Remote Sensing of Environment 267 (2021) 112751

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Mowing event detection in permanent grasslands: Systematic evaluation of input features from Sentinel-1, Sentinel-2, and Landsat 8 time series

Felix Lobert [a,*], Ann-Kathrin Holtgrave [b], Marcel Schwieder [a], Marion Pause [c], Juliane Vogt [d], Alexander Gocht [a], Stefan Erasmi [a]

[a] Thünen Institute of Farm Economics, Bundesallee 63, 38116 Braunschweig, Germany
[b] Thünen Institute of Rural Studies, Bundesallee 64, 38116 Braunschweig, Germany
[c] Faculty of Environmental Sciences, Technische Universität Dresden, Helmholtzstr. 10, 01062 Dresden, Germany
[d] Department of Ecology and Ecosystem Management, Technische Universität München, Hans-Carl-von-Carlowitz-Platz 2, 85354 Freising, Germany

ARTICLE INFO

Editor: Marie Weiss

Keywords:
Convolutional neural networks
Deep learning
Sequential classification
Agriculture
Interferometric coherence
SAR
NDVI
Backscatter coefficient
GLCM
Moving window
Grassland management intensity
Land use intensity

ABSTRACT

The intensity of land use and management in permanent grasslands affects both biodiversity and important ecosystem services. Comprehensive knowledge about these intensities is a crucial factor for sustainable decision-making in landscape policy. For meadows, the management intensity can be described by proxies such as the mowing frequency, usually, a higher number of cuts indicate higher intensities. Dense time series of medium resolution (10–30 m) remote sensing data are suitable for the detection of mowing events. However, existing studies revealed a general lack of consensus about the most appropriate input data set for a consistent and reliable mowing detection.

We systematically evaluated the synergistic use of acquisitions from Sentinel-1, Sentinel-2, and Landsat 8 to detect the occurrence, frequency, and date of mowing events as an indicator of grassland management intensity. Dense time series of NDVI (Sentinel-2 and Landsat 8), γ^0 backscatter, backscatter cross-ratio, backscatter second-order texture metrics as well as 6-day interferometric coherence (Sentinel-1) were used as input features. All possible combinations of input features were tested to train a one-dimensional convolutional neural network, which enables enhanced exploitation of the temporal domain of the data. The evaluation was conducted on 64 meadows for an overall of 257 mowing events from 2017 to 2019 in Germany.

Our results revealed that the combination of input features improves the detection performance. The highest overall accuracy was reached by a combination of NDVI, backscatter cross-ratio, and interferometric coherence with an F1-Score of 0.84. The mowing frequency was predicted with a mean absolute error of 0.38 events per year, while the date of the events was missed by 3.79 days on average. NDVI time series alone mostly underperformed in comparison to optical/SAR combinations but clearly outperformed input-sets that were solely based on SAR features. The proposed model performed well for meadows with low to medium management intensities but further testing is recommended for highly intensive managed parcels.

The results clearly demonstrate the additional value of fusing time series of the three present Earth observation systems that deliver a freely available global coverage of the land surface at medium resolution.

1. Introduction

The increasing intensity of land use and management in agricultural areas of the mid-latitudes is an important driver of biodiversity loss (Benton et al., 2003; Gossner et al., 2016; Hallmann et al., 2017; IPBES, 2019). For conservation of biodiversity and sustainable decision-making in landscape policy and planning, comprehensive knowledge about the parameters that represent land use intensity in agricultural areas is a crucial factor (Billeter et al., 2008; Foley et al., 2005; Tscharntke et al., 2005). In the context of climate change mitigation, permanent grasslands (i.e. land used as grassland for five years or more) are in general treated as soil carbon sinks in comparison to cropland (Poeplau et al., 2020). The management intensity of grassland impacts the quality of the sink where intensively used grasslands are assumed to contribute to soil carbon loss and greenhouse gas emissions (Conant et al., 2001; Hörtnagl et al., 2018; Jones et al., 2005; Sándor et al., 2018). In permanent

* Corresponding author.
E-mail address: felix.lobert@thuenen.de (F. Lobert).

https://doi.org/10.1016/j.rse.2021.112751
Received 1 April 2021; Received in revised form 7 October 2021; Accepted 11 October 2021
Available online 22 October 2021

grasslands, the management intensity can be described by parameters such as the mowing frequency on meadows (Weiner et al., 2011). However, area-wide information on mowing events, e.g., from public statistics or farmer surveys, is missing in most agricultural regions.

Time series of optical and radar imagery provide a valuable data basis for the generation of area-wide information on grassland management intensity, indicated by a growing body of literature (Reinermann et al., 2020). Besides animal grazing, fertilization, and irrigation approaches, the detection of mowing events makes up the largest part of the literature. A comprehensive overview of previous research on remote sensing of grassland management was provided by Reinermann et al. (2020). Several studies investigated the relationship between remote sensing time series and mowing events based on the rationale that cutting of the grass leads to an abrupt change in the remote sensing signal (De Vroey et al., 2021; Kolecka et al., 2018; Stendardi et al., 2019; Voormansik et al., 2020). Based on this relationship, different methods for the automatic detection of mowing events were developed.

Commonly used input features are vegetation indices derived from optical sensors like, e.g., the normalized difference vegetation index (NDVI) (Estel et al., 2018; Griffiths et al., 2020; Kolecka et al., 2018; Lobert et al., 2021) and the Enhanced Vegetation Index (EVI) (Halabuk et al., 2015). Vegetation indices are strongly related to vegetation vitality and show a clear response to changes in biomass, which makes them straightforward to interpret in this context. The indices are mostly derived from medium-resolution Sentinel-2 (S2) and Landsat 8 (L8) imagery and coarser resolved imagery as from the Moderate Resolution Imaging Spectroradiometer (MODIS) (Estel et al., 2018; Halabuk et al., 2015). However, the main drawback of optical satellite data is the imperviousness of clouds towards the used domain of the electromagnetic spectrum. Consequently, data gaps occur in optical time series, and cloud detection algorithms and interpolation strategies are required. For mowing detection, this might lead to the omission of mowing events. Moreover, undetected clouds can be misclassified as mowing events (Griffiths et al., 2020).

This limitation can be overcome by the use of radar systems, which are not affected by clouds and enable to derive equidistant time series. Several studies have shown that changes in time series of the Synthetic Aperture Radar (SAR) backscatter coefficient, e.g., acquired by X-band COSMO-SkyMed (CSM), can be related to mowing events (Grant et al., 2015; Siegmund et al., 2016; Zalite et al., 2016, 2014). The same behavior was also observed in time series acquired by C-band Sentinel-1 (S1) (De Vroey et al., 2021; Grant et al., 2015; Lobert et al., 2021; Siegmund et al., 2016; Stendardi et al., 2019; Taravat et al., 2019). Taravat et al. (2019) have shown that second-order texture metrics derived from S1 backscatter coefficient were also able to improve the results of a mowing detection algorithm. This can be explained by changes in the field (i.e. equal grass heights due to mowing) that affect the texture of the meadows.

Besides radar backscatter, several studies reported the potential of the interferometric coherence as an input feature for mowing detection (De Vroey et al., 2021; Tamm et al., 2016; Voormansik et al., 2020; Zalite et al., 2016, 2014). The temporal decorrelation of the scatterers in a grassland parcel increases with grass height and decreases significantly after a mowing event. This leads to very low coherence between the image pair where one image was taken before and the other one after the event. On the contrary, the next pair, where both images are acquired after the event, has a significantly higher coherence due to higher temporal correlation. Tamm et al. (2016) observed the coherence to remain on a higher level for up to 36 days after an event. However, this pattern is not consistent with every mowing event (Chiboub et al., 2019) and De Vroey et al. (2021) concluded that coherence alone is not sufficient for the detection. Furthermore, interferometric coherence is also influenced by precipitation and dew, which can lead to ambiguities according to Tamm et al. (2016). The effect of dew, however, can be reduced when only imagery acquired in the afternoon is considered.

To overcome the mentioned limitations of each sensor system, Stendardi et al. (2019) investigated the potential of using both optical and radar data as input features for the detection of mowing events. Based on their comparative analysis of S1 backscatter and S2 NDVI time series, they concluded that uncertainties in one input feature caused by, e.g., clouds in case of S2, could be resolved with the complementary data source. Yet, they did not present an approach that combines S1 and S2 data for mowing detection. The detection of mowing events based on S2 LAI and S1 coherence is implemented within the Sentinels for Common Agriculture Policy system (Sen4CAP). The independently detected events are merged ex-post using confidence levels, where S2 is assumed to be more reliable than S1 (de Vendictis et al., 2019). However, only a technical report was provided missing a validation. Lobert et al. (2021) used S1 backscatter cross-ratio, as well as S2 and L8 NDVI as input features. Since their results revealed shortcomings of the chosen feature combination, they suggested further improvements by adding more features and performing a feature selection.

Recent studies explored the analogies between optical and SAR data over agricultural areas for the possibility of using SAR data to fill gaps in optical time series (Holtgrave et al., 2020) and also included the generation of synthetic time series where, e.g., gap-free NDVI time series were produced based on a recurrent neural network (RNN) fed by S1 and S2 data (Garioud et al., 2020, 2019). This may be a promising approach as it uses the strengths of optical vegetation indices for the detection of mowing events while also tackling the weakness of data gaps in optical time series by the use of SAR data.

In literature, we identified two general pathways to detect mowing events, which are either based on classic change detection or machine learning (ML) algorithms. Both pathways assume that mowing events are indicated by significant changes in the time series. The first aims to distinguish these changes by derivatives of time series (Grant et al., 2015; Siegmund et al., 2016), rule sets that are applied to key points within the growing season (Courault et al., 2010; De Vroey et al., 2021; Grant et al., 2015; Kolecka et al., 2018; Stendardi et al., 2019) or idealized annual trajectories (Estel et al., 2018; Griffiths et al., 2020). These approaches partially depend on predefined thresholds that can vary between investigated regions and periods and the number of detected events may differ according to the chosen threshold (Garioud et al., 2019). At the same time, these approaches are based on sound empirical assumptions.

The second pathway uses ML algorithms to detect changes in time series induced by mowing events. For these methods, the derived time series are transformed into a supervised classification problem. This is mostly done by segmenting the time series into smaller sequences that are labeled as *mown* or *not mown*. Halabuk et al. (2015) investigated the performance of the Classification and Regression Trees (CART) algorithm and were able to predict mowing events with an overall accuracy of 85%. However, their test setup did not allow to infer the date and number of mowing events within the year, but only whether a grassland parcel was mown at all. Taravat et al. (2019) trained a multilayer perceptron (MLP) on short sequences of a multivariate time series. They detected 75% of the mowing events correctly and reached an overall accuracy of 85.7%. A shortcoming was the small number of 10 reference meadows and the short study period of one year. Lobert et al. (2021) used a Support Vector Machines (SVM) and a one-dimensional convolutional neural network (1D-CNN) classifier. Both models showed shortcomings in terms of accuracy, yet, the CNN achieved more robust results.

The presented studies show a wide range of opportunities for the detection of mowing events from satellite time series and highlight the overall potential of dense Earth observation time series and ML for the monitoring of grassland management intensity. However, most of the studies were evaluated with little reference data from small regions, which limits the transferability of the proposed approaches (Reinermann et al., 2020). Further, no consensus could be identified about an optimal feature combination that allows for a reliable mowing detection under varying environmental conditions.

