Mapping land-use intensity of grasslands in Germany with machine learning and Sentinel-2 time series

Maximilian Lange[a,b,*], Hannes Feilhauer[b,a], Ingolf Kühn[c,d,e], Daniel Doktor[a,b]

[a] *Department of Remote Sensing, Helmholtz-Centre for Environmental Research - UFZ, Permoserstr. 15, 04318 Leipzig, Germany*
[b] *Remote Sensing Centre for Earth System Research, RSC4Earth, Talstr. 35, 04103 Leipzig, Germany*
[c] *Department of Community Ecology, Helmholtz-Centre for Environmental Research - UFZ, Theodor-Lieser-Str. 4, 06120 Halle, Germany*
[d] *Institute for Biology/Geobotany & Botanical Garden, Martin Luther University Halle-Wittenberg, Große Steinstraße 79/80, 06108 Halle, Germany*
[e] *German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Puschstr. 4, 04103 Leipzig, Germany*

Abstract

Information on grassland land-use intensity (LUI) is crucial for understanding trends and dynamics in biodiversity, ecosystem functioning, earth system science and environmental monitoring. LUI is a major driver for numerous environmental processes and indicators, such as primary production, nitrogen deposition and resilience to climate extremes. However, large extent, high resolution data on grassland LUI is rare. New satellite generations, such as Copernicus Sentinel-2, enable a spatially comprehensive detection of the mainly subtle changes induced by land-use intensification by their fine spatial and temporal resolution. We developed a methodology quantifying key parameters of grassland LUI such as grazing intensity, mowing frequency

*Corresponding author
Email address: `maximilian.lange@ufz.de` (Maximilian Lange)

and fertiliser application across Germany using Convolutional Neural Networks (CNN) on Sentinel-2 satellite data with 20 m x 20 m spatial resolution. Subsequently, these land-use components were used to calculate a continuous LUI index. Predictions of LUI and its components were validated using comprehensive *in situ* grassland management data. A feature contribution analysis using Shapley values substantiates the applicability of the methodology by revealing a high relevance of springtime satellite observations and spectral bands related to vegetation health and structure. We achieved an overall classification accuracy of up to 66% for grazing intensity, 68% for mowing, 85% for fertilisation and an r^2 of 0.82 for subsequently depicting LUI. We evaluated the methodology's robustness with a spatial 3-fold cross-validation by training and predicting on geographically distinctly separated regions. Spatial transferability was assessed by delineating the models' area of applicability. The presented methodology enables a high resolution, large extent mapping of land-use intensity of grasslands.

Keywords: Mowing, Grazing, Fertilisation, Convolutional Neural Networks, Random Forest, Classification, Deep Learning, optical satellite data

---

1. Introduction

Grasslands cover about one third of the global land surface and are the most cultivated biome on Earth (IPCC, 2019). They provide numerous ecosystem services, such as carbon sequestration and food production (FAO, 2009), and have a high importance for conservation as they are often species rich (Gibson, D. J., 2009; Wilson et al., 2012). Biodiversity, ecosystem services and functions in grasslands are strongly affected by the management regime (IPBES, 2019), i.e. grazing, fertilisation, timing and frequency of mowing events (Blüthgen et al., 2012), and its quantitative variation (FAO, 2009; Gibson, D. J., 2009; IPCC, 2019). The optimisation of farming efficiency towards yield maximisation usually leads to an increase in these treatments (hereafter referred to as land-use intensity components) and thus land-use intensity (LUI), which may have numerous negative implications for the environment, such as biodiversity loss (Kruess and Tscharntke, 2002; Wesche et al., 2012; Tsiafouli et al., 2015), water pollution (FAO, 2009), land degradation (IPBES, 2019; IPCC, 2019) and increased carbon emissions (Jones and Donnelly, 2004; Post and Kwon, 2000; Conant et al., 2001). Monitoring LUI on a large spatial extent is thus crucial for Earth system science and environmental monitoring to support decision making and reporting of climate-relevant processes.

Combining several LUI components into a single quantifiable measure is not a trivial issue as these components are often not independent (Blüthgen et al., 2012). For example, meadows are commonly fertilised at higher levels than pastures. However, Blüthgen et al. (2012) proposed a LUI index combining fertiliser amount, mowing frequency and grazing intensity irrespective of the complex interactions between these components. Such a simple combination allows to analyse land-use effects across landscapes in different years without requiring additional (mixed) land-use types accounting for spatio-temporal grassland management variations, while a more mechanistic understanding can be gained by analysing the underlying components (Blüthgen et al., 2012).

Remote sensing methods may be able to assess LUI in grasslands, as mowing, grazing and fertilisation all affect vegetation structure, composition and vitality, albeit to different degrees and over varying time scales (de Bello et al., 2011; Bernhardt-Römermann et al., 2011; Socher et al., 2013). Mowing and grazing remove vegetation and thus influence its structure and vitality in the short-term (Dusseux et al., 2014), and species composition in the long-term, e.g. by inhibiting the growth of species not adapted and facilitate those (pre-)adapted to this management (Chapin III et al., 2000; Socher et al., 2013; Bouchet et al., 2017). Livestock manure and fertilisation provide nutrients, hence influencing vegetation structure and vitality in the short term, while determining the competitive interactions among plant species in the long-term (de Bello et al., 2011; Bernhardt-Römermann et al., 2011; Gibson, D. J., 2009). Land-use has thus a pronounced effect on grassland reflectance properties that varies with its intensity. Analysis of these reflectance characteristics enables the identification of the underlying vegetation structure, composition and vitality (Jacquemoud et al., 2009; Ramoelo et al., 2015; Sakowska et al., 2016) and consequently supports the assessment of grassland management with remote sensing methods. However, the relationships between vegetation and reflected light are complex (Jacquemoud et al., 2009; Doktor et al., 2014). These relationships are commonly inferred by analysing spectral bands or vegetation indices (VI), such as the Normalized Difference Vegetation Index (NDVI), and their temporal changes, hereafter called time series analysis. Further, classification approaches using clustering algorithms or machine learning on mono- or multi-temporal satellite observations are increasingly used in recent studies.

Studies detecting grassland management practices with remote sensing methods mainly focus on one of the above mentioned LUI components (Reinermann et al., 2020), while only a few studies infer LUI by depicting multiple components (Gómez Giménez et al., 2017; Stumpf et al., 2020). The detection of mowing events is commonly performed by VI time series analysis (Estel et al., 2018; Kolecka et al., 2018; Griffiths et al., 2020) or by analysing changes in time series of Earth observation data products like the leaf area index (LAI) derived from radiative transfer model inversion (Dusseux et al., 2014; Asam et al., 2015), and only a few studies use a classification approach, e.g. of Synthetic Aperture Radar backscatter (SAR; Siegmund et al., 2016; Taravat et al., 2019). Grazing intensity was recently mapped by using VI change detection (Blanco et al., 2009; Li et al., 2016) or regression techniques linking VI and above-ground biomass as a proxy for grazing intensity (Ma et al., 2019). Studies identifying fertiliser applications or estimating plant nutrient status in grasslands with remote sensing techniques are rare and commonly use hyperspectral data (Pellissier et al., 2015; Sibanda et al., 2015). However, Hollberg and Schellberg (2017) demonstrated the feasibility to distinguish different fertiliser treatments in grasslands by using vegetation indices based on simulated RapidEye data. Research on nutrient status or supply in cropland also commonly uses hyperspectral data (Cilia et al., 2014; Xia et al., 2016).

The majority of past studies observing grassland management from space used coarse spatial resolution sensors such as NASA Advanced Very High Resolution Radiometer (AVHRR) or Moderate-resolution Imaging Spectroradiometer (MODIS), or coarse temporal resolution sensors such as Landsat

(Reinermann et al., 2020). Contrarily, grassland management activities are often conducted within small areas of fragmented landscapes and their spectral footprint diminishes quickly, especially after mowing events, due to the rapid regrowth of perennials in summer (Griffiths et al., 2020; Stumpf et al., 2020). Additionally, information about the farmer's grassland management on the level of individual fields or farms is commonly not available, but necessary for policy makers (Griffiths et al., 2020) and research would greatly benefit from spatially explicit land management information (Franke et al., 2012; Reinermann et al., 2020). To cope with these issues, recent studies worked with new satellite generations, such as Copernicus Sentinel-2 or RapidEye, which enable the detection of the mainly subtle changes induced by land-use intensification by their fine spatial and temporal resolution (Gómez Giménez et al., 2017; Griffiths et al., 2020; Reinermann et al., 2020). However, the majority of studies often focussed on homogeneous grasslands or on local scales (Franke et al., 2012; Lopes et al., 2017; Taravat et al., 2019), mainly due to the lack or quality of calibration and validation data (Ali et al., 2016; Kuemmerle et al., 2013). The spatial and temporal heterogeneity of cultivated grasslands remains a key issue: management practices often change over time or are conducted only in subsets of the parcels, e.g. piece-wise mowing or grazing, calling for spatially and temporally explicit, high quality, ground truth data (Estel et al., 2018; Reinermann et al., 2020). The lack of these datasets for training and validation severely hampers the development of accurate remote sensing products. Validation is thus often done without *in situ* validation data, e.g. by visual interpretation of time series (Gómez Giménez et al., 2017; Kolecka et al., 2018; Griffiths et al., 2020) or spatially non-explicit by comparing statistical features or the density or frequency distribution of predictions with regional management statistics (Estel et al., 2018; Stumpf et al., 2020).

Here, we aim to infer LUI quantifying the components mowing frequency, grazing intensity and fertilisation with machine learning methods to appropriately account for the complex, e.g. non-linear, relationships between remotely sensed observations and LUI components on the ground. While used for classification of management type or general use intensity, only a low number of studies use machine learning for mowing detection (Halabuk et al., 2015; Taravat et al., 2019), and we found no studies using machine learning for the detection of grassland fertilisation or grazing intensity. We train and validate the models using a spatially explicit grassland management dataset of the DFG (Deutsche Forschungsgemeinschaft - German Research Foundation) Biodiversity Exploratories program (Fischer et al., 2010). The functional relationships between the intensity measures on the ground and satellite observations are evaluated with Shapley additive explanations (SHAP; Lundberg and Lee, 2017; Lundberg et al., 2020) summarising explanations of individual predictions to gain information about the global model structure and to robustly extract variable importances within the underlying model. We evaluated the methodology's robustness with a spatial 3-fold cross-validation by training and predicting on geographically distinctly separated regions. Spatial transferability was assessed by delineating the methodology's area of applicability considering the feature space given by the training data and the models' variable importances (Meyer and Pebesma, 2021). Finally, we aim to apply the method to the national scale by mapping LUI of all grasslands in Germany. In summary, we address the following research questions:

(1) Can mowing frequency, grazing intensity and fertilisation be quantified using CNN by identifying and relating respective spectral and temporal patterns in Sentinel-2 reflectance time series?

(2) Can we translate the LUI index proposed by Blüthgen et al. (2012) into a remote sensing based framework?

(3) Are the LUI component models, trained and validated using three observatories representative for grasslands in Germany, transferable to all grasslands in Germany?

2. Data

2.1. Study area

This study focuses on grassland areas in Germany. Their distribution is shown in Figure 1 according to the digital landscape model (DLM) of the official topographic-cartographic information system (ATKIS; Bundesamt für Kartographie und Geodäsie, 2015) providing information about topographic objects within Germany in vector format. Contrary to croplands, grasslands in Germany are usually scattered small-sized areas. Accumulations of grasslands are found in a) northern Germany in vicinity of the North Sea, b) in fluvial valleys, such as the Elbe Valley, c) in topographically high levels, such as the Central German Uplands and d) at the Alpine Foothills in southern Germany.

Field data were available from three observatories of the DFG Biodiversity Exploratories (Fischer et al., 2010): *Schorfheide* (SCH) is located in the

North-East, *Hainich* (HAI) in the center and *Schwäbische Alb* (ALB) in the South-West of Germany (Figure 1). Germany's climate is dominated by humid westerly winds and the influence of the oceanic climate decreases from western and coastal areas towards the eastern parts (Köppen, 1918; Stanners and Bourdeau, 1995). The increasing altitude from North to South modifies these general climate patterns, resulting in considerably different climate conditions (Doktor, 2008). Thus, the three observatories represent the climatic gradient of Germany (Köppen, 1918; Stanners and Bourdeau, 1995). Further, these observatories represent most of variation in land-use typical for grasslands in Germany, including hardly managed up to intensively used areas (Fischer et al., 2010).

We chose four regions (Figure 1) for demonstration, outside of the DFG Biodiversity Exploratories due to data privacy agreements not allowing for the disclosure of exact locations, with an extent of 10 km x 10 km and a high number of grassland pixels for visual and statistical interpretation: Region (a) is located in northern Germany, in the district Dithmarschen of federal state Schleswig-Holstein composed of mostly fertile tidal marshes and a sandy Geest, (b) covers the northern part of the nature reserve Ohre-Drömling and the surrounding landscape in the federal state Saxony-Anhalt composed of bogs, marshes and farmland, (c) covers parts of the transition areas of the biosphere reserve Rhön in the federal state Thuringia diverse in land-use and morphology and (d) is located in southern Germany in the district Oberallgäu of the federal state Bavaria where large agricultural parts are used for dairy farming.

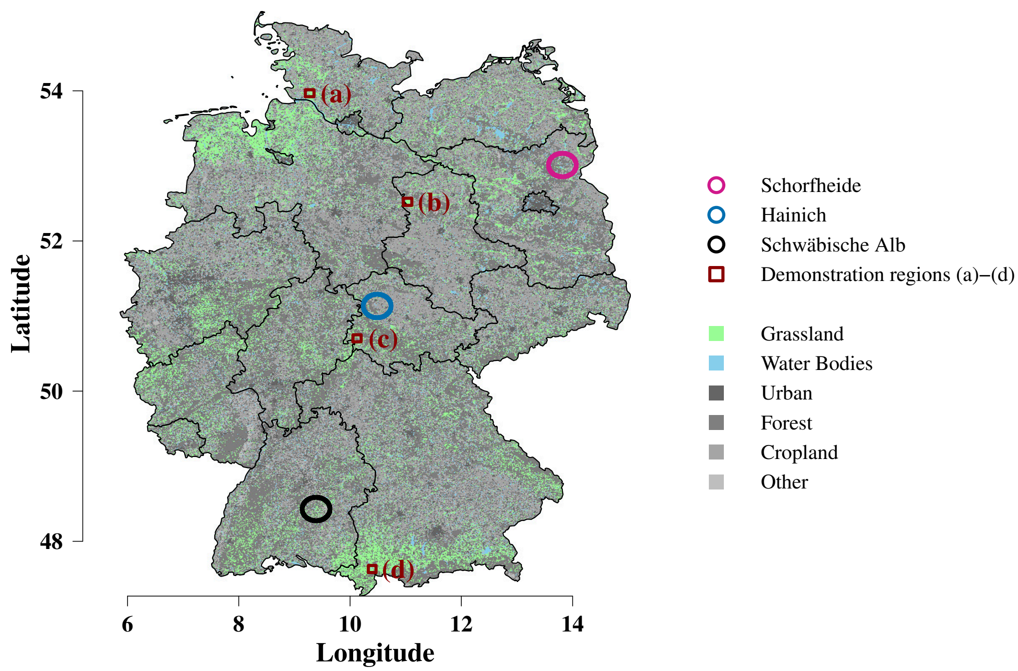

Figure 1: Location of study areas in Germany. Calibration and validation plots from the DFG Biodiversity Exploratories are marked with circles. We are not allowed to publish exact parcel positions due to privacy policies. Exact positions of the four demonstration regions chosen for visual and statistical interpretation are given as red boxes. Background colors show simplified land cover information. State borders are shown as black lines (© GeoBasis-DE / Bundesamt für Kartographie und Geodäsie, 2017).

2.2. *Land-use intensity measures*

The spatially explicit grassland management data of the DFG Biodiversity Exploratories (Fischer et al., 2010) include information about livestock, fertilisation and mowing events (Vogt et al., 2020) in three regions across Germany (see Figure 1) from 2006 to 2018. Here, we only used records from 2017 and 2018 due to the availability of respective satellite data. Each of the three observatories contains 50 plots (Ostrowski et al., 2020). All plots have a size of 50 m x 50 m and are situated in a grassland area with a median size of 10 ha (minimum 0.49 ha, maximum 148 ha). The total size of represented grassland area is around 26 km^2 (see Table 1). Environmental monitoring units are situated within each plot. Several manipulative experiments not related to this study, such as fertilisation, disturbance and climate change experiments, were carried out within 27 of these plots, nine in each observatory (Fischer et al., 2010).

