturally assembled communities. Biol. Rev. brv.12499. https://doi.org/10.1111/brv.12499

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

Villoslada Peciña, M., Bergamo, T.F., Ward, R.D., Joyce, C.B., Sepp, K., 2021. A novel UAV-based approach for biomass prediction and grassland structure assessment in coastal meadows. Ecol. Indic. 122, 107227. https://doi.org/10.1016/j.ecolind.2020.107227

Wadoux, A.M.J.-C., Heuvelink, G.B.M., de Bruin, S., Brus, D.J., 2021. Spatial cross-validation is not the right way to evaluate map accuracy. Ecol. Model. 457, 109692. https://doi.org/10.1016/j.ecolmodel.2021.109692

Wang, J., Rich, P.M., Price, K.P., Kettle, W.D., 2005. Relations between NDVI, Grassland Production, and Crop Yield in the Central Great Plains. Geocarto Int. 20, 5–11. https://doi.org/10.1080/10106040508542350

Wang, J., Xiao, X., Bajgain, R., Starks, P., Steiner, J., Doughty, R.B., Chang, Q., 2019. Estimating leaf area index and aboveground biomass of grazing pastures using Sentinel-1, Sentinel-2 and Landsat images. ISPRS J. Photogramm. Remote Sens. 154, 189–201. https://doi.org/10.1016/j.isprsjprs.2019.06.007

Wang, R., Gamon, J.A., Cavender-Bares, J., Townsend, P.A., Zygielbaum, A.I., 2018. The spatial sensitivity of the spectral diversity–biodiversity relationship: an experimental test in a prairie grassland. Ecol. Appl. 28, 541–556. https://doi.org/10.1002/eap.1669

Wehner, K., Renker, C., Simons, N.K., Weisser, W.W., Blüthgen, N., 2021. Narrow environmental niches predict land-use responses and vulnerability of land snail assemblages. BMC Ecol. Evol. 21, 15. https://doi.org/10.1186/s12862-020-01741-1

Weiss, M., Baret, F., Jay, S., 2020. S2ToolBox Level 2 products: LAI, FAPAR, FCOVER.

Willmott, C.J., 1981. On the validation of models. Phys. Geogr. 2, 184–194. https://doi.org/10.1080/02723646.1981.10642213

Wolanin, A., Mateo-García, G., Camps-Valls, G., Gómez-Chova, L., Meroni, M., Duveiller, G., Liangzhi, Y., Guanter, L., 2020. Estimating and understanding crop yields with explainable deep learning in the Indian Wheat Belt. Environ. Res. Lett. 15, 024019. https://doi.org/10.1088/1748-9326/ab68ac

Wu, H., Li, Z.-L., 2009. Scale Issues in Remote Sensing: A Review on Analysis, Processing and Modeling. Sensors 9, 1768–1793. https://doi.org/10.3390/s90301768

Xu, L., Saatchi, S.S., Yang, Y., Yu, Y., White, L., 2016. Performance of non-parametric algorithms for spatial mapping of tropical forest structure. Carbon Balance Manag. 11, 18. https://doi.org/10.1186/s13021-016-0062-9

Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., Xu, H., Tan, W., Yang, Q., Wang, J., Gao, J., Zhang, L., 2020. Deep learning in environmental remote sensing: Achievements and challenges. Remote Sens. Environ. 241, 111716. https://doi.org/10.1016/j.rse.2020.111716

Zeng, N., Ren, X., He, H., Zhang, L., Zhao, D., Ge, R., Li, P., Niu, Z., 2019. Estimating grassland aboveground biomass on the Tibetan Plateau using a random forest algorithm. Ecol. Indic. 102, 479–487. https://doi.org/10.1016/j.ecolind.2019.02.023

Zhang, D., Pan, Y., Zhang, J., Hu, T., Zhao, J., Li, N., Chen, Q., 2020. A generalized approach based on convolutional neural networks for large area cropland mapping at very high resolution. Remote Sens. Environ. 247, 111912. https://doi.org/10.1016/j.rse.2020.111912

Zhao, Y., Liu, Z., Wu, J., 2020. Grassland ecosystem services: a systematic review of research advances and future directions. Landsc. Ecol. 35, 793–814. https://doi.org/10.1007/s10980-020-00980-3

Zhou, Y., Zhang, L., Xiao, J., Chen, S., Kato, T., Zhou, G., 2014. A Comparison of Satellite-Derived Vegetation Indices for Approximating Gross Primary Productivity of Grasslands. Rangel. Ecol. Manag. 67, 9–18. https://doi.org/10.2111/REM-D-13-00059.1