With this study, we aimed to fill this research gap by systematically evaluating the performance of input features derived from dense time series of S1, S2, and L8 data for detecting mowing events. Therefore, we derived the most common optical and SAR input features from the time series that have been proven useful in previous studies for the detection of mowing events. We used a comprehensive set of reference data for meadows in three different biogeographic regions in Germany to train and evaluate a 1D-CNN model that enables the exploitation of the temporal domain of the satellite data (Wang et al., 2017) with all possible combinations of the input features. Based on this setting we investigated the following three research questions:

1. What is the best combination of optical and SAR input features for the identification of mowing events in dense annual time series?
2. How well can we predict measures of mowing intensity (mowing frequency, date of mowing) from the best input feature combination?
3. Are there temporal and/or regional differences and similarities in the accuracy of the predictions that allow for a generalization of the overall findings?

2. Study area and data

2.1. Study area and reference data

We used reference data from three sites in Germany that are part of the Biodiversity Exploratories (BE) project (Fischer et al., 2010). The BE include 150 standardized grassland parcels that are intensively studied since 2006 regarding their functional biodiversity. The three sites are Schorfheide-Chorin (SCH), Hainich-Dün (HAI), and Schwäbische Alb (ALB) with 50 parcels each (Fig. 1). All three sites are within the continental climate zone Dfb according to Köppen & Geiger, with a humid climate and a warm summer (Beck et al., 2018). The three different regions represent different agricultural management systems and soil qualities.

Fig. 1. Location and bounding-boxes of the of the three study sites.

The BE are focused on land use gradients in grasslands and the grassland parcels that are studied range from extensive to relatively intensive management. The 50 parcels per site consist of pastures that are grazed only, meadows that are mown only, and mown pastures where both activities occur. For our study, we considered all parcels indicated as meadows from 2017 to 2019, which yielded a reference dataset of 257 mowing events on 64 distinct parcels. We digitized the parcel boundaries manually using high-resolution aerial imagery. A buffer to the boundary was omitted to exclude pixels influenced by the boundary and adjacent land cover. The years 2018 and 2019 were dominated by very warm and dry conditions compared to the long-term average, while 2017 was less warm and more humid. Detailed information is provided in Table 1 and Fig. 2.

For each parcel, comprehensive information about the grassland management practices was provided, including the date of mowing, respective cutting heights, grazing intensity, and fertilization practices among many others that are recorded annually in accordance with the farmers (Vogt et al., 2019). The three sites represent the variations of cultivation and management types of grassland in Germany and differ in their management and land use intensity (Fischer et al., 2010). ALB has the highest land use intensities with a mowing frequency of up to five times per year (average 2.4), while the mowing frequencies in HAI and SCH are relatively low with a maximum of 2 mowing events per year, respectively (average 1.3 and 1.5 for HAI and SCH).

2.2. Remote sensing imagery

2.2.1. Sentinel-1

We used two different SAR-based parameters as the basis for this study, namely the interferometric coherence and the gamma naught (γ^0) backscatter coefficient. Data were acquired by the S1 constellation that operates in C-band (5.4 GHz, 5.5 cm) with dual-polarization primarily in VV (vertical transmit and vertical receive) and VH (vertical transmit and horizontal receive). We chose the standard acquisition mode of S1, the interferometric wide swath (IW), which covers a swath of about 250 km (Torres et al., 2012). For the estimation of the interferometric coherence, the full information of the SAR acquisition is required, including both amplitude and phase information. For S1, this information is included in the Single Look Complex (SLC) product type. To process the SAR backscatter coefficient, the phase information of the acquisition is not required. We therefore used the Ground Range Detected (GRD) product type, which is comparably smaller in size.

Since the S1 constellation consists of the two satellites S1A and S1B, the interval between S1 acquisitions from the same orbit is 6 days. We chose a single relative orbit for each site that covers all respective

Table 1
Geographical and meteorological characteristics and the mowing frequency of the study sites (Source: Deutscher Wetterdienst).

	Schwäbische Alb	Hainich-Dün	Schorfheide-Chorin
Elevation [m]	670–785	250–450	15–75
Avg. mowing frequency [a^{-1}]	2.4	1.3	1.5
Avg. parcel size [ha]	3.6	3.7	2.9
Prevailing grassland species	*Alopecurus pratensis*	*Poa pratensis aggr.*	*Poa trivialis*
Mean temperature [°C]			
1981–2010	7.3	8.4	9.0
2017	8.0	9.2	9.8
2018	9.0	10.0	10.6
2019	8.5	9.8	11.0
Precipitation [mm a^{-1}]			
1981–2010	1000	740	530
2017	1030	810	740
2018	770	490	420
2019	970	610	460

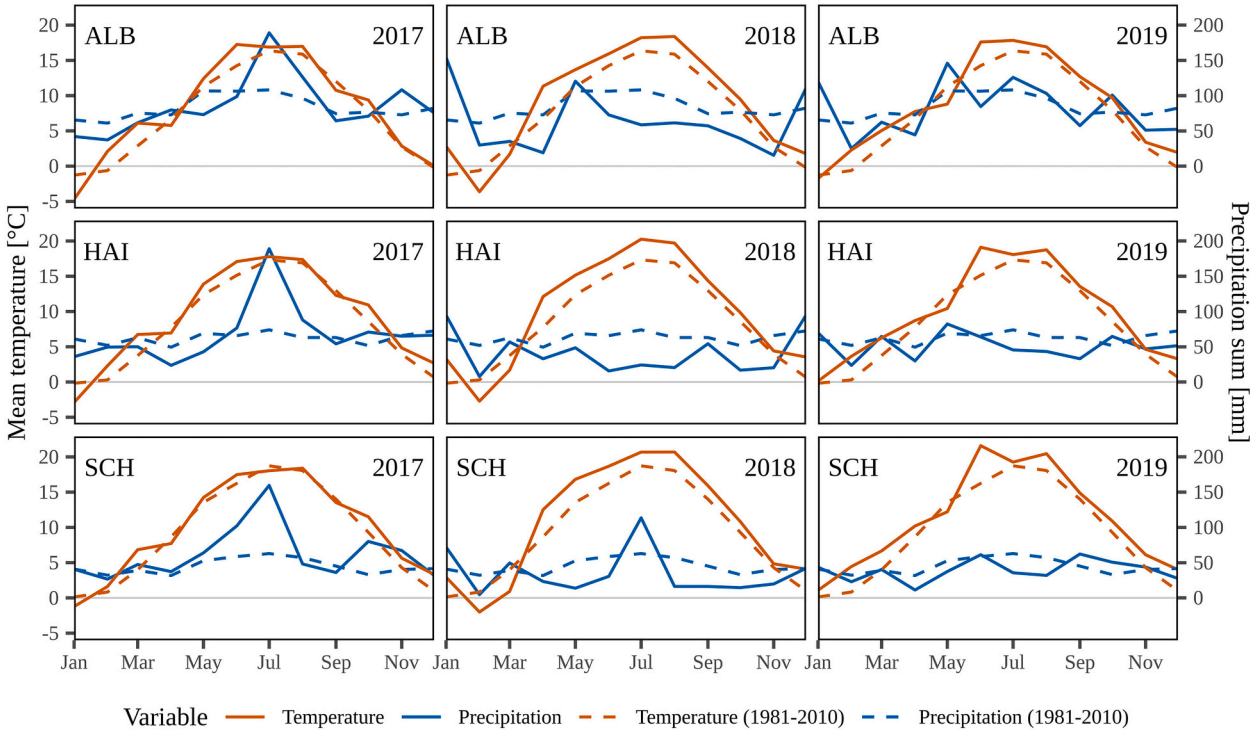

Fig. 2. Variation of temperature and precipitation in the study sites during the studied years and long-term average (1981–2010) of the respective site (Source: Deutscher Wetterdienst).

parcels. This ensures a consistent acquisition geometry of the time series. Furthermore, we only selected ascending orbits, as they are acquired in the late afternoon in local time. This reduces the risk of influence from dew. The chosen orbits were 117 for ALB and HAI, and 146 for SCH. For each of the chosen orbits, we acquired all scenes for the months March to November from 2017 to 2019 for analysis to cover the whole vegetation period in the respective years. This resulted in 414 S1 scenes in total for all three sites. We accessed the S1 data through the Copernicus Data and Exploitation Platform - Deutschland (CODE-DE; Benz et al., 2020).

2.2.1.1. Coherence preprocessing. The magnitude of the interferometric coherence describes the degree of correlation between two complex radar images (Moreira et al., 2013). Besides instrument and acquisition parameters, it is affected by topography and properties of the object that is imaged, e.g., the soil and vegetation structure. We carried out the preprocessing steps to compute the interferometric coherence based on the S1 SLC data using the Graph Processing Tool (GPT) of the Sentinel Application Platform (SNAP). For each image, we selected the corresponding image acquired six days earlier. For both, we updated the exact orbit position. We then coregistered the image pair using back-geocoding to ensure that the image-pairs align with subpixel accuracy. Afterward, we estimated the coherence using a window size of 2 pixels in azimuth and 10 in range direction, which is the SNAP default value, to obtain approximately square pixel size. Finally, we merged the individual bursts of the coherence image followed by a terrain correction to a pixel size of 10 m using the Shuttle Radar Topographic Mission (SRTM) 3 arc-second global digital elevation model (DEM; Farr et al., 2007).

2.2.1.2. SAR backscatter preprocessing. To process the SAR backscatter coefficient, we first applied border and thermal noise removal to the S1 GRD scenes. Then, we calibrated and radiometrically flattened the data to obtain γ^0 backscatter coefficient in VV and VH polarization. Gamma naught represents the ratio between the incident power and the scattered power for a reference area that is perpendicular to the line of sight from the sensor to an ellipsoidal model of the ground surface (Small, 2011). We terrain corrected and resampled the images to a pixel size of 10 m using the SRTM DEM. Finally, we converted γ^0 from linear scale to dB. Again, we used the GPT for these steps.

Furthermore, to exploit the information content of the backscattered signal in both polarizations, we calculated the backscatter cross-ratio (CR):

$$CR = \gamma^0_{VH}[dB] - \gamma^0_{VV}[dB] \quad (1)$$

which was observed to be strongly affected by structural changes in crops like winter cereals (Holtgrave et al., 2020; Schlund et al., 2021; Vreugdenhil et al., 2018). Moreover, Schlund and Erasmi (2020) reported that the CR stays relatively stable in dense time series over longer periods when agricultural areas are observed, because the effect of terrain and soil properties on the radar signal is similar for both polarizations and, thus, the ratio or difference (in dB) reduces the impact of these factors on the CR signal.

Additionally, we calculated second-order texture metrics for γ^0 based on the grey-level co-occurrence matrix (GLCM) (Haralick et al., 1973). The GLCM summarizes the relative frequency distribution of neighboring pixel values. It can be used to augment the input data for several tasks such as image classification and mowing event detection (Haralick et al., 1973; Taravat et al., 2019). As stated by Hall-Beyer (2017), many of the metrics calculated from the GLCM are highly correlated. We therefore only derived a subset of the measures from the GLCM in this study, that we expected to be least correlated and to provide the most information gain (Table 2). For the generation of the GLCM, we quantized the γ^0 pixel values in dB to 32 distinct grey-levels. We used a moving window of 7 × 7 pixels and aggregated the results for all four directions (0°, 45°, 90°, and 135°).