Livestock information is given as livestock units per plot for four grazing periods per year. Livestock units are a conversion of livestock numbers depending on species and age (Table 2; Fischer et al., 2010) and were multiplied with the grazing duration (number of days of the four grazing periods), summed up and divided by the length of the year (days) and the grazing area (ha) to get a normalised value of grazing intensity (livestock units per day and ha). Mowing information is given in up to five mowing dates per plot. Fertilisation information include type and amount of fertiliser, as well as dates and total number of fertilisation events per plot, excluding dung depositions by livestock during grazing.

Table 1: Summary of descriptive statistics on grassland management data of Biodiversity Exploratories for 2017 and 2018. Statistics were generated using 50 plots per observatory (ALB: *Schwäbische Alb*, HAI: *Hainich* and SCH: *Schorfheide*) and year. Each plot is situated in a grassland, the sum ($\sum_s$), first quartile (Q_1), median (Q_2) and third quartile (Q_3) of the size of these grassland areas in the respective region were given in *ha*. Grazing intensities (G_I) were summarised by the expected value ($E[G_I]$) and variance ($V[G_I]$) of their beta distribution in the interval $[0, c_G]$, with c_G representing the maximum value of grazing intensity in the respective area. Mowing (M) and fertilisation (F) counts were given by the λ of their Poisson distribution (λ_M and λ_F, respectively).

		Size (ha)								
Region	**Year**	$\boldsymbol{\sum_s}$	$\boldsymbol{Q_1}$	$\boldsymbol{Q_2}$	$\boldsymbol{Q_3}$	$\boldsymbol{E[G_I]}$	$\boldsymbol{V[G_I]}$	$\boldsymbol{c_G}$	$\boldsymbol{\lambda_M}$	$\boldsymbol{\lambda_F}$
ALB	2017	311	2.4	5.5	8.0	0.20	0.08	1.82	1.36	1.12
	2018	311	2.4	5.5	8.0	0.22	0.10	1.71	1.42	0.96
HAI	2017	1138	6.0	11.1	19.8	0.24	0.10	3.76	0.90	0.64
	2018	1127	6.0	11.1	19.6	0.22	0.12	3.35	0.68	0.70
SCH	2017	1180	10.2	16.0	32.0	0.47	0.31	2.46	0.88	0.10
	2018	1175	9.7	15.3	32.0	0.46	0.28	2.20	0.84	0.04
Overall	2017	2629	5.3	10.0	18.6	0.33	0.24	3.76	1.05	0.62
	2018	2614	5.3	9.8	17.6	0.32	0.24	3.35	0.98	0.57

Table 2: Definition of livestock units based on species and age as given in the land management data of the DFG Biodiversity Exploratories (slightly different from those given in Fischer et al., 2010).

Species	**Age (in years)**	**Livestock units**
Sheep, Goat	< 1	0.05
	$\geqq 1$	0.10
Cattle	< 0.5	0.30
	$0.5 - 2$	0.60
	> 2	1.00
Horse	< 3	0.70
	$\geqq 3$	1.00
Pony		0.70

2.3. Satellite data

We used the complete Copernicus Sentinel-2 A and B data acquisitions for Germany (67 tiles) of 2017 and 2018 with 20 m x 20 m resolution as predictor variables. Although Sentinel-2 data with 10 m x 10 m resolution are potentially beneficial for the assessment of management on small grassland parcels, we used the 20 m x 20 m product including more spectral bands. We expected a high relevance of short-wave infrared (SWIR) bands 11 and 12 for the assessment of grassland management since these bands relate to numerous vegetation properties affected by management practices (Jacquemoud et al., 2009; Jacques et al., 2014; Jenal et al., 2021). Data were downloaded as top-of-atmosphere Level 1C products from the Copernicus Open Access Hub at scihub.copernicus.eu. Subsequently, Level 1C products were atmospherically, terrain and cirrus corrected with the software Sen2Cor (version 2.8.0; Mueller-Wilm et al., 2019), resulting in bottom-of-atmosphere Level 2A products. Scene classification information, retrieved from the Sen2Cor process and available for each pixel of the respective satellite image, was used to discard all observations not taken under clear sky conditions (Lange et al., 2017). Observations, consisting of measurements in nine spectral bands (with radiometric resolution according to the ESA Sentinel-2 User Guide, see sentinel.esa.int), were stored in a data cube, containing 145.572 chunks with 100 x 100 pixels each, to speed up further calculations.

Satellite observation dates differ among regions. Predictions with machine learning algorithms require data with predictors respective to the predictors of the training dataset. Thus, we aligned the data by averaging observations within 16 composite periods according to Table 3, resulting in the same, not equidistant, time steps for each time series.

Composite periods were chosen according to their expected importance with respect to grassland management and number of valid observations. During wintertime, low sun angles, potential snow cover and weather effects limit the number of valid observations. However, we did not completely discard wintertime observations as plant species composition differences may be visible, thus a large period of one to two months was chosen. Satellite observations within the vegetation period were expected to be highly relevant for the assessment of management practices as most management activities occur in this period. However, cloud cover hampers satellite observations frequently during spring and autumn in Germany. Thus, composite periods within spring and autumn were chosen with a length of one month. We chose a period of two weeks during summer as we expect a high number of valid satellite observations. Empirical experiments revealed an increasing number of data gaps when reducing the period length more. An overview about the number of clear-sky observations per composite period is given in the Supplementary Material (Table S.1).

Subsequently, missing values were replaced by interpolating the composite time series per pixel linearly. In 2017, more than 95% of composite time series contain less than 7 interpolated values, while one third contain three and one third contain less than 3 interpolated values (see Supplementary Material, Figure S.1). Data availability increased with the start of Sentinel-2B in Spring 2017. Consequently, the number of interpolated composite periods decreased in 2018, with 99% of composite time series containing less than 4 interpolated values, one third containing one and 36% containing no interpolated values.

Table 3: Definition of composite periods: list of averaged satellite scenes per composite period.

Period	**Name**	**Satellite scenes**
1	Jan/Feb	all of January and February
2	Mar	all of March
3	Apr	all of April
4,6,8,10,12	$Month_1$	first half of months May to September, respectively
5,7,9,11,13	$Month_2$	second half of months May to September, respectively
14	Oct	all of October
15	Nov	all of November
16	Dec	all of December

3. Methods

The procedure for LUI derivation was implemented in the statistical software R (version 3.6.2; R Core Team, 2020) and divided into six main parts: First, data were acquired and pre-processed as described in Section 2. Second, land management data were classified, combined with respective satellite data and sampled to generate training and validation datasets. Third, machine learning methods were used to model LUI components based on which a LUI index was computed. Fourth, results and underlying functional relationships were evaluated. Fifth, the models' areas of applicability were delineated. Finally, we applied the models to satellite data covering all grasslands in Germany.

3.1. Land management data classification

The continuous value of grazing intensity and the ordinal nature of mowing and fertilisation event number would allow for a regression approach. Since mowing frequency and fertilisation are count data, a discrete regression or ordinal classification approach would have been required to deal with these data. Although studies doing discrete regression or ordinal classification with neural networks exist (Cheng et al., 2008; Cao et al., 2020), common frameworks (keras, tensorflow) do, to our knowledge, not include the required algorithms. Thus, implementing such algorithms would require extensive programming efforts and the approach was hence dismissed for the current study. Further, high variability within management information hampers regression approaches. Although information on grazing intensity is given per field and certain time periods, livestock is not evenly distributed throughout the fields and the distribution varies over time. Fertilisation can be carried out with mineral or organic fertilisers. Quantifying the amount of mineral fertiliser is fairly straight forward, whereas liquid manure contains a varying amount of ingredients, such as nitrate, and thus documentation by farmers is hampered. Livestock can be an additional fertiliser source. Further, the effect of fertilisation on the vegetation differs depending on soil, climate and water availability (Gibson, D. J., 2009). An aggregation to classes mitigates this variability and results in more robust classification models. Based on these observations we conclude that the use of regression approaches is technically feasible but limited for practical reasons in the present study. Consequently, we reduced the problem's complexity by aggregating our target variables into classes and aimed for a classification approach instead of a regression. Thus, grazing intensity was used to derive four grazing intensity classes according to Table 4. Mowing count was respectively aggregated into six (zero to five mowing events) and fertilisation information into two classes (fertilised and not fertilised). An overview about the resulting data class distribution is given in Table 5.

Table 4: Definition of grazing classes (G) by grazing intensity (G_I) given in livestock units per ha and day.

G	$\boldsymbol{G_I}$
0	0
1	$0 < G_I \leqq 0.33$
2	$0.33 < G_I \leqq 0.88$
3	$G_I > 0.88$

3.2. Data sampling

Land management data and respective Sentinel-2 composite time series of 2017 and 2018 were combined using the satellite pixels covering the full plot size, namely the pixel with center position nearest to the respective center of each Biodiversity Exploratories plot and the eight adjacent pixels. Data combination was done separately for the years 2017 and 2018, resulting in two respective datasets. Data from 2017 and 2018 were treated separately as weather may influence plant phenology (e.g. the timing of green-up and senescence), the timing of management practices and thus the temporal evolution of the remote sensing signal differently between years. Time series

Table 5: Grassland management class distribution overview for data of 2017 and 2018 showing the number of plots assigned to a certain management class, namely the four grazing classes (G), six mowing count classes (M) and two fertilisation classes (F).

		2017				2018			
Class		**ALB**	**HAI**	**SCH**	**Overall**	**ALB**	**HAI**	**SCH**	**Overall**
G	0	22	11	18	51	21	13	15	49
	1	17	27	8	52	17	27	10	49
	2	9	8	15	32	9	7	18	34
	3	2	4	9	15	3	3	7	13
M	0	18	17	17	52	15	22	22	59
	1	8	22	22	52	12	22	14	48
	2	14	10	11	35	11	6	14	31
	3	9	1	0	10	11	0	0	11
	4	0	0	0	0	1	0	0	1
	5	1	0	0	1	0	0	0	0
F	0	25	27	45	97	26	26	48	100
	1	25	23	5	53	24	24	2	50

examples are shown in Figure 2 with respective management activities, exhibiting e.g. rapid decreases in NDVI after mowing events and a higher variability within the NDVI time series while grazed.

Plots within these datasets were divided into training and validation plots with conditioned latin hypercube sampling using the R-package clhs (Roudier et al., 2020) to preserve the high variance of grassland usage information. Thus, all pixels of a plot were used only for training or for validation, respectively, to preserve the intra-plot variance of satellite observations and to ensure independence of training and validation sets. We tested different training and validation set allocations (30%/70%, 50%/50% and 70%/30% training and validation set size, respectively). Although we found reasonable results for all allocations, higher training set sizes mainly resulted in higher accuracies. Consequently, 70% of the plots were used for training ($n_{plots} = 105$, $n_{pixels} = 945$) and 30% for validation ($n_{plots} = 45$, $n_{pixels} = 405$). Predictors, namely the 144 combinations of Sentinel-2 bands ($n_{bands} = 9$) and composite periods ($n_{periods} = 16$), were scaled to $\mu = 0$ and $\sigma = 1$ to normalise the weight of each variable before training the machine learning models.

3.3. Classification methods

Deep learning became increasingly important in various applications over the last decade due to their ability to model complex behaviors. In remote sensing, especially convolutional neural networks (CNN) and recurrent neural networks, and their special form long short-term memory networks (LSTM), are of use for image analysis and classification (Ienco et al., 2017; Kussul

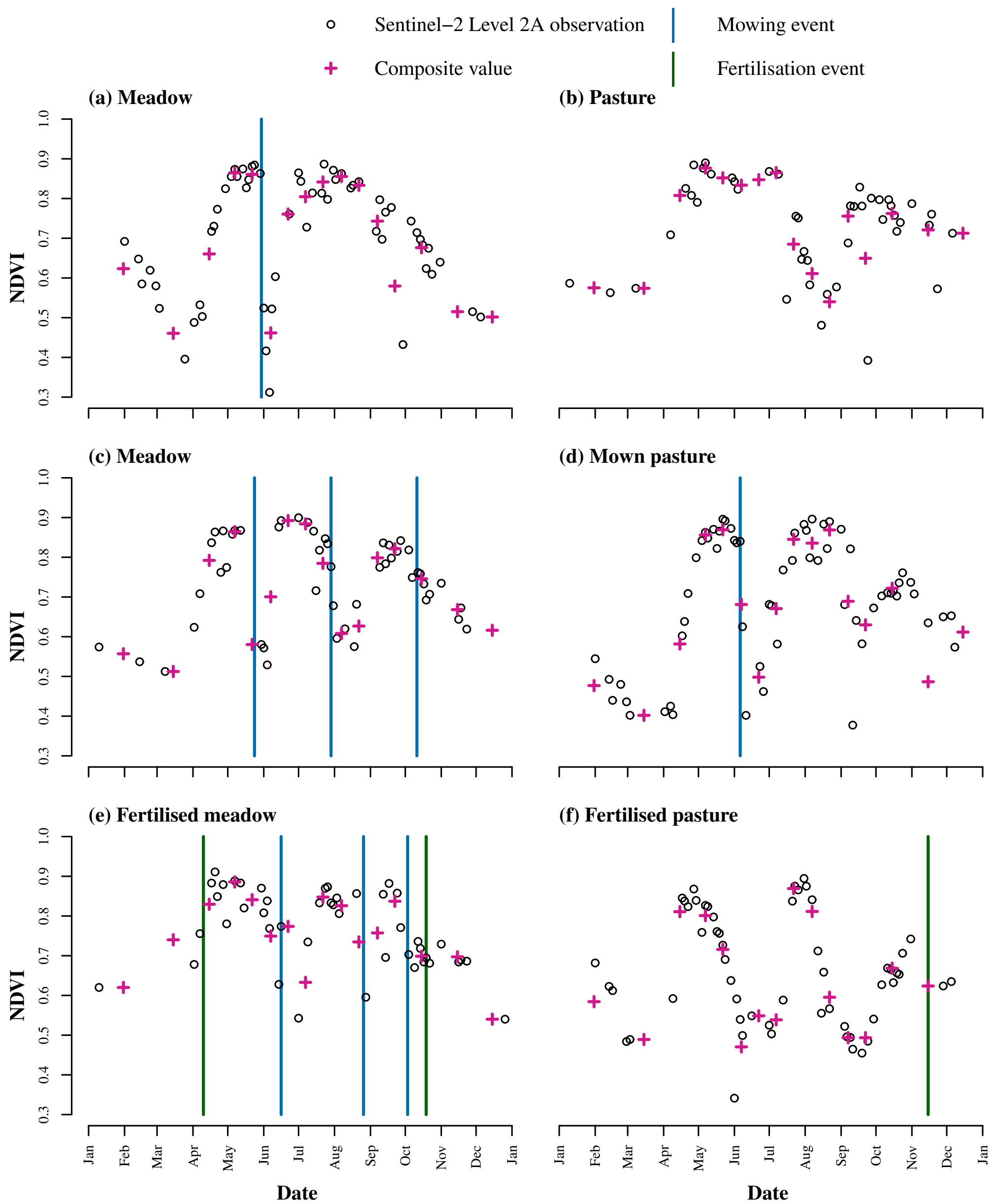

Figure 2: Normalised difference vegetation index (NDVI) time series of 2018 of center pixels in six plots (a-f) within the Biodiversity Exploratories. Pastures (b, d, f) are and meadows (a, c, e) are not grazed.

et al., 2017; Rußwurm and Körner, 2017). Here, we used a CNN to predict three LUI components according to the classes defined in section 2.2. Note that we trained the algorithm pixel-wise instead of the common image-wise application to get pixel-wise predictions. The convolution itself is thus performed on temporal and spectral instead of spatial patterns. The CNN was implemented by using the R package keras (Falbel et al., 2020). The structure of the constructed network is shown in Figure 3. It was trained in 50 epochs with the RMSProp optimiser.

We compared the CNN approach to the state-of-the-art method Random-Forest (RF; Breiman, 2001). RF is often used in remote sensing studies and especially in land cover classification (Rodriguez-Galiano et al., 2012; Preidl et al., 2020). Here, we used the R-package randomForest (Liaw and Wiener, 2002) to train a RF model on same dataset as used in the CNN approach. A McNemar-test (McNemar, 1947) with Yates correction (Yates, 1934) was performed to test for significant differences between CNN and RF classifications. Additionally, overall accuracy of both methods was compared. Parameters of RF were chosen based on suggestions from literature (Svetnik et al., 2003): m_{try} defaults to values in relation to the number of predictors (p), usually $\sqrt{p}$ for classification or $p/3$ for regression. We chose $m_{try} = p/3$ (resulting in $m_{try} = 48$) based on empirical experiments. The number of trees (n_{tree}) usually varies, depending on the number of samples, between 50 and 500. Here, we chose $n_{tree} = 128$ based on empirical experiments.