2.2.2. Sentinel-2 & Landsat 8

S2 acquires imagery in 13 bands with pixel sizes from 10 m to 60 m. The bands relevant for this study were band 4 (red, 665 nm central wavelength) and 8 (near-infrared, 833 nm), both in 10 m resolution. L8

Table 2
Texture measures used in this study (only for S1 γ0; Hall-Beyer, 2017). i and j refer to the respective columns and rows of the GLCM, $P_{i,j}$ is the probability of occurrence for the value pair i and j, N is the number of grey values, and μ and σ denote the mean and standard deviation of the GLCM.

Texture Measure	Formula
Homogeneity (Hom)	$\sum_{i,j=0}^{N-1}\left(\frac{P_{i,j}}{1+(i-j)2}\right)$
Entropy (Ent)	$\sum_{i,j=0}^{N-1} P_{i,j} - \left(lnP_{i,j}\right)$
Correlation (Cor)	$\sum_{i,j=0}^{N-1} P_{i,j}\left[\frac{(i-\mu_i)\left(j-\mu_j\right)}{\sigma_i\sigma_j}\right]$

has 12 bands with pixel sizes from 15 m to 100 m. The corresponding bands to S2 were band 4 (red, 655 nm) and band 5 (near-infrared, 865 nm), both with 30 m pixel size. We obtained L8 as Level-L1TP and S2 as Level-1C data. For further analysis, we considered all available scenes that cover at least one parcel in the study period and have a cloud coverage of less than 75%, which resulted in an overall of 465 scenes for the three study sites. We note that all three sites are within areas where S2 orbits overlap, resulting in two acquisitions each five days instead of one. We corrected all data for radiometric and geometric effects using the Level 2 processing system in FORCE (Frantz, 2019). Clouds, including their shadows, were masked out using the Fmask algorithm (Frantz et al., 2018; Zhu et al., 2015; Zhu and Woodcock, 2012). We organized the data in a data cube structure for which they were tiled and reprojected. We adjusted the pixel size of L8 to the size of S2 using nearest neighbor resampling.

Studies based on optical data so far focused on vegetation indices. The rationale of the most common optical vegetation indices is comparable and mostly based on the relation between VIS and NIR reflectance. The temporal signatures of different vegetation indices of grasslands mostly differ in the range of the index values and are not expected to have a considerable impact on the features of a time series that are relevant in our study (e.g. position of local extrema, inflection points, succession of the signal in general). Halabuk et al. (2015) confirmed the similarity and replaceability of NDVI and EVI for mowing detection and even found the NDVI to yield slightly higher accuracies compared to EVI. Most other studies on the subject used the NDVI. On this basis, we decided to use NDVI and calculated it as follows:

$$NDVI = \frac{(\rho NIR - \rho RED)}{(\rho NIR + \rho RED)} \quad (2)$$

where ρNIR is the reflectance in the near-infrared band and ρRED is the measured reflectance in the red band of the respective satellite (Tucker, 1979).

3. Methods

3.1. Time series composition

The analysis concept in our study builds on the parcel level. Thus, the first step in the time series generation process was to derive median values per parcel for each observation using the meadows' parcel boundaries. A technical prerequisite for the used 1D-CNN classification algorithm is an equidistant time series data set. Since the SAR data already provided an equidistant time interval of six days, we linearly interpolated the NDVI time series to match the dates of the SAR observations. For this step, we fused the NDVI time series from S2 and L8 and treated them like a single data set. An example is shown in the appendix (Fig. A1).

To smooth the time series, we applied a Savitzky-Golay filter to each parcel and year (Savitzky and Golay, 1964) for all optical and SAR features. After iterative testing, we chose a filter order of 5 and a filter length of 7 observations to achieve the best trade-off between remaining noise and smoothing out important information. To test the impact on the result, we ran each model with and without smoothing.

As the last step of time series generation, we applied a linear normalization to the values per input feature, meadow, and year to the value range between zero and one. This improved the comparability of the approach for the different parcels, sites, and years and ensured that all of the input parameters are in the same value range, which can ease the learning process during model training (Bishop, 1995).

3.2. Sequencing & labelling

For the classification approach used in this study, the input dataset had to be transferred to the form of a supervised classification problem. This form consists of multiple samples that hold both a number of features and a categorical label or class. In the case of this method, one sample corresponded to a period of time for a given grassland parcel. The values of the remotely sensed parameters within the period formed the features of each sample. Lastly, the assigned label indicated whether a mowing event occurred during the period or not.

In order to achieve this format, we applied the moving window approach to split the time series into a number of short sequences (Dietterich, 2002). This approach includes the stepwise moving of a window over the time series with a given size and creating new sequences out of the values that fall into the window at each step. Subsequently, we labeled the sequences as *mown* or *not mown* depending on whether a mowing event occurred at the midpoint of the sequence (Fig. 3).

We set the number of the steps by which the window is shifted, the so-called stride, to one timestep (6 days). This ensured, that every observation occurred as the middle time step of exactly one sequence and therefore represents a potential date that can be classified as *mown*. The size of the sliding window, i.e., the number of observations before and after the middle time step of a sequence, controls how much information is available to the model before and after an observation that is to be classified. We tested window sizes of 4, 5, and 6 leading to sequences of 9, 11, and 13 observations, respectively. With regard to the observation interval of six days, these window sizes include 24, 30, and 36 days before and after the middle time step, respectively. To simplify matters, we refer to the window sizes as ±4, ±5, and ±6 hereafter.

Since the reference mowing events do not necessarily fall on the observation interval of the created time series, we assigned the label *mown* to the sequence whose middle step happened on or after the mowing event. However, we applied this displacement of the dates only for the labeling process; the reference data used for the evaluation of the predicted mowing dates remained unaffected by this step.

3.3. Deep learning model

Machine learning algorithms are widely used for classification or regression tasks in remote sensing applications (Khatami et al., 2016; Yu et al., 2014). In the last years, especially Artificial Neural Networks (ANN) started to be applied more and more and achieved competitive results (Ma et al., 2019). A simple form of an ANN is the multilayer perceptron (MLP). MLPs are well-performing in various tasks and were already used for the detection of mowing events in time series of satellite data by Taravat et al. (2019). For a detailed overview of this subject, we refer the reader to Goodfellow et al. (2016).

MLPs share an important lack among many other machine learning algorithms. MLPs do not exploit the sequential order and in this case the temporal dimension of the data. The order of the input values is ignored during the training process as long as it is consistent in any sample (Pelletier et al., 2019). Convolutional Neural Networks (CNN) can overcome this shortcoming and are a well-suited method to exploit the temporal dimension of data (Di Mauro et al., 2017; Pelletier et al., 2019; Zhong et al., 2019). As the name suggests, CNNs convolve the input data in a way, that consecutive values are combined, and higher-level

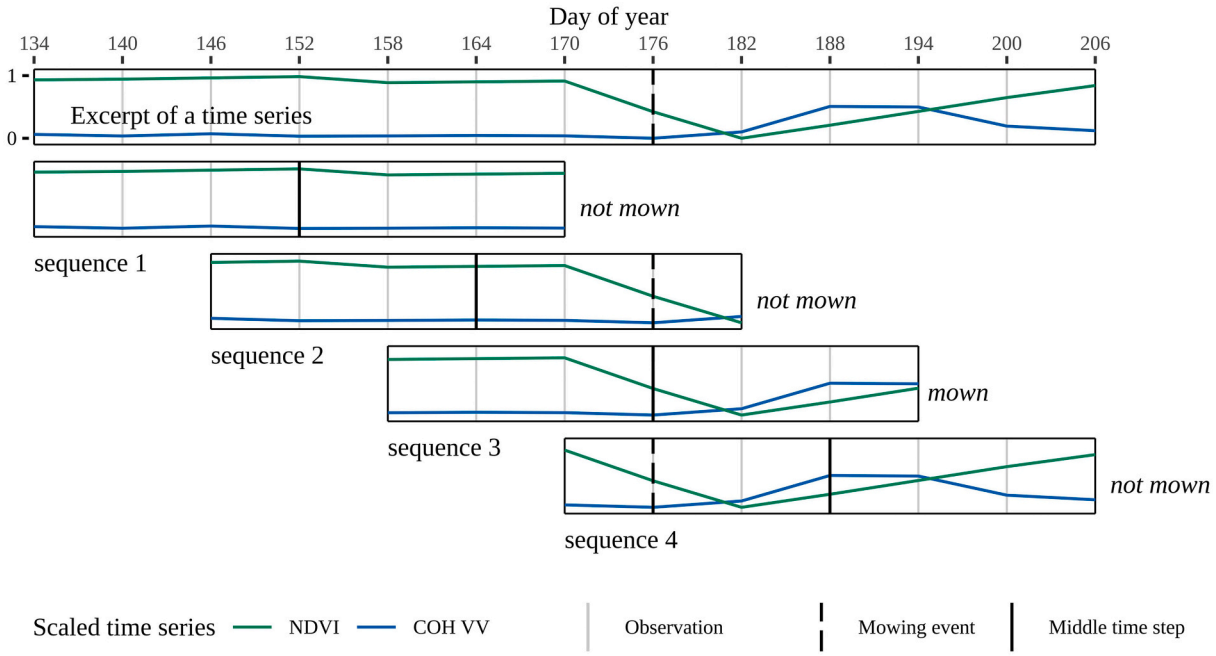

Fig. 3. Exemplary excerpt from the time series for a parcel in 2019 that is segmented into four shorter sequences using the moving window approach. Here, a window size of ±3 observations (±18 days) originating from the middle time step is shown. The window is moved over the time series with a stride of 2 observations (12 days). Depending on whether there was a mowing event on the middle time step of a sequence, it is labeled as *mown* or *not mown*.

features are extracted. This makes them a frequently applied method in the domain of deep learning (Kattenborn et al., 2021).

3.3.1. Implementation

In this study, we used an adaption of a 1D-CNN proposed by Wang et al. (2017) to classify the labeled sequences into mown and not mown. In a comprehensive study, Wang et al. (2017) tested eleven different algorithms on the UCR Time Series Classification Archive (Chen et al., 2015) that consists of 44 datasets belonging to time series classification tasks. The 1D-CNN performed best, while state-of-the-art models for time series classification, as well as an MLP and a Residual Network (ResNet), were among the inferior algorithms.

We implemented the 1D-CNN with Keras (Chollet and others, 2015) and TensorFlow (Abadi et al., 2016) as backend using the R interface to Keras (Allaire and Chollet, 2020; R Core Team, 2021). The original model was developed to classify univariate sequences into multiple classes. To match the given classification problem, we adapted the model to run on multivariate input data and give binary classification output. Since the initial model was built to classify time series from other domains with greater lengths and higher densities, we further reduced the model's depth in terms of convolutional layers (ConvLayers) and the size of the applied filter kernels. This reduction decreased the risk of over-fitting and speeded up the training time significantly.