Model generation and training was repeated 100 times to account for the randomness in the training procedure, evaluate the performance on the data and to select the best model for subsequent variable importance assessment, transferability evaluation and Germany-wide application.

3.4. Land-use intensity index

LUI was inferred applying the index (here: LUI_{region}) proposed by Blüthgen et al. (2012), based on the underlying components grazing class (G), mowing count (M) and fertilisation (F). Pixel values of these components were divided by their mean study area (regional) value ($\overline{G}_{region}$, $\overline{M}_{region}$ and $\overline{F}_{region}$). Subsequently, the results were summed up to obtain the dimensionless non-negative value LUI_{region} (see Equation 1). LUI_{region} estimations based on LUI component predictions or ground observations are further referred to as $LUI^{satellite}_{region}$ or LUI^{ground}_{region}, respectively.

$$LUI_{region} = \frac{G}{\overline{G}_{region}} + \frac{M}{\overline{M}_{region}} + \frac{F}{\overline{F}_{region}} \qquad (1)$$

3.5. Validation of land-use intensity and its components

Validation was done for LUI component models as well as for LUI quantification. LUI component models were validated using the classification's overall accuracy (OA), calculated by dividing the total number of correctly classified pixels by the total number of pixels (Congalton, 1991). Further, we calculated the precision (user accuracy), recall (producer accuracy) and their harmonic mean, namely the F1-score, as measures of classification performance to account for class imbalances (Tharwat, 2020). These measures are calculated for each class. Additionally, we calculated their sample size weighted average across all classes as total measure for each model. Validation of $LUI^{satellite}_{region}$, calculated by using the predictions of the CNN and

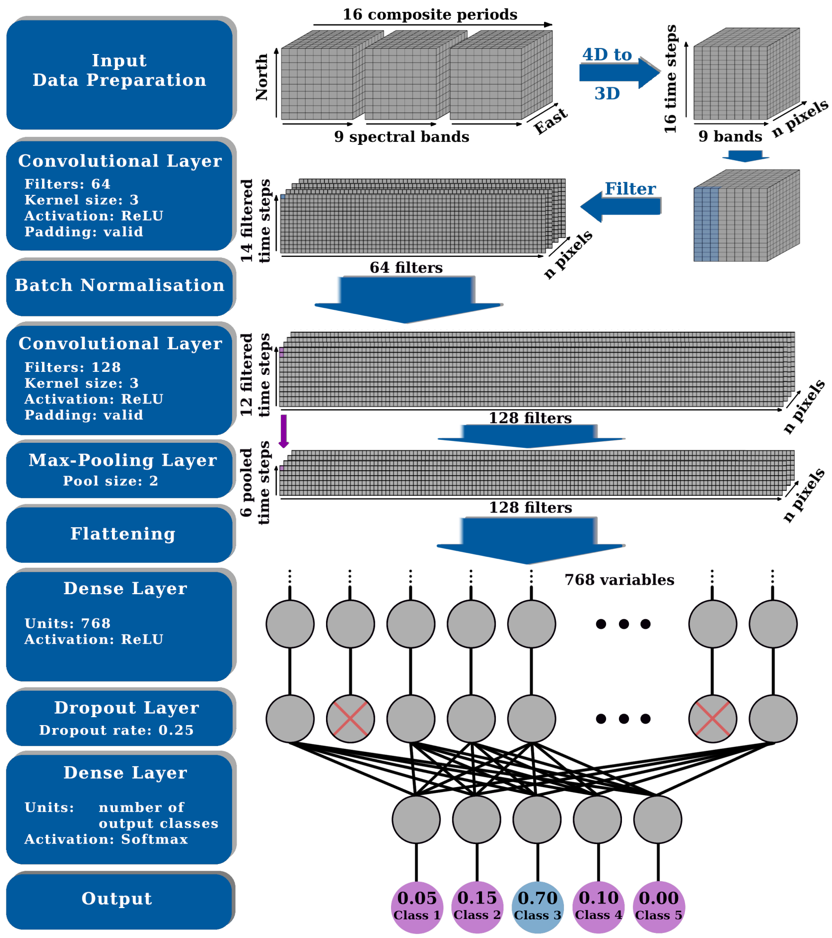

Figure 3: CNN structure: The input was reshaped such that 16 composite periods and nine bands per Sentinel-2 pixel serve as input for the net, which consists of two convolutional layers, with 64 and 128 filters of kernel-size three and valid padding, respectively, a batch normalisation layer in between, a (max-)pooling layer of size two and a subsequent dense neural network with an input layer consisting of 768 neurons with a rectified linear unit (ReLU) activation function, a dropout layer with a dropout rate of 0.25 and an output layer with a softmax activation function and as many neurons as the number of classes in the target variable.

RF model with highest OA for each component, was done for each observatory (ALB, HAI and SCH) respectively and for the union of these three regions (called "*Overall*" in the results section). We utilised the provided ground management data of LUI components to calculate the ground truth LUI^{ground}_{region} (see Equation 1). Subsequently, we calculated the squared Pearson correlation coefficient (r^2) between $LUI^{satellite}_{region}$ and LUI^{ground}_{region} to evaluate the relationship between the remote sensing based LUI quantification and LUI estimated from ground level observations.

3.6. Variable importance evaluation

We performed a variable importance evaluation to assess the relevance of spectral bands and composite periods for the prediction of LUI components. Supervised machine learning algorithms are presented with example inputs, namely our composite time series of spectral bands, and desired outputs, the LUI components, to learn general patterns. Analysing these patterns is essential to substantiate subsequent predictions by supporting the understanding of functional relationships of the process being modelled and by unveiling misconceptions or data biases. However, machine learning models are usually complex and the underlying patterns lack interpretability. Commonly, this problem is addressed by calculating the importance of input variables, here the 144 combinations of nine spectral bands and 16 composite periods, also referred to as features, for the model output. Global feature importance measures are calculated based on the output of the whole feature space, e.g. a performance change resulting from modifying a single input variable, whereas local feature importance measures base on the variable contributions to a single prediction. Global feature importance measures, e.g. the mean decrease in impurity used in RF, may be inconsistent between different models or biased depending on model structure, except for permutation approaches such as the mean decrease in performance after permuting the value of the feature of interest (Lundberg et al., 2019, 2020). However, the latter are highly computationally intensive and may become impractical when explaining predictions from large datasets. Lundberg and Lee (2017) proposed the local approach Shapley Additive Explanations (SHAP), calculating each variables' contribution to a certain prediction. These contribution values are additive and the (normalised) sum of their absolute values represents global variable importance for a given feature space. SHAP values provide the benefit of being consistent with human intuition, thus enabling an accurate analysis of the model behavior (Lundberg and Lee, 2017). We used SHAP's GNU R port shapper (Maksymiuk et al., 2020) to calculate each feature's contribution and to evaluate the importance of thematically meaningful feature subsets, namely satellite bands and composite periods.

3.7. Transferability evaluation

Spatial transferability is required when aiming for large scale applications, enabling the extrapolation into regions not used for training. The evaluation includes i) a spatial 3-fold cross-validation (CV) and ii) the delineation of the models' areas of applicability (AOA) that is defined as the area where training based relationships inherent to a model apply with the respective estimated CV performance (Meyer and Pebesma, 2021).

The spatial 3-fold CV was performed by training the CNN models on spatial subsets of the composite time series. Spatial subsets were built by leaving out all data points of each observatory once. Consequently, each subset contained data points of two observatories only. Models trained on each of these subsets were subsequently validated by calculating the OA using only data points of the left out observatory (referred to as OA^*), i.e. a geographically distinctly separated region, to assess the models' robustness against the lack of certain training areas. Further, we calculated accuracy loss from subtracting OA^* from OA of models where training and validation were done with the dataset containing all observatories.

We delineated the models' AOA following the procedure proposed by Meyer and Pebesma (2021). They define the AOA by first calculating the dissimilarity of new data to the feature space of the training data considering the models' feature importances and second, applying a threshold to these dissimilarity index (DI) values. The threshold is calculated as the outlier-removed maximum DI of the spatially 3-fold cross-validated training data. Here, we used the R-Package CAST (Meyer et al., 2021) to delineate the AOA, resulting in maps highlighting grassland pixels within and outside the CNN models' AOA. Predictions within a model's AOA are considered reliable. The LUI AOA (AOA_{LUI}) was inferred by combining the AOA of grazing intensity (AOA_G), mowing count (AOA_M) and fertilisation (AOA_F). We chose a conservative approach such that if a pixel was outside of only one LUI component's AOA, it was labelled as outside AOA_{LUI}.

3.8. Germany-wide extrapolation of land-use intensity and its components

Germany-wide maps were generated by applying the LUI component models to each Sentinel-2 pixel covering a grassland object of the ATKIS-DLM. Subsequently, $LUI_{Germany}$ was computed for these areas according to Equation 1.

4. Results

4.1. Validation of land-use intensity and its components

OA vary between observation years, with grazing intensity and mowing models showing higher OA for 2018, whereas fertilisation models show a slightly higher OA for 2017 compared to 2018 (Figure 4). The highest OA of CNN for 2017 and 2018 are 59% and 66% for grazing intensity models, 62% and 68% for mowing models, and 85% for fertilisation models. OA of LUI component models are slightly higher for CNN compared to RF. The difference is significant for mowing and fertilisation depiction (McNemar test, $p < 0.002$) and not significant for grazing intensity models. Median OA is higher for CNN than for RF models, except for grazing intensity depiction of 2018, where maximum OA is achieved with CNN. CNN models show a higher variance in OA than RF models.

Class statistics (Table 6) show a balanced distribution of total weighted F1-score, precision and recall. Total weighted F1-score is slightly lower than OA. However, class imbalances are visible and class-based F1-score, precision and recall in general decrease with decreasing sample size, except for grazing intensity classes 0 and 1 of 2018. Grazing intensity class 2 of 2017 and mowing class 2 of 2018 exhibit low F1-score, precision and recall, although sample sizes are larger than their respective class 3. We observe a low precision for grazing intensity class 2 in 2017 and class 3 in 2018. In 2017, the confusion matrix of the model with highest OA (see Supplementary Material, Table S.2) reveals an overestimation in pixels with low grazing intensity (class 0

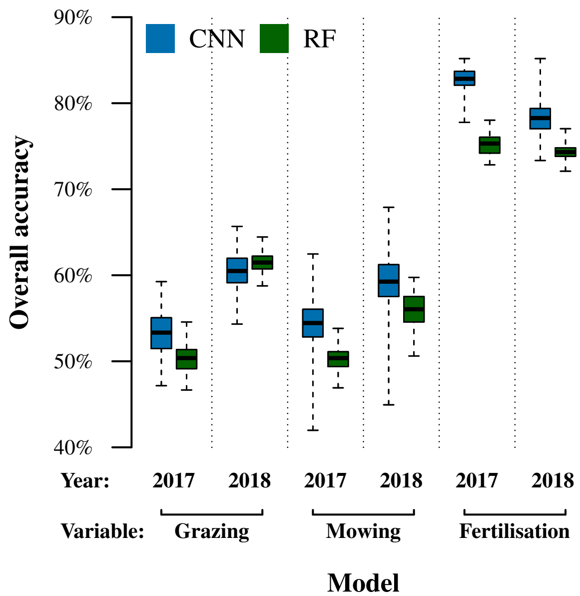

Figure 4: Overall accuracies of 100 runs with CNN, shown as blue boxes, and RF models, shown as green boxes, with whiskers extending to the full range of accuracies. The models were trained and validated with grazing, mowing or fertilisation data of the years 2017 and 2018 (70% training, 30% validation).

and 1) and an underestimation in pixels with grazing intensity class 2. In 2018, the confusion matrix shows an underestimation of grazing intensity class 3 and mowing class 2. Mowing class 2 in 2017 and fertilisation class 1 in 2018 exhibit substantially lower recall than precision values. In both cases, the confusion matrix also shows an overestimation of the lower classes with respectively higher sample size, resulting in a decreased recall. Further, standard deviation of mowing model accuracy is high for classes with low sample sizes.

The overall r^2 between $LUI_{region}^{satellite}$ and LUI_{region}^{ground} is slightly higher for CNN, with 0.82, compared to RF, with 0.75 and 0.76 for 2017 and 2018, respectively (Figure 5). r^2 of observatory SCH in year 2018 is much lower than for the observatories ALB and HAI, as well as SCH in 2017. Root mean squared errors (RMSE) of SCH are higher compared to ALB and HAI, whereas median absolute errors (MedAE) are lower, except for CNN in 2017, where HAI shows the lowest MedAE (Table 7).

4.2. Variable importance of CNN models

Variable importance evaluation (see Figure 6) reveals an at least moderate importance of all spectral bands of Sentinel-2 for one or the other component model. Composite periods of winter months November, December, January and February show a low importance for all models. The most important composite periods are late spring and early summer, followed by late summer and autumn.

Table 6: Class statistics of 100 runs with CNN for data of 2017 and 2018 showing mean (μ) and standard deviation (σ) of user (precision) and producer accuracy (recall), F1 score and number of samples (n_s) for each management class, namely the four grazing classes (G), six mowing count classes (M) and two fertilisation classes (F), as well as the total weighted F1 score, user and producer accuracy.

			F1-score		**Precision**		**Recall**		
Year	**Variable**	**Class**	μ	σ	μ	σ	μ	σ	n_s
2017	**G**	0	59%	3.9%	66%	7.6%	53%	3.8%	108
		1	66%	2.5%	72%	5.0%	62%	3.5%	171
		2	17%	5.7%	13%	4.8%	28%	8.7%	108
		3	41%	6.1%	45%	8.0%	37%	7.8%	18
		total	**50%**	**2.6%**	**53%**	**2.5%**	**49%**	**3.3%**	**405**
	M	0	65%	4.1%	61%	6.2%	71%	4.5%	162
		1	47%	2.4%	45%	3.5%	49%	5.3%	144
		2	52%	5.2%	67%	9.9%	43%	3.8%	72
		3	40%	16.3%	36%	20.7%	43%	17.4%	27
		4 & 5							0
		total	**54%**	**3.1%**	**54%**	**2.9%**	**57%**	**3.1%**	**405**
	F	0	87%	0.9%	90%	1.6%	85%	1.6%	270
		1	72%	2.7%	68%	4.3%	77%	2.3%	135
		total	**82%**	**1.4%**	**83%**	**1.3%**	**82%**	**1.3%**	**405**
2018	**G**	0	68%	3.7%	69%	6.9%	68%	4.7%	126
		1	66%	2.6%	64%	4.0%	70%	4.4%	162
		2	48%	4.0%	56%	7.2%	42%	2.9%	81
		3	34%	5.0%	26%	4.1%	53%	16.7%	36
		total	**60%**	**2.2%**	**60%**	**2.2%**	**62%**	**2.6%**	**405**
	M	0	73%	3.1%	74%	5.4%	72%	5.0%	180
		1	52%	5.5%	53%	8.8%	52%	3.9%	126
		2	24%	5.1%	22%	6.6%	27%	6.4%	63
		3	75%	7.4%	76%	11.4%	75%	8.5%	36
		4 & 5							0
		total	**59%**	**2.9%**	**59%**	**3.1%**	**59**	**2.9%**	**405**
	F	0	84%	1.7%	80%	3.4%	89%	1.8%	297
		1	65%	2.9%	74%	5.8%	58%	3.5%	108
		total	**79%**	**1.8%**	**78%**	**2.0%**	**81%**	**1.6%**	**405**

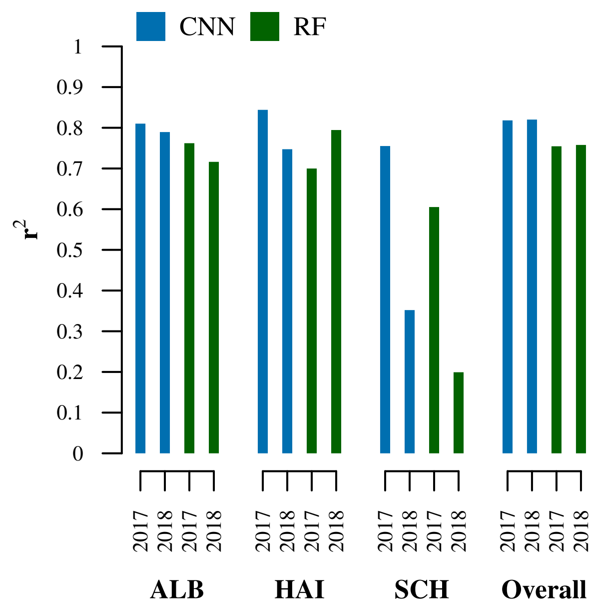

Figure 5: Squared Pearson correlation coefficients (r^2) between LUI_{region}^{ground} and $LUI_{region}^{satellite}$ based on classifications of CNN, shown as blue bars, and RF models, shown as green bars, and using data of each observatory (ALB: *Schwäbische Alb*, HAI: *Hainich* and SCH: *Schorfheide*), as well as the union of all observatories (*Overall*). All correlations are significant ($p < 10^{-20}$).

Table 7: Median absolute error (MedAE) and root mean square error (RMSE) between LUI_{region}^{ground} and $LUI_{region}^{satellite}$ based on classifications of CNN and RF models and using data of each observatory (ALB: *Schwäbische Alb*, HAI: *Hainich* and SCH: *Schorfheide*), as well as the union of all observatories (*Overall*).

			ALB	**HAI**	**SCH**	**Overall**
CNN	**2017**	MedAE	0.21	0.10	0.15	0.18
		RMSE	0.66	0.68	1.87	1.21
	2018	MedAE	0.10	0.15	0.09	0.11
		RMSE	0.73	0.91	5.03	2.97
RF	**2017**	MedAE	0.16	0.32	0.12	0.19
		RMSE	0.77	1.01	2.63	1.69
	2018	MedAE	0.25	0.26	0.07	0.13
		RMSE	0.84	0.86	5.01	2.97

The grazing intensity model is strongly influenced by the two SWIR bands 11 and 12 and by the red-edge band 6 (in 2017) and 5 (in 2018). In 2018, the two near-infrared (NIR) bands 7 and 8a show a high contribution, which is not reflected in the feature importances of 2017. However, the green band 3 shows a higher contribution in 2017. Grazing intensity models' composite period importance shows highest values from April to June and September to October.

In 2017, the mowing count model is most strongly influenced by the SWIR bands 11 and 12 and NIR bands 7 and 8a. In 2018, the red band 4, SWIR band 11 and NIR bands 5 and 7 show the highest contribution. Composite periods in April to May and August to September stem the major contribution.

The fertilisation model is most strongly influenced by band 11. All other bands, except band 6, contribute moderately to the predictions in 2017. In

2018, red-edge bands 5, 6, SWIR band 12 and blue band 2 exhibit a moderate contribution. The composite periods from April to first half of July exhibit the highest variable importance.

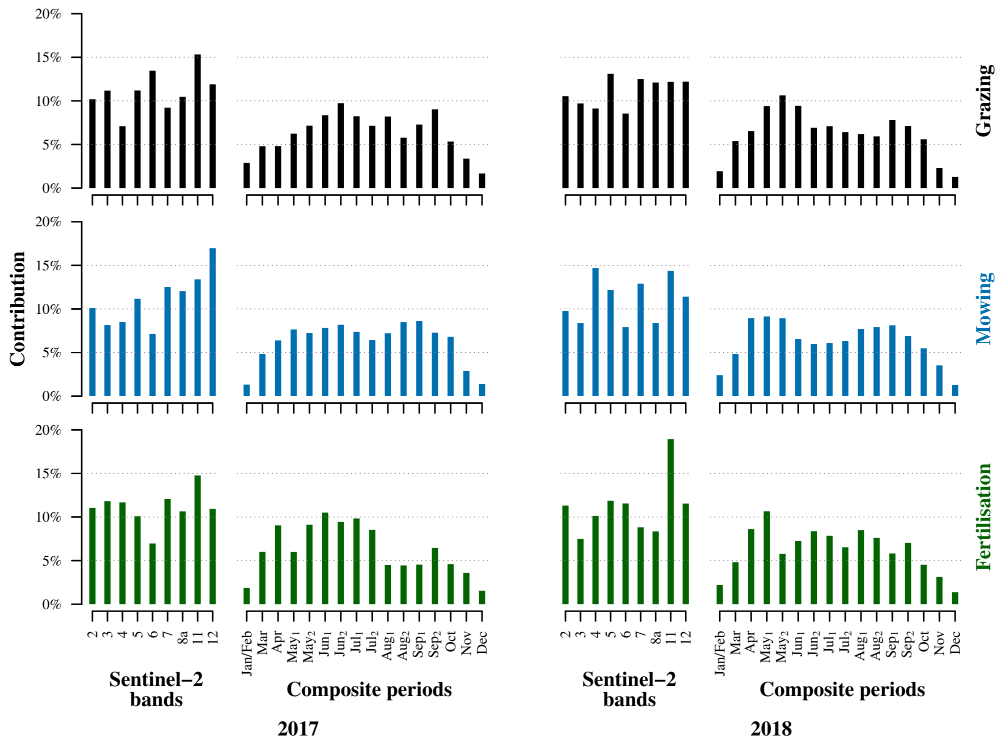

Figure 6: Feature importance: Summary of CNN feature contribution on predictions using data from 2017 (left) and 2018 (right) generated by using the mean absolute SHAP values of features grouped by satellite band (first and third column) or composite period (second and fourth column) for each variable (by row from top to bottom). Mean absolute feature contribution is shown on the ordinate, the abscissa shows the respective bands and composite periods.

4.3. Spatial transferability of land-use intensity component models

The spatial 3-fold CV reveals varying performance depending on observatory, year and LUI component (see Figure 7). The grazing model is robust against leaving out the observatory HAI for training. Leaving out the observatory ALB leads to high accuracy loss in 2018, but not in 2017. Accuracy decreases more than 20% when leaving out the observatory SCH. The mowing model shows a low CV accuracy loss in 2017 and a high accuracy loss in 2018 for all observatories. The fertilisation model is robust against leaving out ALB in 2017 and SCH in 2018. Consequently, all training areas are valuable and leaving out one of them reduces the accuracy of at least one of the models.

The delineation of AOA reveals varying spatial transferability for different years and LUI components. Out of a total number of 146.4 million grassland pixels, we found 29.3% outside AOA_{LUI} in 2017. In 2018, the number of pixels outside AOA_{LUI} decreased to 6.9%. The grazing intensity model is not applicable for 19.7% and 4.1% of pixels in 2017 and 2018, respectively. The mowing model is not applicable for 17.4% and 5.4% of pixels, respectively. The fertilisation model shows the highest amount of pixels outside AOA, with 27.3% and 5.9% in 2017 and 2018, respectively. The spatial distribution of areas outside AOA_{LUI} varies between our regions (a)-(d) (see Figure 8). We observe the lowest amount of pixels outside AOA_{LUI} in region (b) and (c), a moderate amount in region (a) and a high amount in region (d). Pixels outside AOA_{LUI} show spatial small scale patterns resembling fields or parcels with respective management.

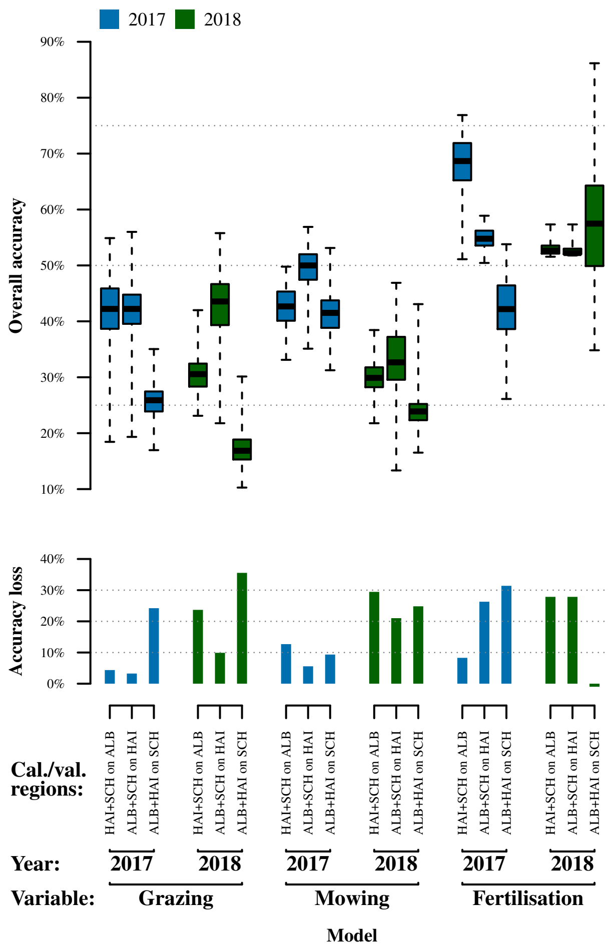

Figure 7: Spatial 3-fold cross-validation: CNN models, 100 per fold, were trained and validated on different spatial subsets for the observation years 2017 (blue) and 2018 (green). Training was done with data from two regions, e.g. *Schwäbische Alb* (ALB) and *Hainich* (HAI), and validation with the remaining third region, e.g. *Schorfheide* (SCH). Boxplots show the OA^* of the resulting 100 models per fold. The bars show the accuracy loss (or, in one case, gain) of the CV model with the highest OA^* compared to the highest OA of the models applied on data of all observatories.

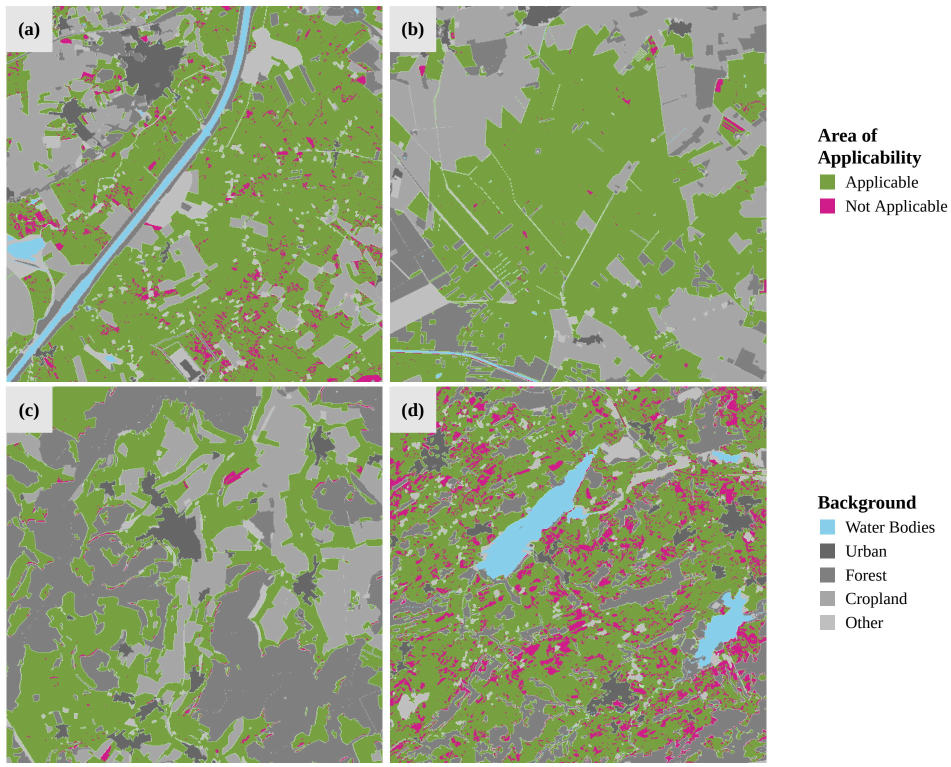

Figure 8: The maps of the four regions (a)-(d) show land cover information with colours respective to the bottom-right legend (© GeoBasis-DE / Bundesamt für Kartographie und Geodäsie, 2015), whereas grassland pixels are overlaid by AOA_{LUI} of 2018, with areas inside AOA_{LUI} in green and areas outside AOA_{LUI} in magenta colours.

4.4. Germany-wide extrapolation of land-use intensity and its components

The dataset of national extent prototype results was made available online (Lange et al., 2021) and provided in a web service (www.ufz.de/land-use-intensity) enabling visual exploration. The maps within the web service show regional patterns, such as a high land-use intensity in the Alpine Foothills with a high grazing intensity and a high share of fertilised grasslands, a high grazing intensity in north-western Germany in vicinity to the Northern Sea, and divers regional patterns of mowing counts. Further, the share of pixels outside AOA is highest for the Alpine regions and, in 2017, for north-western Germany. Here, we show results from the four regions (a)-(d) of 2018, where data availability was high and thus a larger share of pixels within the AOA.

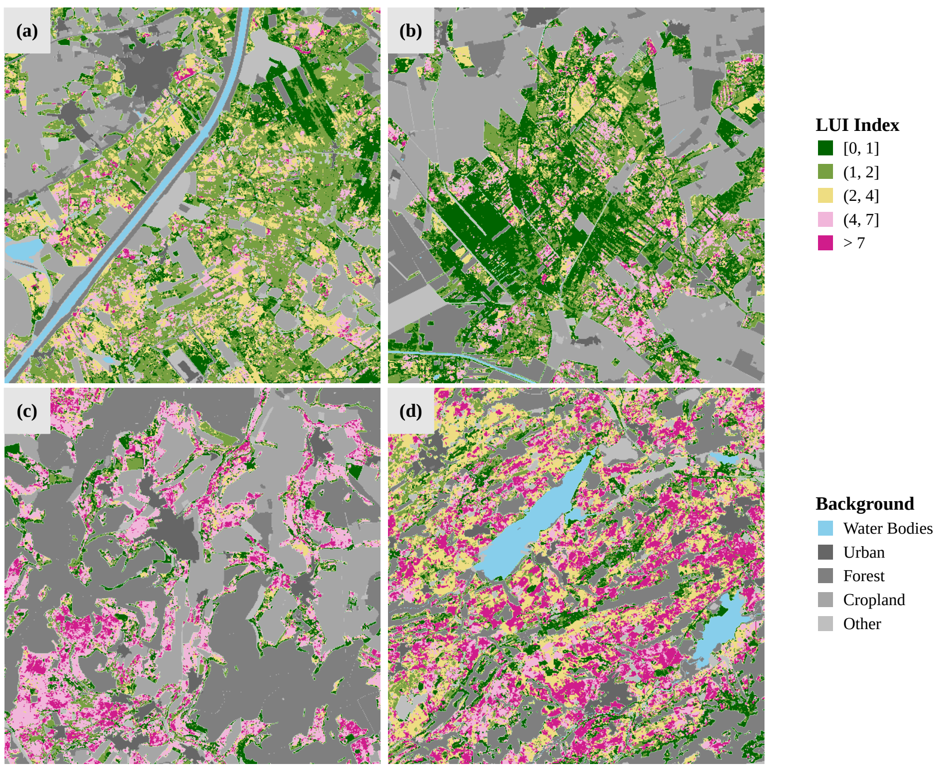

Figure 9: The maps of the four regions (a)-(d) show land cover information with colours respective to the bottom-right legend (© GeoBasis-DE / Bundesamt für Kartographie und Geodäsie, 2015), whereas grassland pixels are overlaid by their $LUI^{satellite}_{Germany}$ of 2018 with colours ranging from green (extensive use) to magenta (intensive use).

Figure 9 reveals large scale grassland management patterns characteristic for each demonstration region most likely due to different abiotic conditions, but also high spatial variations of $LUI^{satellite}_{Germany}$ within (a)-(d). Figures 9a and 9b exhibit a heterogeneous small scale colouring within groups of 10 x 10 pixels.

Figure 10 illustrates the high variability of grassland LUI and its underlying components in region (d). Please note that for mowing class 4 only one training data point and no validation data was available. However, we kept the value within the results as the predictions are not necessarily wrong. Regions with high number of mowing events exhibit lower grazing intensity and vice versa, e.g. in the southwestern part, suggesting a contradictory relation between both management types. Highest $LUI^{satellite}_{Germany}$ is observed in areas with detected fertiliser application and either with three or more mowing events or with grazing intensity class two or three, e.g. in the eastern part, between village Oy-Mittelberg and lake Grüntensee. Lowest $LUI^{satellite}_{Germany}$ are found in areas with steep slopes, e.g. Burgkranzegger Horn, ranging from south-west to north-east between the lakes, where neither mowing nor fertilisation is detected.