The final model proposed here consists of two convolutional layers with kernel sizes five and three, respectively. The output of the convolutional layers was zero-padded to remain the same size as the input. Batch-normalization and a rectified linear unit (ReLU) activation function were applied after each layer. The ReLU function converts all negative values to zero, while positive values are preserved. The utilization of activation functions allows to find non-linear relationships and model non-linear problems (Graves, 2012). The two convolutional layers were followed by a global average pooling and finally a single densely connected and sigmoid-activated neuron to give the final output. To find the best number of filters applied at each convolutional layer, we tested different filter numbers. We chose the values 64, 128, and 256 as possible filter numbers for the first layer, while the second layer always had twice the number of filters. The schematic architecture of the used model is shown in Fig. 4.

3.3.2. Training

For model training, we split the input data set into three parts: training, validation, and test data. The training dataset was used to train the model, while after each training epoch, the current performance was evaluated on the validation set to monitor the convergence and trigger learning parameter changes. The test set was not used by the model during the training and is applied to estimate the final model performance. The splitting was conducted along randomly chosen meadows. This ensured, that time series from different years for the same meadow did not appear in the test and training set at the same time. Additionally, it was always ensured that the same number of meadows was chosen from each of the three test sites in order to minimize the variance between the folds.

Within the training of ANNs, the loss function measures the error between the predictions and the reference data and represents the objective function to be minimized. The loss function we chose for the training process here was binary cross-entropy, which considers not only the correctness of the predicted class but also the certainty of the respective prediction. This makes it the function of choice for most DL models (Goodfellow et al., 2016). We used the Adam optimization algorithm (Kingma and Ba, 2015) to minimize the loss function with a batch size of 64 and an initial learning rate of $1e^{-4}$. The batch size is the number of samples that are fed to the model before updating the weights, the learning rate controls the degree to which these weights are updated. The other parameters were left as the Keras default values.

Furthermore, we implemented an early stopping mechanism that stops the training after not improving the loss function for the validation set for ten epochs to avoid over-fitting. To overcome plateaus of the loss function without improvement, the learning rate was multiplied by 0.1 if the loss function for the validation set did not improve for five epochs. We set the maximum number of epochs to train a model to 100.

A problem with the classification of mowing events is the frequency of their occurrence. The studied meadows are mown between one and four times per year making the class of mowing events very rare compared to the class of not mown sequences. If a model is trained on

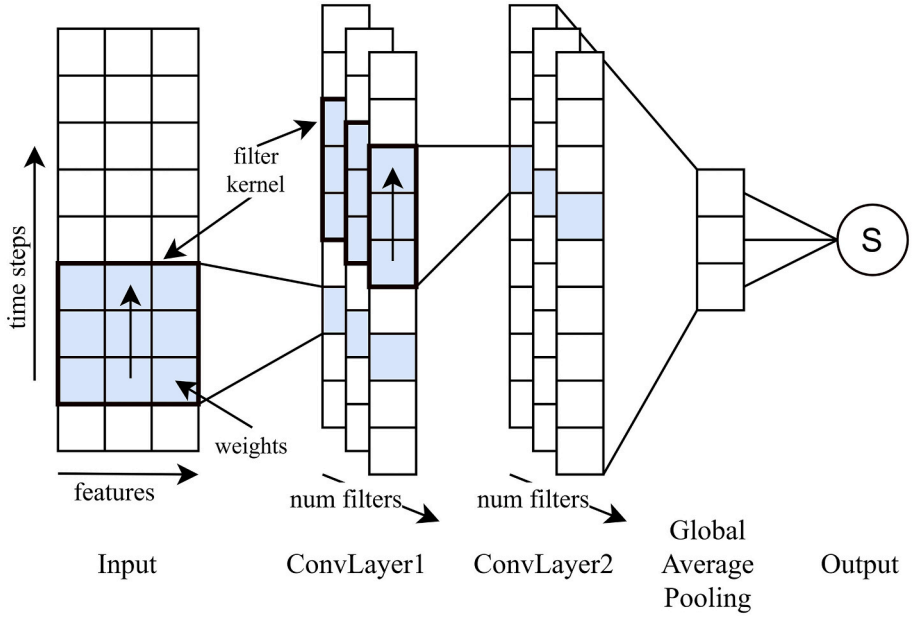

Fig. 4. Schematic architecture of the proposed 1D-CNN adapted from (Wang et al., 2017). The model shown here consists of an input layer with three input features of nine timesteps each. The input layer is followed by two convolutional Layers (ConvLayers) with three filters, each with a filter kernel of length three. Next is a global pooling layer, summarizing each filter to its average. The last layer is a densely connected and sigmoid-activated output neuron.

such a highly imbalanced dataset, it is very likely to develop a prediction bias towards the majority class. The random oversampling method (ROS) randomly duplicates the samples belonging to the minority class until the class-distribution is balanced, whereas the values of the sequences remain unchanged. Despite its simplicity, ROS has shown to perform competitively to other state-of-the-art methods when applied for the training of CNNs (Buda et al., 2017). We therefore applied ROS to the training set before the training of the model, while this was not done for the validation and test data.

3.4. Evaluation

With this study, we aimed to reveal the best combination of remotely sensed input features for the detection of mowing events. Exhaustive testing of feature combinations, however, is time and resource-consuming. We therefore subsumed several of the processed time series into one feature each. For example, the time series for coherence in VV and VH polarizations were subsumed to the feature COH and the time series of homogeneity, entropy, and correlation in VV and VH polarizations to the feature GLCM (Table 3). I. e., where the feature COH is included in a model, the time series of coherence in VV and VH are meant, and so on.

Based on these five features, we formed all possible combinations from one up to all five features. This resulted in 31 different input-sets (five single-input sets, ten double-input sets, ten triple-input sets…). We identified the best performing input-set by a comparative evaluation after training a model for each input-set.

Different hyperparameters such as smoothing of the time series, the window size chosen for the moving window approach, or the number of filters within the ConvLayers may lead to the optimal results for each input-set under test. We therefore conducted a grid-search, testing all possible combinations to find the best hyperparameters for every input-set (Table 4).

The results slightly vary with each training cycle due to the stochastic behavior of the learning process in DL. Furthermore, the chosen split into training, validation, and test data can influence the results and introduce bias. We conducted a repeated ten-fold cross-validation to tackle both problems. For every model, we split the whole dataset into ten equally sized folds. We then trained the model ten times while every time a different fold was chosen as the test set. The nine remaining folds became the training set (80%) and validation set (20%). We repeated this procedure three times to further account for the stochastic effects of the fold definition. We summarized the results of the trained models per model setup from the grid-search afterward (Fig. 5).

Before evaluation, we conducted one more postprocessing step. Due to the ability of CNNs to extract higher-level information, successive sequences for the same meadow may appear similar to the CNN. Operations like global pooling can even increase this. Hence, it was very likely that not only the sequence with the mowing event in the middle time step is classified as mown but also one or multiple previous and following sequences. To solve this ambiguity, we clustered successive predictions of the model and defined the mean of the dates as the final

Table 3
List of the five tested features and the individual time series they subsume.

Feature	Individual time series
BSC	γ0 backscatter coefficient VV
	γ0 backscatter coefficient VH
COH	interferometric coherence VV
	interferometric coherence VH
CR	backscatter cross-ratio
GLCM	homogeneity VV
	homogeneity VH
	entropy VV
	entropy VH
	correlation VV
	correlation VH
NDVI	Normalized Difference Vegetation Index

Table 4
Hyperparameters tuned in the grid-search and respective values. Window size is expressed in observations before and after the middle time step of the time series sequences. Filter numbers refer to the first and second ConvLayers in the CNN, respectively.

Hyperparameter	Values
Smoothing	$\in\{True, False\}$
Window size	$\in\{\pm4, \pm5, \pm6\}$
Filter numbers	$\in\{(64,128);(128,256);(256,512)\}$

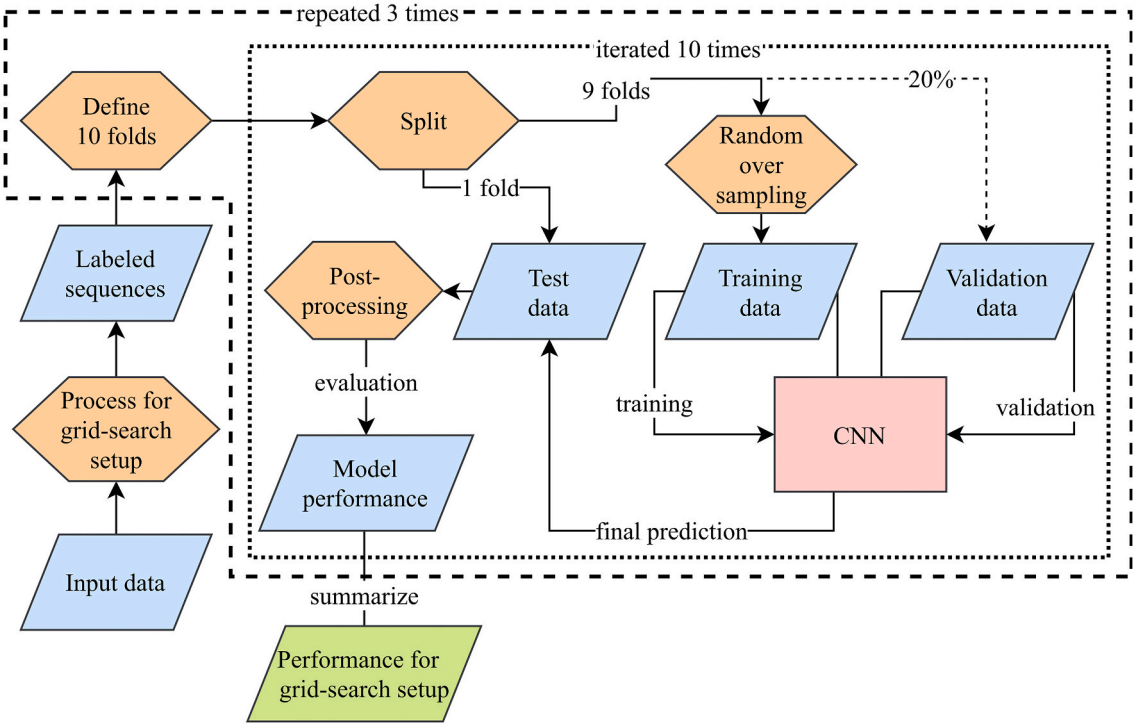

Fig. 5. Model training cycle and validation scheme including cross-validation (dotted line) and the three repetitions (dashed line).

prediction.

The grid-search resulted in 558 (31*2*3*3) different model setups consisting of input-sets and the other hyperparameters defined by the grid-search. For each of the model setups, we conducted three 10-fold cross-validations which made 16,740 training cycles in total. We then averaged the results of the cross-validation for each of the 558 model setups. Since the main goal of the study was to identify the best performing input-set, only the best performing hyperparameter combination for the 31 different input-sets in terms of the F1-Score was considered hereafter.

The following model evaluation included the overall accuracy assessment, the number of mowing events that were detected (mowing frequency) as well as the dates of the detected individual mowing events. First, we conducted an accuracy assessment of the classification results for the entity of all mowing events using binary classification metrics. The recall metric states how many of the mowing events in the test set were detected correctly by the model. The precision measures how many of the events predicted by the model were correct. To sum up both, we chose the F1-Score, as it is a better choice for imbalanced classification problems compared to the frequently chosen overall accuracy (Sokolova and Lapalme, 2009).