Figure 11 shows the density distribution of the spatial patterns visible in Figure 9 and Figure 10. Region (b) exhibits lowest overall $LUI^{satellite}_{Germany}$ due to a low grazing intensity, zero to one mowing events in 2018 and a low number of fertilised pixels. Nature reserve core zones are located in the southwestern part of the extracted extent and exhibit lowest $LUI^{satellite}_{Germany}$ (see Figure 9b).

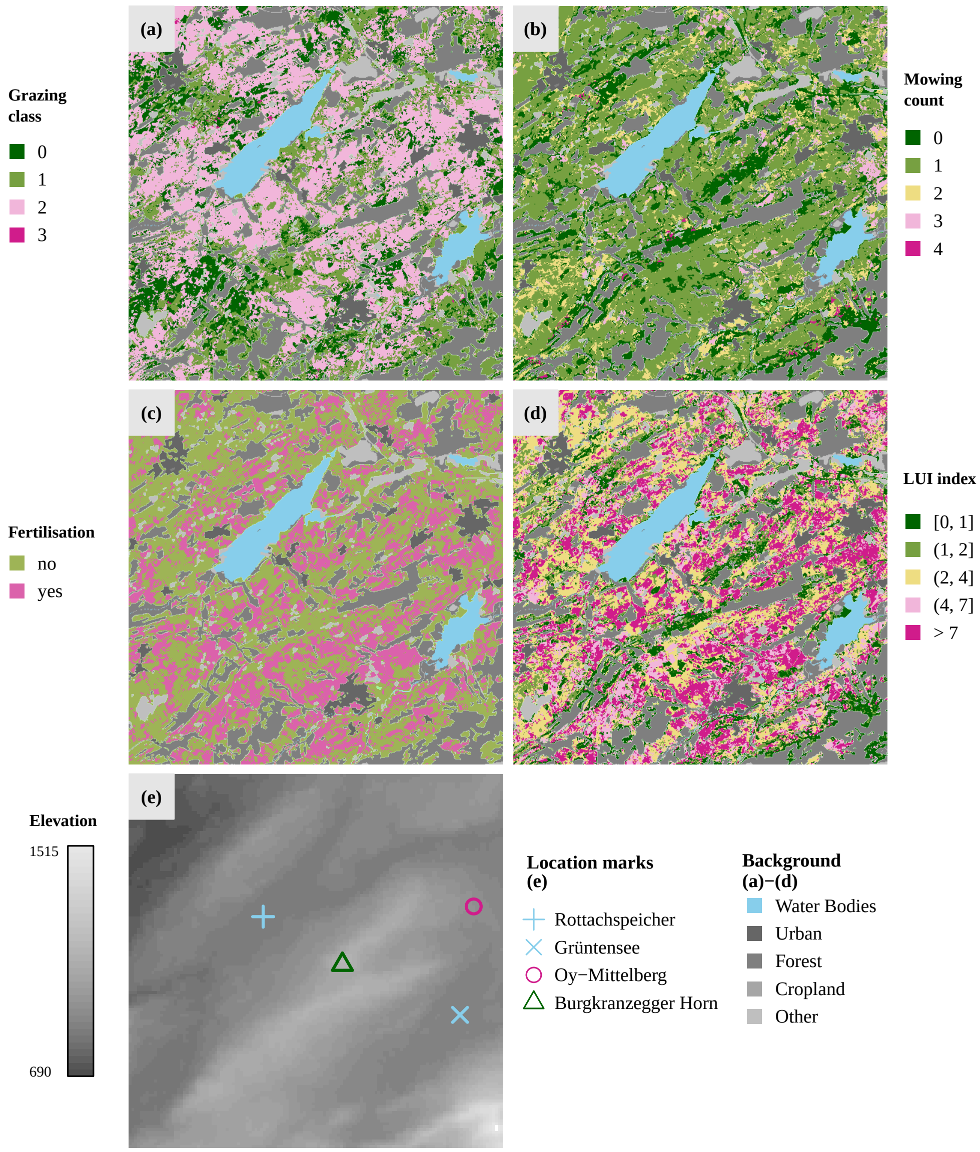

Figure 10: Subsets of national extent maps: (a)-(d) grazing classes, mowing counts, fertilisation and $LUI^{satellite}_{Germany}$, respectively, of 2018 for an area of 10 km x 10 km in the district Oberallgäu (see Figure 9d) in the federal state Bavaria. All non-grassland pixels use colours respective to the legend of background values in the right. Grazing classes range from zero (low grazing intensity, green) to three (high grazing intensity, magenta) and mowing counts from zero (green) to four (magenta). Fertilisation (no/yes) is indicated in green and light magenta, respectively. $LUI^{satellite}_{Germany}$ values are aggregated into five classes with colours ranging from green (extensive use) to magenta (intensive use). Map (e) displays the digital elevation model of the region based on NASA's Shuttle Radar Topographic Mission (SRTM; Jarvis et al., 2008) and highlights specific locations.

Region (a) shows a moderate $LUI_{Germany}^{satellite}$ with a comparably high grazing intensity, a low number of mowing events and lowest number of fertilised areas. However, a moderate number of pixels shows distinctively higher $LUI_{Germany}^{satellite}$ values, reflected in high small scale variations of respective maps (see Figure 9a). Region (c) shows a high $LUI_{Germany}^{satellite}$ based on high grazing intensity and around 70% of fertilised pixels, as well as a moderate number of mowing events. Although the shown extent is situated in a biosphere reserve, most of the pixels are in transition areas and not in core or buffer zones. Further, the histogram reveals a contrasting partition with fields either of high or of low $LUI_{Germany}^{satellite}$, which translate into spatially coherent respective areas in map 9c. Analysis of region (d) reveals highly diverse $LUI_{Germany}^{satellite}$ values based on a high grazing intensity, a high number of mowing events and around 40% of fertilised pixels.

National scale statistics of $LUI_{Germany}^{satellite}$ and its components (see Figure 12) reveal ample differences between years and classes. We observe an increase in grazing intensity in 2018, with the majority of pixels in class two compared to class one in 2017. Contrarily, the number of mowing events decreased. We detected mowing events on 66% of all grassland pixels in 2017 compared to only 38% in 2018. We found one mowing event in 45% of pixels in 2017 and in 27% in 2018, and two mowing events in 13% resp. 8%. 9% and 3% of pixels were mown three times in 2017 and 2018, and less than 0.1% and 0.05% four or more times, respectively. Fertilisation was detected on around 39% and 31% of all pixels in 2017 and 2018, respectively. $LUI_{Germany}^{satellite}$ values

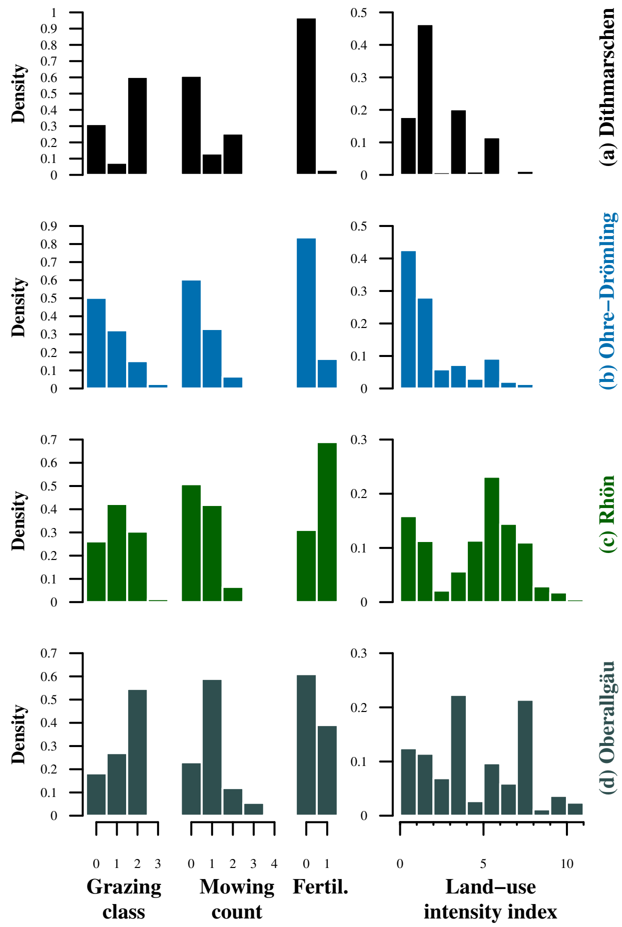

Figure 11: $LUI^{satellite}_{Germany}$ and its components (grazing class, mowing count and fertilisation) within the AOA for 2018 in grasslands in the four demonstration regions (a)-(d).

mainly remain in the range from one to ten, with the majority of pixels with values less than three, a moderate amount with values between three and six, a low amount between six and ten and less than 1% above ten. The amount of pixels with $LUI^{satellite}_{Germany}$ between zero and one increased from 3% in 2017 to 27% in 2018, whereas the amount of pixels with values between one and five decreased from 66.5% to 45.5%. Further, the number of pixels with $LUI^{satellite}_{Germany}$ between five and nine decreased from 29.5% to 24.5% and the number of pixels with $LUI^{satellite}_{Germany}$ above nine increased from less than 1% to 3%.

5. Discussion

5.1. Validation of land-use intensity and its components

We quantitatively estimated LUI of grasslands in three regions situated across Germany with fine spatial resolution of 20 m x 20 m with an r^2 of 0.82 and subsequently applied the methodology on national extent. Other studies inferring LUI on a similar scale exist (Gómez Giménez et al., 2017; Stumpf et al., 2020) but either focus on only one LUI component or lack overall quality measures. The CNN approach showed slightly higher OA estimating LUI components than the state-of-the-art method Random Forest with significant better classifications of mowing count and fertiliser application. The variance in 100 training repetitions was higher for CNN than for RF, potentially revealing the demand for a higher number of training data when using CNN. However, both approaches revealed reasonable results and

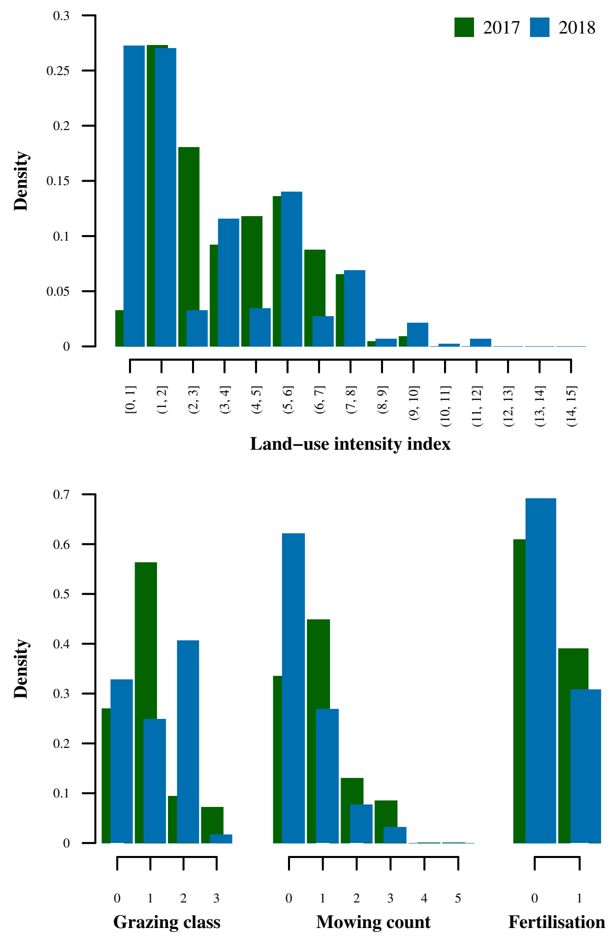

Figure 12: $LUI^{satellite}_{Germany}$ and its components (grazing class, mowing count and fertilisation) for 2017 and 2018 in grasslands of Germany within the AOA.

consequently demonstrate the feasibility of inferring LUI by combining remote sensing based quantifications of mowing frequency, grazing intensity and fertilisation.

Each LUI component may be addressed in more detail in separate studies, e.g. using discrete regression or ordinal classification. We dismissed this for the current study as i) such approaches are not readily available within common AI frameworks and ii) an aggregation to classes mitigated the high variability within the land management data. However, information is lost during this aggregation, potentially leading to an underestimation of classification model performance. Further, we did not provide uncertainty estimates. Using class probabilities instead of discrete classes when modelling LUI components allows to assess prediction uncertainty and could potentially increase the accuracy of $LUI^{satellite}_{region}$ depictions.

Mowing count prediction is usually done by counting local extrema or features in time series of optical or SAR data. However, validation data is rare. Most of these studies rely on visual interpretation of time series for validation data acquisition. Consequently, they lack proper ground truth. We found one recent study using in situ data for the validation of a SAR-based mowing count estimation (de Vroey et al., 2021). They correctly identified 56% of all parcels in terms of mowing dynamics, with grazing events as major confounding factor. Griffiths et al. (2020) did no quantitative validation due to the low sample size, but showed a validity check. They detected a total of 25% of grasslands not mown in Germany in 2016, whereas our study found 33% grasslands with no mowing events in 2017 and 62% in 2018. This discrepancy may stem from climatic differences between 2016, 2017 and 2018, as in 2018 a severe drought occurred in Germany during summer potentially leading to lower biomass production in grasslands. However, we would expect to identify at least one mowing event before the drought event. Consequently, the model seems to underestimate mowing. This is not only visible in decreased model precision and recall of mowing classes with increasing mowing counts, but also in spatial patterns in regions with intensively managed grasslands, e.g. the alpine foothills, where we would expect more than three mowing events per year. This is presumably caused by the lack of grasslands mown more than three times a year in our calibration data. Further, national scale statistics (Figure 12) show a decrease in mowing events and an increase in grazing intensity from 2017 to 2018, potentially reflecting a confounding of mowing and grazing as found in de Vroey et al. (2021). Contrarily, counting local extrema or features in time series may overestimate mowing, as extrema or features may stem from several reasons, such as weather effects, misclassification in cloud detection in satellite imagery (Lange et al., 2017), flooding, fire and land management or conversion (Griffiths et al., 2020). Our compositing approach eliminates short-term changes, such as weather effects and misclassifications in cloud detection, in the input data. Consequently, we assume a less pronounced effects of these factors on our results. However, long-term changes, such as fire, flooding and land conversion, alter vegetation vitality, structure and composition or remove vegetation completely and thus alter the overall reflectance properties captured by our models. Further, the compositing approach might also have negative impacts on our results as it might blur certain management activities, leading to an underestimation of management activities when their effect on vegetation diminishes quickly, e.g.

rapid regrowth after a mowing event. Estel et al. (2018) depicted mowing frequencies from MODIS data and did a quantitative validation by comparing predictions to visually interpreted time series. They achieved an OA of 77% and 80% in years 2009 and 2012, respectively. We found an OA of mowing count prediction of 62% and 68% in 2017 and 2018, respectively. However, OA comparison is limited due to different validation data, spatial resolution of satellite data and climatic differences between years. Kolecka et al. (2018) used Sentinel-2 data of 2017 to infer mowing frequency in northern Switzerland, although validating results with visually interpreted time series, and achieved an OA of 79.7%. They found 27% of observed pixels not mown in this fairly intensively managed area. We found 33% of grassland pixels not mown in Germany in 2017, including extensively managed areas and pastures. Taravat et al. (2019) used a field campaign to acquire validation data and SAR data for the classification of mowing practices and achieved an OA of 85.7%. The discrepancies between literature and our results may be due to different approaches, regions, managements, climates and spatial scales. Furthermore, we used *in situ* ground truth for validation and applied the methodology to a diverse landscape. Here, we did not assess the timing of mowing events as it is not included in the LUI definition proposed by Blüthgen et al. (2012). However, it is of high ecological relevance (Bernhardt-Römermann et al., 2011; Blüthgen et al., 2012; Franke et al., 2012) and may be included in future studies improving remote sensing LUI products.

Grazing intensity is usually inferred by VI time series analysis. Validation data is commonly acquired in field campaigns. Li et al. (2016) depicted four grazing intensity classes by estimating above ground biomass from Landsat-8 images of a steppe region in China with 170 samples in 2014 and achieved an OA of 57.65%. Conception and results compare well to our approach, which achieved an OA of 59% and 66% in 2017 and 2018, respectively. Ma et al. (2019) used MODIS images of 2016 and 2017 to relate vegetation indices, above ground biomass and grazing intensity for grazing areas of three herders in China. They found a power regression model of NDVI and grazing intensity with an r^2 of around 0.56. Studies on the classification of grazing intensity using more spectral information than given by vegetation indices are rare and our results suggest potential for large scale applications. However, the decrease in precision and recall of higher classes of grazing intensities in our models suggest an underestimation of grazing intensity, although patterns observed in regions (a)-(d) match our empirical knowledge. The potential underestimation of grazing intensity and mowing frequency is substantiated by the probability of each pixel to belong to the class assigned, which is mostly lower for grazing and mowing models than for the fertilisation models (see Supplementary Material, Figure S.3 and S.4).

Although hyperspectral remote sensing is used to study plant nutrient status and supply on crops (Cilia et al., 2014; Xia et al., 2016), the depiction of grassland nutrient status or fertilisation from remote sensing data is seldom addressed in literature (Pellissier et al., 2015; Sibanda et al., 2015; Hollberg and Schellberg, 2017). However, Hollberg and Schellberg (2017) simulated RapidEye data from ground-based hyperspectral data to subsequently differentiate five different fertiliser treatments in grasslands and achieved OA of up to 91%. Sibanda et al. (2015) discriminated grasses under different fertilizer treatments using ground-based hyperspectral data with an OA of 85%. Our approach achieved an OA of 85% differentiating between fertilised and unfertilised plots, supported by respective high model precision and recall and an F1-score for the detection of fertilised plots of 72% in 2017 and 65% in 2018. Consequently, this demonstrates the feasibility differentiating between fertilised and unfertilised plots by classifying optical satellite data in three geographically distinctly separated regions varying in fertilisation intensity. The combination of the national-extent extrapolation and the models' AOA is a first step towards a large-scale grassland fertilisation assessment and consequently supports LUI depiction. Further research might include extended calibration data to e.g. depict different levels of fertilisation.

Generally, the validation of LUI component models demonstrated the feasibility to quantify mowing frequency, grazing intensity and fertilisation with CNN with limitations stemming from calibration and validation data availability and from using a classification instead of a regression approach.

5.2. Variable importance analysis

The feature contribution analysis revealed reasonable functional relationships between LUI components and temporal (composite periods) as well as spectral (satellite bands) predictors. Satellite images outside the vegetation period have low impact on the predictions, whereas images at the beginning and end of the vegetation period are highly relevant. This highlights the strong relationship of grassland management and land-surface phenology, e.g. a potentially stronger increase of plant growth due to fertilisation in early spring or possibly multiple regrowth events due to mowing in early summer. However, satellite images of Germany in spring and autumn are often affected by cloud contamination. This could influence our results, e.g.

leading to an underestimation of mowing events or grazing intensity, although our compositing scheme was set-up to mitigate these effects. The inclusion of SAR data could be beneficial by providing additional predictors unaffected by cloud contamination.

Grazing intensity models' feature importance revealed a high relevance of composite periods from April to June and medium importance of periods in September to October, potentially reflecting farmers' livestock number adjustments per site as fodder quantity in grasslands is highest during the vegetation period and fodder quality peaks in spring and fall (Gilhaus and Hölzel, 2016). Mowing is carried out during spring and summer months, with first cuts in late spring (Griffiths et al., 2020). Feature contribution of composite periods underpins this usage behavior: main feature contribution stems from periods in April to May, followed by periods of August to September. Fertilisation, usually done in late winter or spring, accelerates plant growth at the beginning of the vegetation period. This is reflected in the two highest feature contribution values of composite periods: April and first half of May.

The spectral dimension of feature contributions showed a high relevance of the SWIR, RED and NIR bands. SWIR bands are highly correlated with vegetation structural parameters, such as dry vegetation masses or cover fraction (Jacques et al., 2014; Jenal et al., 2021), average leaf angle (Jacquemoud et al., 2009) and leaf dry mass per area (LMA; Rossi et al., 2020). E.g., they can be used to monitor forage mass in grasslands and semi-arid areas (Jacques et al., 2014; Jenal et al., 2021). Further, SWIR bands contribute significantly to estimations of leaf equivalent water thickness (Jacquemoud et al., 2009) and leaf nitrogen (Pellissier et al., 2015; Ramoelo et al., 2015), and consequently plant vitality. RED and NIR bands are highly correlated with biophysical parameters related to vegetation vitality, such as chlorophyll and carotenoid content (Tucker, 1979; Gitelson et al., 2003), but also to vegetation structure and structural parameters such as biomass, LMA and LAI (Tucker, 1979; Jacquemoud et al., 2009; Rossi et al., 2020).

Changes in vegetation composition may not be visible within single bands, as many plant species feature similar reflectance characteristics. However, optical traits vary between different species compositions (Feilhauer et al., 2017). These differences may only be reliably detected by analysing whole time series, accounting for temporal variations of optical traits (Feilhauer et al., 2017; Gholizadeh et al., 2020). Contrarily, other management-induced short-term changes might be obvious. Fertilisation is usually carried out to influence plant nutrient status and to accelerate its growth, thus to increase vegetation structure and vitality (Gibson, D. J., 2009; Sibanda et al., 2015; Rossi et al., 2020). This is reflected in the high contribution of SWIR bands to the fertilisation assessment. Other studies discriminating fertilizer application (Sibanda et al., 2015) or plant nutrient status (Pellissier et al., 2015; Ramoelo et al., 2015) with remote sensing methods ranked red-edge or NIR bands more important than SWIR bands, although the latter was still found highly important. This discrepancy might stem from i) the mono-temporal approach used in the given studies, ii) different study design as two of them assessed plant nutrient status and not management, iii) different ecosystems as Sibanda et al. (2015) and Ramoelo et al. (2015) analysed a South African rangeland and iv) spectral composition as Sibanda et al. (2015) and Pellissier et al. (2015) worked with hyperspectral data.

Grazing reduces vegetation biomass until regrowth occurs. Further, large animals such as cattle may cause structural damage by trampling (Jantunen, 2003). Analogically, mowing leads to short-term reductions of vegetation biomass. This is reflected in the high contribution of SWIR and red-edge bands to the grazing intensity and mowing count models. Other studies assessing grazing intensity (Ma et al., 2019) and mowing count (Estel et al., 2018; Kolecka et al., 2018; Griffiths et al., 2020) commonly use the NDVI and thus red-edge and NIR bands. Li et al. (2016) assessed grazing intensity by estimating above ground biomass with artifical neural networks using Landsat-8 images. These images include also red-edge, NIR and SWIR bands. However, Li et al. (2016) does not give any variable importance measures.

Although grazing intensity and mowing models show fairly similar absolute values of feature contribution, the sign of their contribution may be contradictory: often mown grasslands are mainly used for fodder production, whereas fields with livestock are mown less frequently to facilitate grazing. This effect is visible in Figure 10. Results of Dusseux et al. (2014) support this thesis, since they could differentiate grazing and cutting with 80% accuracy, suggesting major differences in the spectral signature of these management practices.

The variable importance analysis supports the underlying assumption that management practices change vegetation reflectance characteristics mainly by modifying the vegetation composition, structure and vitality. Generally, the variable importance analysis demonstrates the CNN's ability to recognize spectral and temporal patterns related to grassland management.

5.3. Spatial transferability

We evaluated the spatial transferability by using two different approaches: a spatial 3-fold CV and the delineation of the methodology's AOA. The spatial 3-fold CV revealed the value of each of the three observatories for the training of our models. While removing one of the observatories from the training set, one or the other model's accuracy decreased substantially. Thus, the low number of training areas affects the models' robustness, as machine learning algorithms usually require large diverse training datasets when applied to large scale remote sensing data (Blatchford et al., 2021). This also refers to the temporal dimension, as the accuracy loss for each fold varies between years. We can conclude that the data available to this study was sufficient to demonstrate the feasibility of the approach, but the national-extent application may be limited by the methodology's sensitivity to the selection and number of training areas. Further, the sampling strategy for the division of data into training and validation sets is crucial. Thus, results should be carefully examined and interpreted. We tried to cover different climates and management practices by using the most comprehensive grassland management data available to science in Germany and by sampling with a method preserving the high natural variance of environmental data. However, the grassland plots used here are equipped with long-term environmental monitoring units and several manipulative experiments are carried out within some of these plots. Although the intruments and the spatial extent of experiments cover only a fraction of a Sentinel-2 pixel (Fischer et al., 2010), they potentially influence our models and the transferability to other regions.

The 3-fold CV reveals a high variance in OA* and a conspicuously high accuracy loss when applying grazing intensity models trained on ALB and HAI on SCH. We also observe a lower predictive power of $LUI_{region}^{satellite}$ in SCH (see Figure 5) and a higher RMSE than in ALB and HAI (see Table 7). Contrarily, MedAE in SCH are low. Further, OA of component models per region (see Supplementary Material, Figure S.2) are in line with OA shown in Figure 4. Consequently, the r^2 is not related to an inferior LUI component prediction, but may be related to outliers in the prediction of $LUI_{region}^{satellite}$. Visual inspection of individual results reveals an overestimation of $LUI_{region}^{satellite}$ for fertilised pixels in this region. The equation proposed by Blüthgen et al. (2012) reacts highly sensible to the low overall number of fertilised fields (see Table 1) as the regional mean of fertilisation is used as denominator in the fertilisation term. This effect causes extremely high values in the $LUI_{region}^{satellite}$ estimate for a low number of pixels with false positives in fertilisation classification, subsequently hampering the validation. Blüthgen et al. (2012) found a similar issue when studying LUI on ground level: they found no significant relationship between LUI_{region}^{ground} and their five response variables (nitrogen indicator, nitrogen and phosphorus in plant biomass, soil phosphorus and soil C/N ratio) in SCH and assume a relation to low plant diversity and drainage effects of peat soils. This is supported by results of Socher et al. (2013), suggesting the explanation of regional differences in the relationship of land-use and plant species diversity by soil types.

The extrapolation of machine learning algorithms is usually limited by only being applicable within the feature space defined by the training data.

Consequently, we delineated the methodology's AOA accounting for the feature space given by the training data and the models' feature importances. Analysis of the AOA show ample differences between 2017 and 2018, potentially related to the increasing Sentinel-2 data availability. The number of pixels outside AOA differs substantially between regions (a)-(d) and visual inspections reveal spatial patterns potentially reflecting fields or parcels. This points towards regional differences in land management practices, abiotic conditions or vegetation compositions not reflected by our training data. However, we see no obvious relation between areas outside the AOA and distance to calibration plots, suggesting a comparably stronger influence of different land management practices than abiotic conditions on the models' AOA. Further research is needed to reveal the major drivers of model applicability. The AOA maps are valuable tools to determine areas where additional training data is required and to chose regions for data acquisition campaigns. Our products would benefit from data from e.g. i) north-western Germany in vicinity of the Northern Sea and ii) the Alpine Foothills in southern Germany by increasing the feature space of the training data. Despite providing the AOA maps, we cannot answer our third research question entirely.

5.4. National extent estimation of land-use intensity

Our nation-wide LUI estimates, grazing intensity and mowing event depictions show large differences between 2017 and 2018, but only minor differences in fertilisation (see Figure 12). This effect is also visible in the differences in OA of land-use component models between 2017 and 2018: whereas grazing intensity and mowing model OA increased and their transferability decreased in 2018, fertilisation model OA is almost unaffected. This suggests a climatic influence on recognising mowing numbers and grazing intensity. In 2018, Germany was effected by a drought, hampering the growth of grass species and thus potentially impacting the yield of grasslands, the number of mowing events (see Table 1) and grazing intensity. Fertilisation of grasslands is prohibited in Germany until January 31 (German fertiliser regulation, § 6 VIII DüV) and thus usually carried out in late winter and spring. Consequently, the impact on vegetation is potentially most pronounced at the start of the vegetation period, reflected in our models by a high relevance of springtime satellite imagery, and droughts during the summer months may have less impact on fertilisation models than on mowing or grazing intensity models.

Our four demonstration regions (see Figure 9a-d) exhibit different management regimes. They revealed lowest $LUI^{satellite}_{Germany}$ in region (d), covering parts of the low-intensity nature reserve Ohre-Drömling (Figure 9b; Unteneckeret al., 2016), and highest $LUI^{satellite}_{Germany}$ in region (c), covering parts of the transition areas of the biosphere reserve Rhön, and (d), in Oberallgäu (see Figure 11c-d). For the former an increase in fertilisation and mowing intensity was reported (Jedicke, 2013), whereas in the latter intensive grassland management is common (Haas et al., 2001). In Figure 10 we could clearly identify livestock dominated as well as mowing-oriented regions with different fertilisation patterns. The maps reveal a high small scale variance, especially 9a and 9b. Visual inspection of both regions, using satellite imagery from Google Earth (map data from © Google 2021 and © GeoBasis-DE / BKG 2021) and field parcel geometry from the Integrated Administration and Control System (IACS) of the EU provided by the respective federal state, reveals field sizes below 10 ha in (a) and up to 90 ha in (b), containing up to 250 and 2250 Sentinel-2 pixels with 20 m x 20 m resolution, respectively. Hence, the detected variances within 10 x 10 pixels are within common field sizes found in these regions and may be related to features within fields, potentially indicating piece-wise management practices. We witnessed these practices in the Biodiversity Exploratories datasets occasionally referring to meadows fractionally managed, as well as in our preceding field work on meadows and pastures in the context of the terrestrial environmental observatories (TERENO; Wollschläger et al., 2016). We found it challenging to work with such ground truth data containing piece-wise management practices, as it is often recorded only on field level and not on the level of pieces actually managed, resulting in discrepancies between high resolution satellite observations within the field and recorded management practice. Further, data availability, especially cloud cover in the satellite data, may cause such heterogeneities. We expect a less pronounced effect from cloud cover, especially in 2018 where data availability was high, as we i) discarded satellite observations not taken under clear sky conditions, ii) applied a broad compositing scheme and iii) filled the remaining gaps by linear interpolation, assuming that changes between two composites occur gradually. These remaining gaps are mainly situated in winter time (see Supplementary Material, Figure S.1 and Table S.1) where features contribution is low (see Figure 6). However, an effect of cloud cover cannot be ruled out completely. The highly diverse nature of grassland management demands for future studies extending ground truth data and analysing the generated map in detail.

6. Conclusion & Outlook

We presented a framework enabling large-scale LUI mapping and verified its feasibility in three observatories across Germany. The products generated (www.ufz.de/land-use-intensity; Lange et al., 2021), namely national scale maps of mowing frequency, grazing intensity, fertilisation and a LUI quantification, may contribute to the timely endeavour of continuous grassland monitoring and to the scientifically valuable assessment of interactions of grassland management with biodiversity, hydrology and climate change. Thus, they support decision making in land management and conservation. However, the scarcity of validation data is a challenging obstacle in grassland remote sensing. We found a high sensitivity of our models to the selection of calibration areas and a resulting potential underestimation of mowing frequency and grazing intensity. Studies on grassland LUI would greatly benefit from ground truth data from different regions with varying climates, vegetation, soil and topography. Inter-annual model predictions may be used to fill data gaps in studies relying on LUI data, such as hydro- and ecological models, by training the machine learning algorithm on one or more observation years and applying it to years or regions where ground data is missing. This would facilitate large-scale applications by increasing the pool of available data. First tests of temporal transferability were promising, but a detailed analysis can only be done with longer time series and thus more observation years. Further, implementing discrete regression or ordinal classification approaches into current machine learning frameworks would allow, together with an increased amount of spatially and temporally explicit ground truth data, for a more detailed analysis of LUI components and would improve the presented methodology. Spatial resolution of satellite imagery is another issue in grassland remote sensing. Parcels are often small or managed piecewise, thus signals are blurred when pixels cover more than one piece. Improvements in positional accuracy, data fusion of different sensors and higher spatial resolution of upcoming satellite missions may solve this issue. Last, but not least, the differentiation between climate impacts, e.g. droughts, and management changes remains challenging and could be topic of future scientific studies. Combining the products of this study, especially the integrative LUI index generalized across land-use types and unprecedented in spatial extent, with auxiliary climatic data may reveal knowledge about the interactions of land-use, environment and climate.