We calculated the metrics as follows:

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \tag{3}$$

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \tag{4}$$

$$F1 - Score = 2*\frac{Precision*Recall}{Precision + Recall} \tag{5}$$

The F1-Score reaches from 0, where either the precision or the recall of the model is zero, to 1, where both precision and recall are 1, which would mean that all events are detected without any false positives. Due to the sensing interval of the satellites and occurring cloud gaps in the NDVI time series, the probability of predicting an event to the exact day was low. Therefore, a tolerance of twice the time series interval (12 days) was considered to be sufficient to evaluate a prediction as true. Based on the F1 score, the best input set was selected, to which further evaluations were limited.

Second, the predicted mowing frequency for the reference meadows was evaluated. The mowing frequency is the sum of the mowing events within a time period (here: one year). It is a measure of the general management intensity of a meadow. We derived the predicted mowing frequency by summarizing all predictions made by the model for each meadow and year regardless if they were true or false. Various metrics were used for this purpose, namely the mean error (ME), mean absolute error (MAE), and normalized mean absolute error (nMAE):

$$ME = \frac{1}{n}\sum_{i=1}^{n} \widehat{Y}_i - Y_i \tag{6}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n} \left|\widehat{Y}_i - Y_i\right| \tag{7}$$

$$nMAE = \frac{1}{n}\sum_{i=1}^{n} \left|\frac{\widehat{Y}_i - Y_i}{Y_i}\right| \tag{8}$$

where n is the number of meadows, $\widehat{Y}_i$ is the predicted mowing frequency for each meadow, and Y_i is the reference mowing frequency.

The ME is the plain average of all residuals. While it is not suitable to determine the magnitude of the error, it indicates a general tendency of the model to over- or underestimate the target value. The MAE measures the average error regardless of its sign and gives the magnitude of the error in the same unit as the prediction, in this case, the mowing frequency. Since the mowing frequencies of the meadows in the reference data vary, the nMAE was calculated as well. In addition to the MAE, the nMAE is related to the reference mowing frequency of the predicted meadow and therefore allows for a better comparison between meadows or test sites of different management intensities.

Finally, the exact prediction of the date of mowing was evaluated. The date of mowing is of relevance for applications that are related to the checking of compliance measures in agricultural policy as well as for

the monitoring of nature protection schemes in grassland in general. This was done by calculating the MAE between the predicted and the reference date for each individual mowing event instead of the mowing frequency. This gives an overview of the temporal error of the predictions.

In summary, Fig. 6 illustrates the processed features from the optical and SAR time series together with the main attributes of the reference data that were used for the evaluation of the predicted mowing events. For reasons of clarity, not all tested remote sensing features are shown. The four shown graphs represent the most common patterns of true/false positive and negative predictions that are likely to occur.

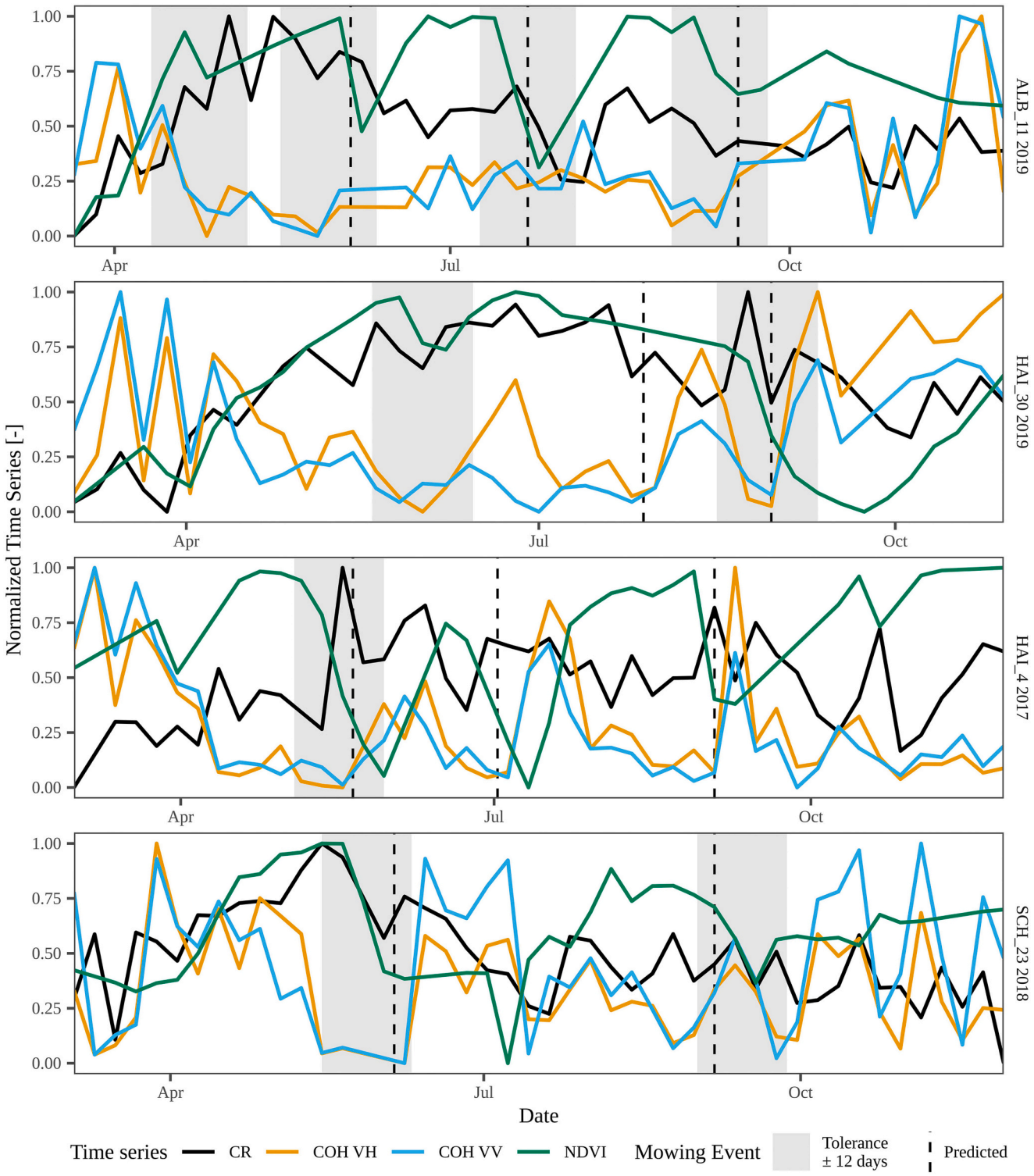

Fig. 6. Time series for a subset of Sentinel-1/-2 and Landsat-8 features on four different meadows that represent frequent patterns of true and false predictions. The reference events took place at the center of the grey-shaded areas.

4. Results

4.1. Accuracy assessment

The main results of the overall accuracy assessment of the grid-search are summarized in Table 5. For each tested input-set, only the best hyperparameter setup in terms of the F1-score is shown. A full list of the performance parameters for all 588 model setups is compiled in the appendix to this publication (Table A1). It can be observed from the summary that the model based on NDVI, BSC, and COH was able to detect the highest number of events in the reference dataset correctly with a recall of 88.8% (true positives). The least false detections (false positives), were achieved with the model based on NDVI, COH, and GLCM, with a precision of 87.1%. However, the best trade-off between detecting a high number of events and predicting only a few false positives was reached, when NDVI, CR, and COH were used in combination, which is stated by the highest F1-Score of 0.839. This input-set was therefore identified as the best-performing combination of input features.

When looking at the input-sets consisting of only one feature, NDVI performed superior to the other features with an F1-Score of 0.762. All SAR-based features yielded lower F1-Scores, led by COH (0.674) and followed by BSC (0.576) and CR (0.519). GLCM reached the worst result with an F1-Score of 0.363.

The models built solely on the SAR-based input features resulted in lower F1-Scores than for any input set that included the NDVI. While improvements through the combinations of multiple SAR-based features were observed, not a single one outperformed the NDVI alone. However, adding SAR-based parameters to the NDVI improved the performance in most cases. Only four combinations of NDVI and SAR-based features reduced the performance of the NDVI used alone. These four combinations included GLCM without exception.

BSC alone outperformed CR. Yet, when combined with NDVI and COH, the input-set containing CR yielded a clearly higher F1-Score (0.839) compared to the input-set containing BSC (0.828).

When combined with NDVI and BSC, CR and COH reached the same F1-Score. However, the input-set including COH was able to detect 88.8% of the mowing events compared to 82.9% (NDVI, BSC, CR). At the same time, the latter yielded more precise predictions with 83.2% of the events predicted by the model being true positives, compared to the input set of NDVI, BSC, and COH with a precision of 77.8%. In every case, adding GLCM to an input-set reduced the F1-Score. The input-set including all tested input features yielded lower F1-Scores than input sets with fewer features. The average standard deviation of the F1-Score per model setup was 0.083.

Due to its best performance, the remainder of the results is presented only for the model based on NDVI, CR, and COH. The confusion matrix of the predictions is presented in Table 6.

4.2. Mowing frequency

Meadows with a mowing frequency of up to two cuts per year showed a positive ME of 0.31 and 0.083 (Table 7). The sign of the ME changed for meadows mown more than two times. Meadows with three cuts showed an ME of −0.3 and meadows with four cuts a much lower ME of −1.44. This indicated an overestimation of mowing events by the model on meadows mown up to two times, while meadows mown three or four times tended to be underestimated.

The MAE was in a comparable order of magnitude for meadows with one to three mowing events per year, with 0.369, 0.321, 0.420 events per year, respectively. Yet, it was considerably higher for meadows mown four times a year with 1.44.

Since the three study sites differ in their average mowing frequency (compare Table 1), we compared the predicted mowing frequencies between them (Fig. 7). In terms of ME, ALB was the only study site where the yearly number of mowing events tended to be underestimated,

Table 5
Main results of the grid-search sorted by F1-Score. For each combination of input features (feature-sets), the best model setup in terms of F1-Score is shown. Filter number denotes the number of filters in the first ConvLayer of the CNN. The highest value of each metric is printed in bold.

Input Features	Window size	Smoothing	Filter number	Recall [−]	Precision [−]	F1-Score [−]
NDVI, CR, COH	±5		128	0.859	0.824	**0.839**
NDVI, BSC, CR	±5	Yes	128	0.829	0.832	0.828
NDVI, BSC, COH	±4		64	**0.888**	0.778	0.828
NDVI, BSC, CR, COH	±5		256	0.798	0.864	0.827
NDVI, COH	±4		128	0.855	0.788	0.815
NDVI, BSC, CR, COH, GLCM	±4	Yes	64	0.798	0.803	0.799
NDVI, CR, COH, GLCM	±4		64	0.800	0.795	0.795
NDVI, CR	±5	Yes	128	0.852	0.735	0.786
NDVI, BSC, COH, GLCM	±4	Yes	256	0.732	0.850	0.781
NDVI, BSC	±4		128	0.859	0.716	0.777
NDVI, CR, GLCM	±4		128	0.777	0.770	0.768
NDVI	±6	Yes	128	0.794	0.742	0.762
NDVI, COH, GLCM	±4		256	0.693	**0.871**	0.755
NDVI, BSC, CR, GLCM	±4	Yes	64	0.797	0.720	0.755
NDVI, BSC, GLCM	±4	Yes	64	0.784	0.725	0.748
NDVI, GLCM	±4		64	0.793	0.707	0.739
BSC, CR, COH	±4		128	0.745	0.734	0.736
CR, COH	±5		128	0.756	0.695	0.720
BSC, COH	±4	Yes	256	0.696	0.708	0.698
COH	±4		256	0.696	0.665	0.674
BSC, CR, COH, GLCM	±4	Yes	128	0.647	0.686	0.660
BSC, COH, GLCM	±4		128	0.602	0.709	0.643
CR, COH, GLCM	±5		256	0.568	0.745	0.638
COH, GLCM	±5	Yes	256	0.575	0.662	0.605
BSC, CR	±5		256	0.597	0.584	0.581
BSC	±4	Yes	256	0.639	0.535	0.576
BSC, CR, GLCM	±4		64	0.571	0.492	0.524

(continued on next page)

Table 5 (*continued*)

Input Features	Window size	Smoothing	Filter number	Recall [−]	Precision [−]	F1-Score [−]
CR	±4		128	0.679	0.424	0.519
BSC, GLCM	±5	Yes	256	0.455	0.544	0.487
CR, GLCM	±4		128	0.533	0.448	0.477
GLCM	±6		256	0.369	0.374	0.363

Table 6
Confusion matrix of the best performing model, based on NDVI, CR, and COH with window size ±5, 128 filters in the first convolutional layer of the CNN and no smoothing applied. Values represent the whole-numbered average of the three cross-validations.

Prediction	Reference		Total
	Not mown	Mown	
Not mown	4334	35	4368
Mown	45	214	259
Total	4380	249	4629

Table 7
Mean error (ME), mean absolute error (MAE), and normalized mean absolute error (nMAE) of the predicted number of mowing events grouped by the reference number of mowing events.

Reference mowing frequency [a^{-1}]	ME [a^{-1}]	MAE [a^{-1}]	nMAE [−]
1	0.310	0.369	0.369
2	0.083	0.321	0.161
3	−0.304	0.420	0.140
4	−1.440	1.440	0.360

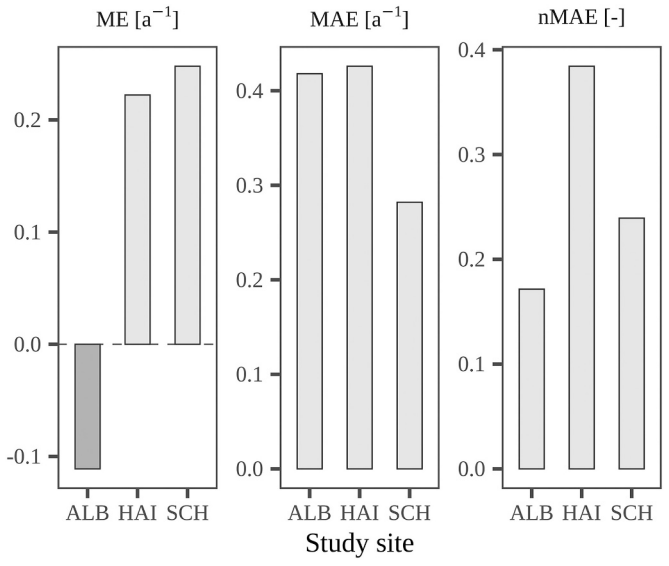

Fig. 7. Mean error (ME), mean absolute error (MAE), and normalized mean absolute error (nMAE) between the predicted and reference mowing frequency.

whereas for HAI and SCH the opposite was the case.

The largest MAE was observed for HAI and ALB and was similar with slightly above 0.4 events per year. For SCH, the MAE was considerably smaller with under 0.3 events per year. To compare the error between the sites, it had to be put into the context of the test sites' mowing frequency, therefore the nMAE was used. The nMAE revealed that the model predicted the mowing frequency with the least nMAE for ALB (17%), followed by SCH (24%), and lastly HAI (38%).

The observed pattern, however, was not stable throughout the full study period (Fig. 8). The underestimation for ALB was clear for 2018 (−0.317) whereas for 2017 and 2019 the ME was close to 0.

Similarly, the MAE and nMAE varied over time for the individual sites. The average MAE and nMAE decreased from 0.52 (2017) to 0.23 (2019) and from 39.1% (2017) to 15.8% (2019), respectively, and thus both have more than halved. Considering the nMAE, the mowing frequency was estimated with the highest error for HAI in 2017 (60%) while the least error was observed for SCH in 2019 (5.6%).

4.3. Date of mowing

Fig. 9 shows the deviation between the day of the year (DOY) from the reference events and the DOY of the respective events predicted by the model based on the best performing input-set from the accuracy assessment (NDVI, CR, COH). A list of the performance parameters for all 32 input-sets is compiled in the annex to this publication (Table A2). When the exact dates were compared, no systematic error depending on the DOY or the number of cuts was observed. The ME was 2.35 days, which shows that the prediction was usually slightly behind the actual event. However, the MAE for all predicted events was 3.79 days, which can be considered low in the context of the observation interval of 6 days. It has to be noted, that these deviations were only calculated for reference events and their nearest matching prediction within 12 days tolerance (compare section 3.4).

The observed MAE for the predicted date of the mowing events did not show a generally applicable pattern for the individual study sites and years (Fig. 10). Yet, the average MAE differed between the sites with HAI having the lowest MAE (3.37 days) and a high variance between the years, and SCH having the highest MAE (4.48 days) with a comparably higher MAE in 2018 (6.41 days). The MAE in ALB was relatively stable throughout the study period with an average of 3.61 days.

5. Discussion

5.1. Main findings

5.1.1. Overall accuracy

With regard to the overall accuracy in detecting mowing events from dense time series of Earth Observation data, our findings were in line with other studies when we trained our models only on single features of optical or SAR data (Halabuk et al., 2015; Kolecka et al., 2018; Taravat et al., 2019). Here, the single use of the NDVI yielded higher accuracies compared to all tested single SAR features and also to all possible combinations of SAR features (F1-score 0.762). This outcome confirms the broad perception in literature that states the general potential of time series of optical visible / near infrared indices for the assessment of phenological patterns and abrupt changes in vegetation state, e.g. by management activities (Gao et al., 2020; Sakamoto et al., 2005; Vrieling et al., 2018; Zeng et al., 2020). We also assume that the combined use of S2 and L8 contributed to the performance of the NDVI in comparison to only using S2 (Kolecka et al., 2018; Stendardi et al., 2019), as already reported by Griffiths et al. (2020). By using data from both sensors, the number of potential observations was substantially increased, which allowed detection of events even in frequently clouded periods when recordings from one sensor alone would not have been sufficient to detect the event. Such a situation where observations from only one sensor could have missed events can also be observed in Fig. A1 in the appendix.

Looking at SAR-only input-sets, the model based on COH achieved the highest overall accuracy of all tested single SAR input-features (F1-score 0.674). Comparable results were reported by De Vroey et al. (2021) who also underlined the importance of repeat-pass coherence time series from Sentinel-1 for detecting mowing events. The general information content of Sentinel-1 coherence data for mowing detection

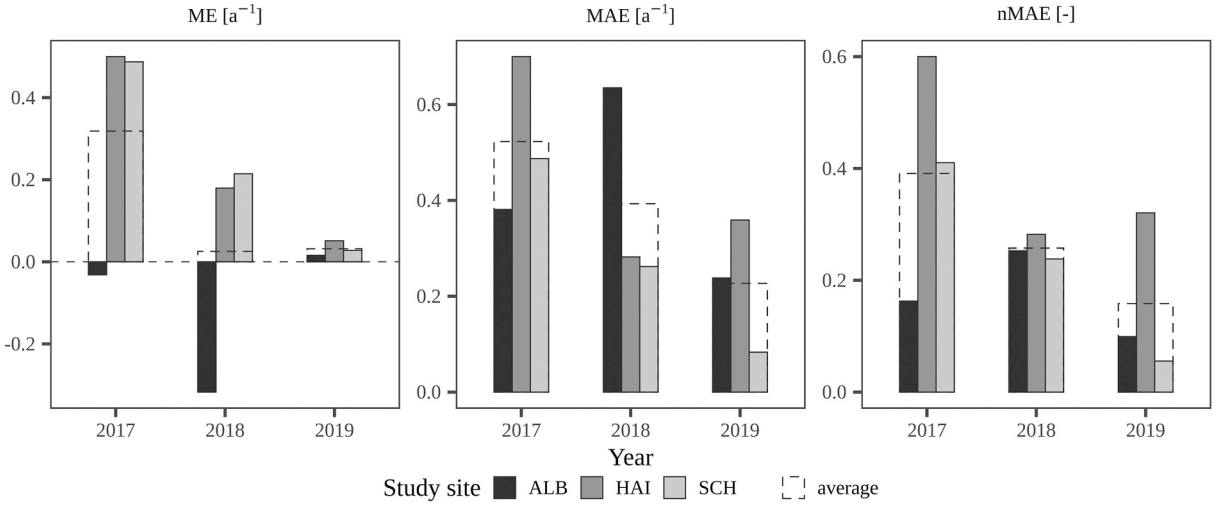

Fig. 8. Mean error (ME), mean absolute error (MAE), and normalizes mean absolute error (nMAE) between the predicted and reference mowing frequency.

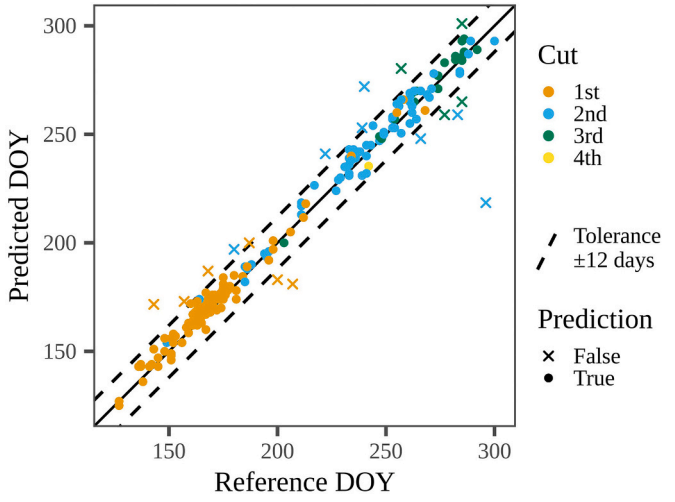

Fig. 9. Reference day of the year (DOY) for the mowing events and the DOY predicted by the model. Predictions for all three sites combined are shown.