Acknowledgements

This work is based on data obtained within the DFG Priority Program 1374 "Infrastructure-Biodiversity-Exploratories". We thank the managers of the three Exploratories, Julia Bass, Anna K. Franke, Sonja Gockel, Martin Gorke, Andreas Hemp, Katrin Lorenzen, Franca Marian, Max Müller, Kirsten Reichel-Jung, Swen Renner, Florian Staub, Iris Steitz, Miriam Teuscher, Juliane Vogt, Sandra Weithmann, Konstans Wells, Kerstin Wiesner and all former managers for their work in maintaining the plot and project infrastructure; Christiane Fischer, Maren Gleisberg, Victoria Grießmeier, Jule Mangels and Simone Pfeiffer for giving support through the central office, Jens Nieschulze, Andreas Ostrowski and Michael Owonibi for managing the central data base, 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. Nico Blüthgen, Steffen Both, Markus Fischer, Stephan Getzin, Sonja Gockel, Ralf Lauterbach, Metke Lilienthal, Birgitta König-Ries, Andreas Ostrowski, Niclas Otto, Daniel Prati, Uta Schumacher, Claudia Seilwinder, Jan Thiele and Wolfgang W. Weisser are acknowledged for their valuable work interviewing farmers and setting up the land management dataset. We thank the administration of the Hainich national park, the UNESCO Biosphere Reserve Swabian Alb and the UNESCO Biosphere Reserve Schorfheide-Chorin as well as all land owners for the excellent collaboration. Field work permits were issued by the responsible state environmental offices of Baden-Württemberg, Thüringen, and Brandenburg (according to § 72 BbgNatSchG). We thank the administrators of the UFZ high performance computing system EVE, especially Ben Langenberg, Toni Harzendorf and Tom Strempel, for their support regarding scientific computation at the UFZ in Leipzig. We also wish to acknowledge the UFZ spatial data group, especially Andreas Böhme, for setting up a Web Map Service enabling the visual exploration of the presented products. As pre-processing of Sentinel-2 data was partly accomplished at the supercomputer JUWELS, the authors gratefully acknowledge the Earth System Modelling Project (ESM) for funding this work by providing computing time on the ESM partition of the supercomputer JUWELS at the Jülich Supercomputing Centre (JSC).

Description of author's responsibilities

All authors contributed significantly to the work presented in this paper. Maximilian Lange contributed to study design, drafted the manuscript, acquired and processed the data and performed the analysis. Hannes Feilhauer and Ingolf Kühn contributed to data analysis and reviewed the manuscript. Daniel Doktor designed the present study and reviewed the manuscript. All authors wrote and edited the manuscript.

References

Ali, I., Cawkwell, F., Dwyer, E., Barrett, B., Green, S., 2016. Satellite remote sensing of grasslands: from observation to management. Journal of Plant Ecology 9, 649–671. doi:10.1093/jpe/rtw005.

Asam, S., Klein, D., Dech, S., 2015. Estimation of grassland use intensities based on high spatial resolution LAI time series. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XL-7/W3, 285–291. doi:10.5194/isprsarchives-XL-7-W3-285-2015.

Bernhardt-Römermann, M., Römermann, C., Sperlich, S., Schmidt, W., 2011. Explaining grassland biomass - the contribution of climate, species and functional diversity depends on fertilization and mowing frequency. Journal of Applied Ecology 48, 1088–1097. doi:10.1111/j.1365-2664.2011.01968.x.

Blanco, L.J., Ferrando, C.A., Biurrun, F.N., 2009. Remote Sensing of Spatial and Temporal Vegetation Patterns in Two Grazing Systems. Rangeland Ecology & Management 62, 445–451. doi:10.2111/08-213.1.

Blatchford, M.L., Mannaerts, C.M., Zeng, Y., 2021. Determining representative sample size for validation of continuous, large continental remote sensing data. International Journal of Applied Earth Observation and Geoinformation 94, 102235. doi:10.1016/j.jag.2020.102235.

Blüthgen, N., Dormann, C.F., Prati, D., Klaus, V.H., Kleinebecker, T., Hölzel, N., Alt, F., Boch, S., Gockel, S., Hemp, A., Müller, J., Nieschulze, J., Renner, S.C., Schöning, I., Schumacher, U., Socher, S.A., Wells, K., Birkhofer, K., Buscot, F., Oelmann, Y., Rothenwöhrer, C., Scherber, C., Tscharntke, T., Weiner, C.N., Fischer, M., Kalko, E.K.V., Linsenmair, K.E., Schulze, E., Weisser, W.W., 2012. A quantitative index of land-use intensity in grasslands: Integrating mowing, grazing and fertilization. Basic and Applied Ecology 13, 207–2020. doi:10.1016/j.baae.2012.04.001.

Bouchet, D.C., Cheptou, P., Munoz, F., 2017. Mowing influences community-level variation in resource-use strategies and flowering phenology along an ecological succession on Mediterranean road slopes. Applied Vegetation Science 20, 376–387. doi:10.1111/avsc.12311.

Breiman, L., 2001. Random forests. Machine Learning 45, 5–32. doi:10.1023/A:1010933404324.

Bundesamt für Kartographie und Geodäsie, 2015. Digitales Basis-Landschaftsmodell (AAA-Modellierung). GeoBasis-DE. Geodaten der deutschen Landesvermessung. Bundesamt für Kartographie und Geodäsie.

Bundesamt für Kartographie und Geodäsie, 2017. Verwaltungsgebiete 1:250 000 - Stand 01.01.2017. GeoBasis-DE. Geodaten der deutschen Landesvermessung. Bundesamt für Kartographie und Geodäsie. Available online: https://www.geodatenzentrum.de (accessed on 06 Dec 2017).

Cao, W., Mirjalili, V., Raschka, S., 2020. Rank consistent ordinal regression for neural networks with application to age estimation. Pattern Recognition Letters 140, 325–331. doi:10.1016/j.patrec.2020.11.008.

Chapin III, F., Zavaleta, E., Eviner, V., Naylor, R.L., Vitousek, P.M., Reynolds, H.L., Hooper, D.U., Lavorel, S., Sala, O.E., Hobbie, S.E., Mack, M.C., Díaz, S., 2000. Consequences of changing biodiversity. Nature 405, 234–242. doi:10.1038/35012241.

Cheng, J., Wang, Z., Pollastri, G., 2008. A neural network approach to ordinal regression, in: 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), pp. 1279–1284. doi:10.1109/IJCNN.2008.4633963.

Cilia, C., Panigada, C., Rossini, M., Meroni, M., Busetto, L., Amaducci, S., Boschetti, M., Picchi, V., Colombo, R., 2014. Nitrogen Status Assessment for Variable Rate Fertilization in Maize through Hyperspectral Imagery. Remote Sensing 6, 6549–6565. doi:10.3390/rs6076549.

Conant, R.T., Paustian, K., Elliott, E.T., 2001. Grassland management and conversion into grassland: Effects on soil carbon. Ecological Applications 11, 343–355. doi:10.1890/1051-0761(2001)011[0343:GMACIG]2.0.CO;2.

Congalton, R.G., 1991. A review of assessing the accuracy of classifications of remotely sensed data. Remote Sensing of Environment 37, 35–46. doi:10.1016/0034-4257(91)90048-B.

de Bello, F., Lavorel, S., Albert, C.H., Thuiller, W., Grigulis, K., Dolezal, J., Janeček, Š., Lepš, J., 2011. Quantifying the relevance of intraspecific trait variability for functional diversity. Methods in Ecology and Evolution 2, 163–174. doi:10.1111/j.2041-210X.2010.00071.x.

de Vroey, M., Radoux, J., Defourny, P., 2021. Grassland Mowing Detection Using Sentinel-1 Time Series: Potential and Limitations. Remote Sensing 13. doi:10.3390/rs13030348.

Doktor, D., 2008. Using satellite imagery and ground observations to quantify the effect of intra-annually changing temperature patterns on spring time phenology. PhD thesis. University of London, London, UK.

Doktor, D., Lausch, A., Sprengler, D., Thurner, M., 2014. Extraction of Plant Physiological Status from Hyperspectral Signatures Using Machine Learning Methods. Remote Sensing 6, 12247–12274. doi:10.3390/rs61212247.

Dusseux, P., Vertès, F., Corpetti, T., Corgne, S., Hubert-Moy, L., 2014. Agricultural practices in grasslands detected by spatial remote sensing. Environmental Monitoring and Assessment 186, 8249–8265. doi:10.1007/s10661-014-4001-5.

Estel, S., Mader, S., Levers, C., Verburg, P.H., Baumann, M., Kuemmerle, T., 2018. Combining satellite data and agricultural statistics to map grassland management intensity in Europe. Environmental Research Letters 13, 074020. doi:10.1088/1748-9326/aacc7a.

Falbel, D., Allaire, J.J., Chollet, F., RStudio, Google, Tang, Y., Van Der Bijl, W., Studer, M., Keydana, S., 2020. keras: R Interface to 'Keras'. R package version 2.3.0.0, https://CRAN.R-project.org/package=keras (accessed on 05 Nov 2020).

FAO, 2009. The State Of Food And Agriculture - Livestock In The Balance. Technical Report. Food And Agriculture Organization Of The United Nations. Rome, Italy.

Feilhauer, H., Somers, B., van der Linden, S., 2017. Optical trait indicators for remote sensing of plant species composition: Predictive power and seasonal variability. Ecological Indicators 73, 825–833. doi:10.1016/j.ecolind.2016.11.003.

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., Weisser, W.W., 2010. Implementing large-scale and long-term functional biodiversity research: The Biodiversity Exploratories. Basic and Applied Ecology 11, 473–485. doi:10.1016/j.baae.2010.07.009.

Franke, J., Keuck, V., Siegert, F., 2012. Assessment of grassland use intensity by remote sensing to support conservation schemes. Journal for Nature Conservation 20, 125–134. doi:10.1016/j.jnc.2012.02.001.

Gholizadeh, H., Gamon, J.A., Helzer, C.J., Cavender-Bares, J., 2020. Multi-temporal assessment of grassland α- and β-diversity using hyperspectral imaging. Ecological Applications 30, e02145. doi:10.1002/eap.2145.

Gibson, D. J., 2009. Grasses and Grassland Ecology. Oxford University Press, Oxford, UK.

Gilhaus, K., Hölzel, N., 2016. Seasonal variations of fodder quality and availability as constraints for stocking rates in year-round grazing schemes. Agriculture, Ecosystems & Environment 234, 5–15. doi:10.1016/j.agee.2016.03.013.

Gitelson, A.A., Gritz, Y., Merzlyak, M.N., 2003. Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. Journal of Plant Physiology 160, 271–282. doi:10.1078/0176-1617-00887.

Gómez Giménez, M., de Jong, R., Della Peruta, R., Keller, A., Schaepman, M.E., 2017. Determination of grassland use intensity based on multi-temporal remote sensing data and ecological indicators. Remote Sensing of Environment 198, 126–139. doi:10.1016/j.rse.2017.06.003.

Griffiths, P., Nendel, C., Pickert, J., Hostert, P., 2020. Towards national-scale characterization of grassland use intensity from integrated Sentinel-2 and Landsat time series. Remote Sensing of Environment 238, 111124. doi:10.1016/j.rse.2019.03.017.

Haas, G., Wetterich, F., Kopke, U., 2001. Comparing intensive, extensified and organic grassland farming in southern Germany by process life cycle assessment. Agriculture Ecosystems & Environment 83, 43–53. doi:10.1016/S0167-8809(00)00160-2.

Halabuk, A., Mojses, M., Halabuk, M., David, S., 2015. Towards Detection of Cutting in Hay Meadows by Using of NDVI and EVI Time Series. Remote Sensing 7, 6107–6132. doi:10.3390/rs70506107.

Hollberg, J.L., Schellberg, J., 2017. Distinguishing Intensity Levels of Grassland Fertilization Using Vegetation Indices. Remote Sensing 9, 81. doi:10.3390/rs9010081.

Ienco, D., Gaetano, R., Dupaquier, C., Maurel, P., 2017. Land Cover Classification via Multitemporal Spatial Data by Deep Recurrent Neural Networks. IEEE Geoscience and Remote Sensing Letters 14, 1685–1689. doi:10.1109/LGRS.2017.2728698.

IPBES, 2019. Summary for policymakers of the global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. S. Díaz, J. Settele, E. S. Brondízio, H. T. Ngo, M. Guèze, J. Agard, A. Arneth, P. Balvanera, K. A. Brauman, S. H. M. Butchart, K. M. A. Chan, L. A. Garibaldi, K. Ichii, J. Liu, S. M. Subramanian, G. F. Midgley, P. Miloslavich, Z. Molnár, D. Obura, A. Pfaff, S. Polasky, A. Purvis, J. Razzaque, B. Reyers, R. Roy Chowdhury, Y. J. Shin, I. J. Visseren-Hamakers, K. J. Willis, and C. N. Zayas (eds.). IPBES secretariat, Bonn, Germany. doi:10.5281/zenodo.3553579.

IPCC, 2019. Climate Change and Land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems. [P. R. Shukla, J. Skea, E. Calvo Buendia, V. Masson-Delmotte, H.-O. Pörtner, D. C. Roberts, P. Zhai, R. Slade, S. Connors, R. van Diemen, M. Ferrat, E. Haughey, S. Luz, S. Neogi, M. Pathak, J. Petzold, J. Portugal Pereira, P. Vyas, E. Huntley, K. Kissick, M. Belkacemi, J. Malley, (eds.)].

In press. Available online: https://www.ipcc.ch/srccl (accessed on 23 Mar 2021).

Jacquemoud, S., Verhoef, W., Baret, F., Bacour, C., Zarco-Tejada, P.J., Asner, G.P., François, C., Ustin, S.L., 2009. PROSPECT plus SAIL models: A review of use for vegetation characterization. Remote Sensing Of Environment 113, S56–S66. doi:10.1016/j.rse.2008.01.026.

Jacques, D.C., Kergoat, L., Hiernaux, P., Mougin, E., Defourny, P., 2014. Monitoring dry vegetation masses in semi-arid areas with MODIS SWIR bands. Remote Sensing of Environment 153, 40–49. doi:10.1016/j.rse.2014.07.027.

Jantunen, J., 2003. Vegetation changes in a semi-natural grassland during mowing and grazing periods. Annales Botanici Fennici 40, 255–263. Available online: http://www.jstor.org/stable/23726842 (accessed on 02 Sep 2021).

Jarvis, A., Reuter, H., Nelson, A., Guevara, E., 2008. Hole-filled SRTM for the globe. Version 4, available from the CGIAR-CSI SRTM 90m Database (http://srtm.csi.cgiar.org, accessed on 31 Aug 2021).

Jedicke, E., 2013. Report for the Periodic Review of Rhön UNESCO Biosphere Reserve 2013. Available online: http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/SC/pdf/Periodic_Review_Rhon_2013.pdf (accessed on 17 Mar 2021).

Jenal, A., Hüging, H., Ahrends, H.E., Bolten, A., Bongartz, J., Bareth, G., 2021. Investigating the Potential of a Newly Developed UAV-Mounted

VNIR/SWIR Imaging System for Monitoring Crop Traits—A Case Study for Winter Wheat. Remote Sensing 13. doi:10.3390/rs13091697.

Jones, M.B., Donnelly, A., 2004. Carbon sequestration in temperate grassland ecosystems and the influence of management, climate and elevated CO2. New Phytologist 164, 423–439. doi:10.1111/j.1469-8137.2004.01201.x.

Kolecka, N., Ginzler, C., Pazur, R., Price, B., Verburg, P.H., 2018. Regional Scale Mapping of Grassland Mowing Frequency with Sentinel-2 Time Series. Remote Sensing 10, 1221. doi:10.3390/rs10081221.

Köppen, W., 1918. Klassifikation der Klimate nach Temperatur, Niederschlag und Jahresablauf (Classification of climates according to temperature, precipitation and seasonal cycle). Petermanns Geogr. Mitt. 64, 193–203, 243–248.

Kruess, A., Tscharntke, T., 2002. Grazing intensity and the diversity of grasshoppers, butterflies, and trap-nesting bees and wasps. Conservation Biology 16, 1570–1580. doi:10.1046/j.1523-1739.2002.01334.x.

Kuemmerle, T., Erb, K., Meyfroidt, P., Müller, D., Verburg, P.H., Estel, S., Haberl, H., Hostert, P., Jepsen, M.R., Kastner, T., Levers, C., Lindner, M., Plutzar, C., Verkerk, P.J., van der Zanden, E.H., Reenberg, A., 2013. Challenges and opportunities in mapping land use intensity globally. Current Opinion in Environmental Sustainability 5, 484–493. doi:10.1016/j.cosust.2013.06.002.

Kussul, N., Lavreniuk, M., Skakun, S., Shelestov, A., 2017. Deep Learning Classification of Land Cover and Crop Types Using Remote Sensing Data. IEEE Geoscience and Remote Sensing Letters 14, 778–782. doi:10.1109/LGRS.2017.2681128.