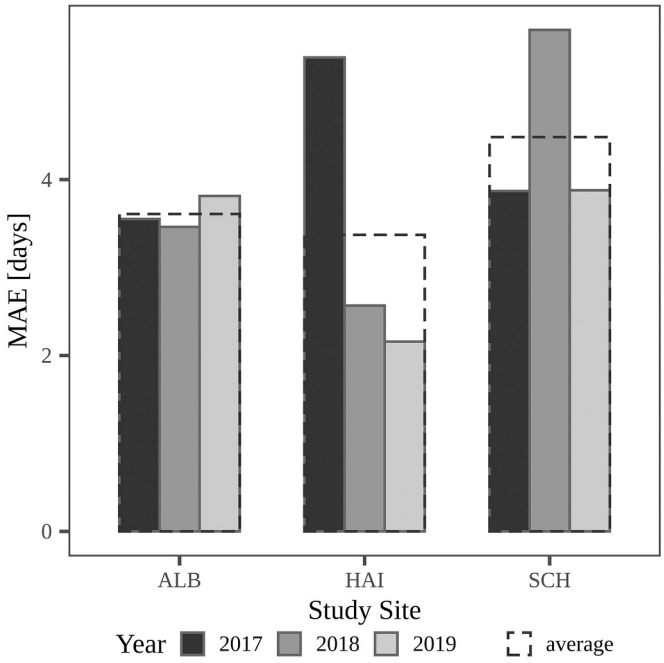

Fig. 10. MAE between predicted and reference date of the individual mowing events by study site and year.

was also investigated by Tamm et al. (2016) who stated a significantly higher coherence values for both polarizations (VH and VV) for up to 36 days after the first image pair after mowing compared to the period before the event. Similar patterns of a sharp increase in coherence after mowing could in general be observed in our study as exemplarily shown in Fig. 6. However, on the other hand, the closer evaluation of the temporal signatures for both polarizations also depicts that the timing and level of increase are variable, which might be a consequence of the revisit cycle of the Sentinel-1 satellites and environmental factors (Tamm et al., 2016). Compared to COH, the other tested SAR features (BSC, CR, GLCM) performed considerably worse with GLCM leading to the overall poorest result (F1-score 0.363). This is also reflected exemplarily by the signatures for CR in Fig. 6 where no clear patterns of signal evolvement are visible after a mowing event.

A considerable increase in the overall detection accuracies was achieved when we combined features from both, optical and SAR data. This general outcome further strengthens the assumption that the combination of optical and SAR input features is beneficial for the performance of mowing detection algorithms, as already implicated by other studies (De Vroey et al., 2021; Stendardi et al., 2019).

Our results revealed that except for input-sets that contained GLCM features, all combinations of NDVI and SAR-based features outperformed models that were built on NDVI or SAR features alone. The best combination of NDVI and SAR features (NDVI, COH, CR) improved the model based on NDVI by more than 10%, and the best model only based on SAR features (BSC, CR, COH) by more than 12%. This clearly demonstrates the synergy of both sensor types. In addition, four out of the top five input-sets with regard to the overall accuracy of feature combinations contained the COH (compare Table 5). This confirms the overall sensitivity of the interferometric coherence to abrupt changes in vegetation physiognomy and phenology that was also reported for cropland management events (e.g. harvest) (Kavats et al., 2019; Nasirzadehdizaji et al., 2021; Schlund and Erasmi, 2020; Shang et al., 2020). This overall improvement in accuracy could give evidence that certain mowing events that are missed by models based on optical data caused by, for example, cloudiness show a response in SAR time series, as suggested by Stendardi et al. (2019). The other way around, events missed by SAR time series, e.g. induced by soil moisture changes

or wind, still might show a detectable signal in the NDVI. Yet, the recall of the best model still indicates that not all mowing events can be detected with the combined data sources. These omissions could in general be a consequence of a combination of confounding factors for both systems, optical and SAR.

With particular regard to the SAR features that are built from the backscatter signal (BSC, CR, GLCM), CR outperformed BSC when included in optical / SAR input-sets. A possible cause is that the learned signal in CR is more unique to mowing events. This goes in line with the findings of Schlund and Erasmi (2020) that CR is more stable than BSC and less prone to short-term signals (shifts and noise) since both polarizations are affected equally by, e.g., changes in soil moisture. Shifts and noise in BSC are at risk to be misinterpreted by the model, which reduces the precision.

The GLCM features derived from the backscatter coefficient even weakened the performance of the input-sets. This is in opposite to the findings of Taravat et al. (2019) who significantly improved their model performance by adding GLCM parameters to their input features. Thus, our results do not confirm their findings and rather imply S1 texture metrics to be unsuitable for mowing detection.

Our findings are based on the average accuracy measures for all cross-validation runs. Even if the standard deviation of the F1-Score indicated noticeable variation in the results within the cross-validation runs, we observed it to be comparably stable throughout all input-sets and the variations seemed normally distributed. We therefore assume the average of the cross-validation runs to be representative for our results.

5.1.2. Measures of mowing intensity

With regard to the prediction of the mowing frequency within one season, we determined a clear positive connection between increasing prediction errors and increasing management intensity. An increasing mowing frequency reduces the temporal distance between individual mowing events and consequently the number of available satellite observations. From the grassland perspective, it leads to varying signals in the remotely sensed time series since the phenology between two cuts and the state of the meadow right before a cut might be different according to the time span since the last cut.

In general, mowing frequency in extensive meadows was overestimated and intensive meadows faced an underestimation. Where the underestimation might be a consequence of the reduced number of time steps between two events, the overestimation could be caused by the long time span within the observed period where no management activities occur. Here, long grass is prone to weather extremes like heavy rainfall, wind, or drought. These extremes could have effects on the temporal signature of the time series that resemble mowing events and, hence, could increase the probability of false detections. This is particularly of relevance for input features of SAR data that are known to be sensitive to changes in the vertical and horizontal structure of the grassland volume (Taravat et al., 2019; Voormansik et al., 2016). The impact of management intensity on the prediction of mowing frequency was also identified by Griffiths et al. (2020). However, their findings were contrary to our results and, in general, showed an underestimation in extensive meadows compared to a general tendency of overestimation in meadows that were mown more than three times.

Considering the inter-annual variation of our results, an overall trend of decreasing prediction errors is noticeable from 2017 to 2019. Overall, the absolute error of the predicted mowing frequency was highest in 2017. In this year, summer precipitation was clearly above the long-term annual mean (compare Fig. 2) and thus, the frequent occurrence of clouds, reduced the number of valid observations in optical time series and led to higher omissions. On the other hand, limitations in the cloud masking process can also lead to false positive detections when the masking process fails to remove a cloud-affected pixel (Griffiths et al., 2020; Kolecka et al., 2018). Another factor that might explain the overall higher errors for the mowing frequency in 2017 is the generally lower number of available observations from Sentinel-2 due to the start of the operational phase in July 2017, only. Hence, the first cut in all meadows could only be detected based on Sentinel-2A data which halved the observation frequency from 5 to 10 days.

This constraint in data availability also had a potential impact on the detection of the exact mowing date of a single event. However, a clear increase in prediction error for 2017 could only be observed in one out of the three test sites. The MAE of 3.79 days over all sites and years can be considered remarkably low with regard to the temporal interval of S1 and the interpolated time series of S2 and L8 ($\Delta t = 6$ days). However, it has to be noted that this error only accounts for correctly detected events (true positives), since predicted events with a distance of more than 12 days from a reference event were assumed to no longer relate to the reference event itself. In comparison, Griffiths et al. (2020) reported a MAE between 53 and 94 days which is a multiple of the error presented in this study. However, it has to be mentioned that this included all detected events, not only true positive predictions. Stendardi et al. (2019) evaluated the end-of-season (EOS) in meadows in Northern Italy and got an MAE of 20 for S1 and 8 for S2 time series, respectively. Considering the lower resolution of the underlying time series (12 days for S1; 10 days for S2) the results are comparable to our results but the numbers base upon a very small sample.

In general, mowing events were detected slightly later than the reference date (i.e. positive ME). This is a consequence of the preprocessing where mowing events from the reference data were assigned to the next following sequence. This necessary technical constraint led to the overall late detections. The time lag is relatively stable throughout the three years and the three sites, indicating that there is no major influence of management intensity or climatic conditions on the prediction of mowing dates.

5.2. Limitations and outlook

A number of limitations and potential error sources of the data and methods that were used in this systematic evaluation of optical and SAR features for mowing detection have to be specified. The most obvious limitation was the inferior potential of all tested SAR features compared to the NDVI in spite of the consistent acquisition interval and the insensitivity of SAR sensors to clouds.

Many studies that use SAR data for vegetation monitoring, reported humidity (atmosphere, plant) and moisture (soil) as possible sources for uncertainties in the detection of vegetation characteristics. This included both backscatter coefficient and interferometric coherence (Grant et al., 2015; Siegmund et al., 2016; Taravat et al., 2019; Voormansik et al., 2020; Zalite et al., 2016). In our study, we minimized the risk of influence by dew on plants in choosing S1 orbits with afternoon local acquisition time. Meteorological data in general were not considered as a feature in our study. For future studies, e.g., the precipitation sum of a number of days before a S1 acquisition could be introduced to the CNN as additional feature and provide a basis for the model to weigh or compensate for changes in SAR features in relation to rainfall (Garioud et al., 2020). For COH, wind is a factor that might cause temporal decorrelation due to movement of the grass and lead to false positives, as mentioned by Tamm et al. (2016) and Voormansik et al. (2020), that could also be tackled with suitable data.

Another aspect that has an impact on predicting mowing events in general is the composition of the reference data that is used for training. In a pre-test to our study, we found that reference data sets that contain more extensively managed meadows introduce a bias in the trained model. This led to a gain in omissions of mown sequences in more intensively used meadows. We accounted for this issue by stratifying the reference data to ensure that meadows from different intensity levels were included in equal proportions in each training set (compare Section 3.3.2). However, even though we used a comprehensive reference data set, we did not cover the far end of highly intensively used grasslands in Germany with up to 6 mowing events. Wadoux et al. (2021) observed

accuracy assesments of spatial predictions conducted with cross-validation to be biased. We must therefore assume that our results are only an estimate of the true error of a map that is produced with the model.

One constraint of the proposed processing and analysis framework was the definition of the parcel as the smallest prediction unit. This enables a direct evaluation of grassland management intensity in terms of agricultural monitoring purposes, e.g., for compliance regulations in agricultural policy. On the other hand, it creates dependencies on the availability of data on field boundaries. These data, however, are available for various countries or alternatively can be derived from remote sensing as well (Gómez Giménez et al., 2017; Tetteh et al., 2020). Another issue of using parcels is that grassland management activities are not always homogenous in space and time throughout the parcel which might lead to false detections.

Another limitation that has to be addressed is the restriction of our study to meadows. Future work should include the separation of meadows, pastures, and mixed grassland types (mown pastures) into the analysis framework to be able to account for other land use activities (e.g. grazing) in the analysis of the temporal signature of grasslands. This is important not least because meadows only made up about 40% of permanent grassland in Germany in 2020 (Federal Statistical Office, 2021). De Vroey et al. (2021) stated grazing as one of the major sources of false positive detections. Since grazing reduces the grass height, changes in NDVI or SAR features are expected as well. However, the amount of livestock and lengths of grazing periods are likely to show variations in the observed signals. Another source of false positive detections can be introduced by fertilization practices on the parcels (e.g., manure) as reported by Kolecka et al. (2018). Here, the signal of the time series could be influenced by a layer of manure or changes of the grass after vehicles entered the meadow. De Vroey et al. (2021) also reported that the impact of varying regrowth rates of the grass between the seasons can further impact the detection performance. Yet, no systematical behavior indicating this was found in our study.