Lange, M., Dechant, B., Rebmann, C., Vohland, M., Cuntz, M., Doktor, D., 2017. Validating MODIS and Sentinel-2 NDVI Products at a Temperate Deciduous Forest Site Using Two Independent Ground-Based Sensors. Sensors 17. doi:10.3390/s17081855.

[dataset] Lange, M., Feilhauer, H., Kühn, I., Doktor, D., 2021. Land-use intensity quantification and management classifications in grasslands of Germany 2017/2018. Mendeley Data. V1. doi:10.17632/m9rrv26dvf.1.

Li, F., Zheng, J., Wang, H., Luo, J., Zhao, Y., Zhao, R., 2016. Mapping grazing intensity using remote sensing in the Xilingol steppe region, Inner Mongolia, China. Remote Sensing Letters 7, 328–337. doi:10.1080/2150704X.2015.1137987.

Liaw, A., Wiener, M., 2002. Classification and regression by randomforest. R News 2, 18–22. URL: https://CRAN.R-project.org/doc/Rnews/.

Lopes, M., Fauvel, M.M., Girard, S., Sheeren, D., 2017. Object-based classification of grasslands from high resolution satellite image time series using Gaussian mean map kernels. Remote Sensing 9, 688. doi:10.3390/rs9070688.

Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S., 2020. From local explanations to global understanding with explainable AI for trees. Nature Machine Intelligence 2, 2522–5839. doi:10.1038/s42256-019-0138-9.

Lundberg, S.M., Erion, G.G., Lee, S., 2019. Consistent individualized feature attribution for tree ensembles. arXiv:1802.03888. Conference paper, follow-up to 2017 ICML Workshop, v3 (accessed on 16 Mar 2021).

Lundberg, S.M., Lee, S., 2017. A Unified Approach to Interpreting Model Predictions, in: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (Eds.), Advances in Neural Information Processing Systems 30. Curran Associates, Inc.. volume 30, pp. 4765–4774. URL: http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf.

Ma, Q., Chai, L., Hou, F., Chang, S., Ma, Y., Tsunekawa, A., Cheng, Y., 2019. Quantifying Grazing Intensity Using Remote Sensing in Alpine Meadows on Qinghai-Tibetan Plateau. Sustainability 11, 1–14. doi:10.3390/su11020417.

Maksymiuk, S., Gosiewska, A., Biecek, P., Staniak, M., Burdukiewicz, M., 2020. shapper: Wrapper of Python Library 'shap'. R package version 0.1.3, https://CRAN.R-project.org/package=shapper (accessed on 03 Dec 2020).

McNemar, Q., 1947. Note on the sampling error of the difference between correlated proportions or percentages. Psychometrika 12, 153–157. doi:10.1007/BF02295996.

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

Meyer, H., Reudenbach, C., Ludwig, M., Nauss, T., Pebesma, E., 2021. CAST: 'caret' Applications for Spatial-Temporal Models. R package version 0.5.1, https://CRAN.R-project.org/package=cast (accessed on 16 Jun 2021).

Mueller-Wilm, U., Devignot, O., Pessiot, L., 2019. S2 MPC Sen2Cor Configuration and User Manual. Reference S2-PDGS-MPC-L2A-SUM-V2.8, Issue 2, Available online: http://step.esa.int/main/third-party-plugins-2/sen2cor/sen2cor_v2-8/ (accessed on 21 Oct 2020).

Ostrowski, A., Nieschulze, J., Schulze, E., König-Ries, B., 2020. Coordinates and Inventory Overview of all Grid Plots (GPs). Biodiversity Exploratories Information System. Dataset. https://www.bexis.uni-jena.de/ (accessed on 27 Aug 2021). Dataset ID=20907.

Pellissier, P.A., Ollinger, S.V., Lepine, L.C., Palace, M.W., McDowell, W.H., 2015. Remote sensing of foliar nitrogen in cultivated grasslands of human dominated landscapes. Remote Sensing of Environment 167, 88–97. doi:10.1016/j.rse.2015.06.009. Special Issue on the Hyperspectral Infrared Imager (HyspIRI).

Post, W.M., Kwon, K.C., 2000. Soil carbon sequestration and land-use change: processes and potential. Global Change Biology 6, 317–327. doi:10.1046/j.1365-2486.2000.00308.x.

Preidl, S., Lange, M., Doktor, D., 2020. Introducing APiC for regionalised land cover mapping on the national scale using Sentinel-2A imagery. Remote Sensing of Environment 240, 111673. doi:10.1016/j.rse.2020.111673.

R Core Team, 2020. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. Available online: http://www.R-project.org/ (accessed on 05 Nov 2020).

Ramoelo, A., Cho, M., Mathieu, R., Skidmore, A.K., 2015. Potential of Sentinel-2 spectral configuration to assess rangeland quality. Journal of Applied Remote Sensing 9, 1–11. doi:10.1117/1.JRS.9.094096.

Reinermann, S., Asam, S., Kuenzer, C., 2020. Remote Sensing of Grassland Production and Management – A Review. Remote Sensing 12, 1949. doi:10.3390/rs12121949.

Rodriguez-Galiano, V.F., Ghimire, B., Rogan, J., Chica-Olmo, M., Rigol-Sanchez, J.P., 2012. An assessment of the effectiveness of a random forest classifier for land-cover classification. ISPRS Journal of Photogrammetry and Remote Sensing 67, 93–104. doi:10.1016/j.isprsjprs.2011.11.002.

Rossi, C., Kneubuehler, M., Schuetz, M., Schaepman, M., Haller, R., Risch, A., 2020. From local to regional: Functional diversity in differently managed alpine grasslands. Remote Sensing of Environment 236, 111415. doi:10.1016/j.rse.2019.111415.

Roudier, P., Brugnard, C., Beaudette, D., Louis, B., 2020. clhs: Conditioned Latin Hypercube Sampling. R package version 0.7-3, https://CRAN.R-project.org/package=clhs (accessed on 30 Oct 2020).

Rußwurm, M., Körner, M., 2017. Multi-Temporal Land Cover Classification with Long Short Term Memory Neural Networks. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLII-1/W1, 551–558. doi:10.5194/isprs-archives-XLII-1-W1-551-2017.

Sakowska, K., Juszczak, R., Gianelle, D., 2016. Remote Sensing of Grassland Biophysical Parameters in the Context of the Sentinel-2 Satellite Mission. Journal of Sensors 2016, 1–16. doi:10.1155/2016/4612809.

Sibanda, M., Mutanga, O., Rouget, M., Odindi, J., 2015. Exploring the potential of in situ hyperspectral data and multivariate techniques in discriminating different fertilizer treatments in grasslands. Journal of Applied Remote Sensing 9, 1–19. doi:10.1117/1.JRS.9.096033.

Siegmund, R., Grant, K., Wagner, M., Hartmann, S., 2016. Satellite-based monitoring of grassland: assessment of harvest dates and frequency using SAR. Remote Sensing for Agriculture, Ecosystems, and Hydrology XVIII 9998, 9–25. doi:10.1117/12.2240947.

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., 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 and Applied Ecology 14, 126–136. doi:10.1016/j.baae.2012.12.003.

Stanners, D.A., Bourdeau, P., 1995. Europe's environment: the Dobris assessment. European Environment Agency, Office for the Official Publications of the European Communities, Luxembourg, LU. Available online: https://www.eea.europa.eu/publications/92-826-5409-5/ (accessed on 17 Oct 2017).

Stumpf, F., Schneider, M.K., Keller, A., Mayr, A., Rentschler, T., Meuli, R.G., Schaepman, M., Liebisch, F., 2020. Spatial monitoring of grassland management using multi-temporal satellite imagery. Ecological Indicators 113, 106201. doi:10.1016/j.ecolind.2020.106201.

Svetnik, V., Liaw, A., Tong, C., Culberson, J.C., Sheridan, R.P., Feuston, B.P., 2003. Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling. Journal of Chemical Information and Computer Sciences 43, 1947–1958. doi:10.1021/ci034160g.

Taravat, A., Wagner, M.P., Oppelt, N., 2019. Automatic Grassland Cutting Status Detection in the Context of Spatiotemporal Sentinel-1 Imagery Analysis and Artificial Neural Networks. Remote Sensing 11, 711. doi:10.3390/rs11060711.

Tharwat, A., 2020. Classification assessment methods. New England Journal of Entrepreneurship 17, 168–192. doi:10.1016/j.aci.2018.08.003.

Tsiafouli, M.A., Thebault, E., Sgardelis, S.P., de Ruiter, P.C., van der Putten, W.H., Birkhofer, K., Hemerik, L., de Vries, F.T., Bardgett, R.D., Brady, M.V., Bjornlund, L., Jorgensen, H.B., Christensen, S., D' Hertefeldt, T., Hotes, S., Hol, W.H.G., Frouz, J., Liiri, M., Mortimer, S.R., Setala, H., Tzanopoulos, J., Uteseny, K., Pizl, V., Stary, J., Wolters, V., Hedlund, K., 2015. Intensive agriculture reduces soil biodiversity across Europe. Global Change Biology 21, 973–985. doi:10.1111/gcb.12752.

Tucker, C.J., 1979. Red and Photographic Infrared Linear Combinations for Monitoring Vegetation. Remote Sensing of Environment 8, 127–150. doi:10.1016/0034-4257(79)90013-0.

Untenecker, J., Tiemeyer, B., Freibauer, A., Laggner, A., Braumann, F., Luterbacher, J., 2016. Fine-grained detection of land use and water table changes on organic soils over the period 1992-2012 using multiple data sources in the Drömling nature park, Germany. Land Use Policy 57, 164–178. doi:10.1016/j.landusepol.2016.04.016.

Vogt, J., Weisser, W.W., Ayasse, M., Fischer, M., Schumacher, U., Schreiber, C., Lauterbach, R., Frank, A., Ostrowski, A., Teuscher, M., 2020. Land use in grasslands: raw data of yearly owner interviews. Biodiversity Exploratories Information System. Dataset. https://www.bexis.uni-jena.de/ (accessed on 27 Aug 2021). Dataset ID=26487.

Wesche, K., Krause, B., Culmsee, H., Leuschner, C., 2012. Fifty years of change in Central European grassland vegetation: Large losses in species richness and animal-pollinated plants. Biological Conservation 150, 76–85. doi:10.1016/j.biocon.2012.02.015.

Wilson, J., Peet, R., Dengler, J., Pärtel, M., 2012. Plant species richness: the world records. Journal of Vegetation Science 23, 796–802. doi:10.2307/23251355.

Wollschläger, U., Attinger, S., Borchardt, D., Brauns, M., Cuntz, M., Dietrich, P., Fleckenstein, J.H., Friese, K., Friesen, J., Harpke, A., Hildebrandt, A., Jäckel, G., Kamjunke, N., Knöller, K., Kögler, S., Kolditz, O., Krieg, R., Kumar, R., Lausch, A., Liess, M., Marx, A., Merz, R., Mueller, C., Musolff, A., Norf, H., Oswald, S.E., Rebmann, C., Reinstorf, F., Rode, M., Rink, K., Rinke, K., Samaniego, L., Vieweg, M., Vogel, H., Weitere, M., Werban, U., Zink, M., Zacharias, S., 2016. The Bode hydrological observatory: a platform for integrated, interdisciplinary hydro-ecological research within the TERENO Harz/Central German Lowland Observatory. Environmental Earth Sciences 76, 29. doi:10.1007/s12665-016-6327-5.

Xia, T., Miao, Y., Wu, D., Shao, H., Khosla, R., Mi, G., 2016. Active Optical Sensing of Spring Maize for In-Season Diagnosis of Nitrogen Status Based on Nitrogen Nutrition Index. Remote Sensing 8, 605. doi:10.3390/rs8070605.

Yates, F., 1934. Contingency tables involving small numbers and the χ^2 test. Supplement to the Journal of the Royal Statistical Society 1, 217–235. doi:10.2307/2983604.

List of Figure Captions

Figure 1: Location of study areas in Germany. Calibration and validation plots from the DFG Biodiversity Exploratories are marked with circles. We are not allowed to publish exact parcel positions due to privacy policies. Exact positions of the four demonstration regions chosen for visual and statistical interpretation are given as red boxes. Background colors show simplified land cover information. State borders are shown as black lines (© GeoBasis-DE / Bundesamt für Kartographie und Geodäsie, 2017).

Figure 2: Normalised difference vegetation index (NDVI) time series of 2018 of center pixels in six plots (a-f) within the Biodiversity Exploratories. Pastures (b, d, f) are and meadows (a, c, e) are not grazed.

Figure 3: CNN structure: The input was reshaped such that 16 composite periods and nine bands per Sentinel-2 pixel serve as input for the net, which consists of two convolutional layers, with 64 and 128 filters of kernel-size three and valid padding, respectively, a batch normalisation layer in between, a (max-)pooling layer of size two and a subsequent dense neural network with an input layer consisting of 768 neurons with a rectified linear unit (ReLU) activation function, a dropout layer with a dropout rate of 0.25 and an output layer with a softmax activation function and as many neurons as the number of classes in the target variable.

Figure 4: Overall accuracies of 100 runs with CNN, shown as blue boxes, and RF models, shown as green boxes, with whiskers extending to the full range of accuracies. The models were trained and validated with grazing, mowing or fertilisation data of the years 2017 and 2018 (70% training, 30% validation).

Figure 5: Squared Pearson correlation coefficients (r^2) between LUI_{region}^{ground} and $LUI_{region}^{satellite}$ based on classifications of CNN, shown as blue bars, and RF models, shown as green bars, and using data of each observatory (ALB: *Schwäbische Alb*, HAI: *Hainich* and SCH: *Schorfheide*), as well as the union of all observatories (*Overall*). All correlations are significant ($p < 10^{-20}$).

Figure 6: Feature importance: Summary of CNN feature contribution on predictions using data from 2017 (left) and 2018 (right) generated by using the mean absolute SHAP values of features grouped by satellite band (first and third column) or composite period (second and fourth column) for each variable (by row from top to bottom). Mean absolute feature contribution is shown on the ordinate, the abscissa shows the respective bands and composite periods.

Figure 7: Spatial 3-fold cross-validation: CNN models, 100 per fold, were trained and validated on different spatial subsets for the observation years 2017 (blue) and 2018 (green). Training was done with data from two regions, e.g. *Schwäbische Alb* (ALB) and *Hainich* (HAI), and validation with the remaining third region, e.g. *Schorfheide* (SCH). Boxplots show the OA^* of the resulting 100 models per fold. The bars show the accuracy loss (or, in one case, gain) of the CV model with the highest OA^* compared to the highest OA of the models applied on data of all observatories.

Figure 8: The maps of the four regions (a)-(d) show land cover information with colours respective to the bottom-right legend (© GeoBasis-DE / Bundesamt für Kartographie und Geodäsie, 2015), whereas grassland pixels are overlaid by AOA_{LUI} of 2018, with areas inside AOA_{LUI} in green and areas outside AOA_{LUI} in magenta colours.

Figure 9: The maps of the four regions (a)-(d) show land cover information with colours respective to the bottom-right legend (© GeoBasis-DE / Bundesamt für Kartographie und Geodäsie, 2015), whereas grassland pixels are overlaid by their $LUI^{satellite}_{Germany}$ of 2018 with colours ranging from green (extensive use) to magenta (intensive use).

Figure 10: Subsets of national extent maps: (a)-(d) grazing classes, mowing counts, fertilisation and $LUI^{satellite}_{Germany}$, respectively, of 2018 for an area of 10 km x 10 km in the district Oberallgäu (see Figure 9d) in the federal state Bavaria. All non-grassland pixels use colours respective to the legend of background values in the right. Grazing classes range from zero (low grazing intensity, green) to three (high grazing intensity, magenta) and mowing counts from zero (green) to four (magenta). Fertilisation (no/yes) is indicated in green and light magenta, respectively. $LUI^{satellite}_{Germany}$ values are aggregated into five classes with colours ranging from green (extensive use) to magenta (intensive use). Map (e) displays the digital elevation model of the region based on NASA's Shuttle Radar Topographic Mission (SRTM; Jarvis et al., 2008) and highlights specific locations.

Figure 11: $LUI^{satellite}_{Germany}$ and its components (grazing class, mowing count and fertilisation) within the AOA for 2018 in grasslands in the four demonstration regions (a)-(d).

Figure 12: $LUI^{satellite}_{Germany}$ and its components (grazing class, mowing count and fertilisation) for 2017 and 2018 in grasslands of Germany within the AOA.