From the technical perspective, the linear interpolation of the optical data to the S1-interval was a convenient step. However, this modified the actual recorded values and time points and might lead to the omission of mowing events. Our systematic evaluation concept allowed new insights on the capabilities, synergy, and limitations of multi-system and multi-sensor Earth Observation time series and provides a benchmark for future monitoring activities for grassland management intensity. Yet, due to the vast amount of data and possible variable combinations, not all single features could be treated separately. E.g., a separation of the individual polarizations of the SAR time series or the texture metrics within the evaluation could give deeper insight. Here, De Vroey et al. (2021) found coherence in VV to show better results in comparison to the VH polarized time series.

All our study sites are located in the orbit overlap areas of S2. This fact, in addition to the combination of both sensors (S2 & L8), will have contributed to a large extent to the high density of the NDVI time series. This has to be considered for wide-area monitoring tasks, that are likely to include non-overlap areas, since the observation density could be considerably reduced. Therefore, the evaluation of the model on reference data from non-overlap areas is an important task to study the impact on mowing detection.

For future studies, it is worth mentioning that alternative optical (e.g., PlanetScope) and SAR Earth Observation systems (e.g., ICEYE) as well as future missions (e.g., Landsat Next, ESA BIOMASS, and CHIME) could complement the data that was used. Especially the improved revisit-times that reach up to (potentially) daily coverage could substantially reduce the probability of missing rapidly changing phenomena like mowing events (Roy et al., 2021). But, at present and for the time being the three sensors that provided the basis for this work represent the state-of-art in freely available Earth Observation data that are able to capture land use processes at high temporal and medium spatial resolution.

6. Conclusion

We demonstrated the impact of choice and combination of optical and SAR input features for the parametrization of management practices in permanent grasslands in Germany. In general, our study verified the capability of dense earth observation time series data and machine learning algorithms to detect mowing events. The expected synergy of optical and SAR satellite data was confirmed by the overall improvement in classification performance when multiple input features from different sensors were combined. Here, the input-set of NDVI (from S2 / L8) together with the 6-day interferometric coherence and backscatter cross-ratio (from S1) yielded the overall best performance. Single mowing events could be detected with a temporal mean absolute error of 3.79 days, which can be considered low in the context of the temporal resolution of the underlying satellite time series. Limitations of the approach were recognized on intensively managed meadows where the mowing frequency tended to be underestimated.

This study was the first to systematically investigate the impact of the combination of the most common input features from optical and SAR satellite systems for the detection of mowing events. Even if the evaluation of a particular classification algorithm was not the main aspect, our work offered insights into the classification power of satellite time series with 1D CNNs and provides a basis for further evaluation regarding both the optimization of classification algorithms for dense satellite time series analysis and the complementarity of auxiliary input data for further improvement of the prediction accuracy.

A key strength of the study was the availability of a consistent and extensive reference data set that represents different biogeographical regions and management intensity levels. This allowed for a detailed evaluation of the model performance and an outlook towards the transferability of the proposed approach to larger regions and other agricultural landscapes.

Overall, the outcome of this study can support the development of satellite-based monitoring strategies for grassland management intensity on state or national levels. Possible applications are targeted towards the evaluation of compliance schemes in agricultural policy or the assessment of grassland use intensity as an indicator of environmental quality and biodiversity in agricultural landscapes. In terms of transferability, the general outcome of our work underlines the additional value of combining multiple optical and SAR time series features for Earth Observation applications.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank the anonymous reviewers and the editors for critically reading the manuscript and improving its precision, clarity, and relevance with their helpful suggestions.

We thank the managers of the three Exploratories, Kirsten Reichel-Jung, Iris Steitz, Sandra Weithmann, Florian Staub, Juliane Vogt, Anna K. Franke, Miriam Teuscher and all former managers for their work in maintaining the plot and project infrastructure; Christiane Fischer and Victoria Grießmeier for giving support through the central office, Andreas Ostrowski 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. 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. The work has been (partly)

funded by the DFG Priority Program 1374 “Biodiversity- Exploratories” (DFG-Refno.). Field work permits were issued by the responsible state environmental offices of Baden-Württemberg, Thüringen, and Brandenburg.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.rse.2021.112751.

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., others, 2016. Tensorflow: A system for large-scale machine learning. In: 12th USENIX Symposium on Operating Systems Design and Implementation, pp. 265–283.

Allaire, J., Chollet, F., 2020. keras: R Interface to “Keras”. R Package Version 2.3.0.0. https://cran.r-project.org/package=keras.

Beck, H.E., Zimmermann, N.E., McVicar, T.R., Vergopolan, N., Berg, A., Wood, E.F., 2018. Present and future köppen-geiger climate classification maps at 1-km resolution. Sci. Data 5, 1–12. https://doi.org/10.1038/sdata.2018.214.

Benton, T.G., Vickery, J.A., Wilson, J.D., 2003. Farmland biodiversity: is habitat heterogeneity the key? Trends Ecol. Evol. https://doi.org/10.1016/S0169-5347(03)00011-9.

Benz, U., Banovsky, I., Cesarz, A., Schmidt, M., 2020. CODE-DE Portal Handbook, Version 2.0, DLR.

Billeter, R., Liira, J., Bailey, D., Bugter, R., Arens, P., Augenstein, I., Aviron, S., Baudry, J., Bukacek, R., Burel, F., Cerny, M., De Blust, G., De Cock, R., Diekötter, T., Dietz, H., Dirksen, J., Dormann, C., Durka, W., Frenzel, M., Hamersky, R., Hendrickx, F., Herzog, F., Klotz, S., Koolstra, B., Lausch, A., Le Coeur, D., Maelfait, J. P., Opdam, P., Roubalova, M., Schermann, A., Schermann, N., Schmidt, T., Schweiger, O., Smulders, M.J.M., Speelmans, M., Simova, P., Verboom, J., Van Wingerden, W.K.R.E., Zobel, M., Edwards, P.J., 2008. Indicators for biodiversity in agricultural landscapes: a pan-European study. J. Appl. Ecol. 45, 141–150. https://doi.org/10.1111/j.1365-2664.2007.01393.x.

Bishop, C.M., 1995. Neural Networks for Pattern Recognition. Oxford University Press, Inc., USA.

Buda, M., Maki, A., Mazurowski, M.A., 2017. A systematic study of the class imbalance problem in convolutional neural networks. Neural Netw. 106, 249–259. https://doi.org/10.1016/j.neunet.2018.07.011.

Chen, Y., Keogh, E., Hu, B., Begum, N., Bagnall, A., Mueen, A., Batista, G., 2015. The UCR Time Series Classification Archive.

Chiboub, O., Kallel, A., Frison, P.-L., Lopes, M., 2019. Monitoring of grasslands management practices using Interferometric products Sentinel-1. In: El-Askary, H. M., Lee, S., Heggy, E., Pradhan, B. (Eds.), Advances in Remote Sensing and Geo Informatics Applications. Springer International Publishing, Cham, pp. 239–242. https://doi.org/10.1007/978-3-030-01440-7_56.

Chollet, F., others, 2015. Keras. https://keras.io.

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

Courault, D., Hadria, R., Ruget, F., Olioso, A., Duchemin, B., Hagolle, O., Dedieu, G., 2010. Combined use of FORMOSAT-2 images with a crop model for biomass and water monitoring of permanent grassland in Mediterranean region. Hydrol. Earth Syst. Sci. 14, 1731–1744. https://doi.org/10.5194/hess-14-1731-2010.

de Vendictis, L., Sciaretta, C., Tutunaru, F., Zavagli, M., Bontemps, S., 2019. Sen4CAP - Sentinels Agricultural Policy for Common Design Justification File Benchmarking for L4B Grassland Mowing Detection Product.

De Vroey, M., Radoux, J., Defourny, P., 2021. Grassland mowing detection using Sentinel-1 time series: potential and limitations. Remote Sens. 13, 1–19. https://doi.org/10.3390/rs13030348.

Di Mauro, N., Vergari, A., Basile, T.M.A., Ventola, F.G., Esposito, F., 2017. End-to-end learning of deep spatio-temporal representations for satellite image time series classification. CEUR Workshop Proc. 1972.

Dietterich, T.G., 2002. Machine learning for sequential data: a review. Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) 2396, 15–30. https://doi.org/10.1007/3-540-70659-3_2.

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. Environ. Res. Lett. 13 https://doi.org/10.1088/1748-9326/aacc7a.

Farr, T.G., Rosen, P.A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D., Alsdorf, D.E., 2007. The shuttle radar topography mission. Rev. Geophys. 45 https://doi.org/10.1029/2005RG000183.

Federal Statistical Office, 2021. Permanent Grassland by Type of Use over Time [WWW Document]. URL. https://www.destatis.de/EN/Themes/Economic-Sectors-Enterprises/Agriculture-Forestry-Fisheries/Field-Crops-Grassland/Tables/grassland-comparison.html.

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

Foley, J.A., DeFries, R., Asner, G.P., Barford, C., Bonan, G., Carpenter, S.R., Chapin, F.S., Coe, M.T., Daily, G.C., Gibbs, H.K., Helkowski, J.H., Holloway, T., Howard, E.A., Kucharik, C.J., Monfreda, C., Patz, J.A., Prentice, I.C., Ramankutty, N., Snyder, P.K., 2005. Global consequences of land use. Science 80 (309), 570–574. https://doi.org/10.1126/science.1111772.

Frantz, D., 2019. FORCE-Landsat + Sentinel-2 analysis ready data and beyond. Remote Sens. 11 https://doi.org/10.3390/rs11091124.

Frantz, D., Haß, E., Uhl, A., Stoffels, J., Hill, J., 2018. Improvement of the Fmask algorithm for Sentinel-2 images: separating clouds from bright surfaces based on parallax effects. Remote Sens. Environ. 215, 471–481. https://doi.org/10.1016/j.rse.2018.04.046.

Gao, F., Anderson, M.C., Hively, W.D., 2020. Detecting cover crop end-of-season using venμs and sentinel-2 satellite imagery. Remote Sens. 12, 1–22. https://doi.org/10.3390/rs12213524.

Garioud, A., Giordano, S., Valero, S., Mallet, C., 2019. Challenges in grassland mowing event detection with multimodal sentinel images. In: 2019 10th International Workshop on the Analysis of Multitemporal Remote Sensing Images, MultiTemp 2019. Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/Multi-Temp.2019.8866914.

Garioud, A., Valero, S., Giordano, S., Mallet, C., 2020. On the joint exploitation of optical and SAR satellite imagery for grassland monitoring. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. - ISPRS Arch. 43, 591–598. https://doi.org/10.5194/isprs-archives-XLIII-B3-2020-591-2020.

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 Sens. Environ. 198, 126–139. https://doi.org/10.1016/j.rse.2017.06.003.

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press.

Gossner, M.M., Lewinsohn, T.M., Kahl, T., Grassein, F., Boch, S., Prati, D., Birkhofer, K., Renner, S.C., Sikorski, J., Wubet, T., Arndt, H., Baumgartner, V., Blaser, S., Blüthgen, N., Börschig, C., Buscot, F., DIekötter, T., Jorge, L.R., Jung, K., Keyel, A.C., Klein, A.M., Klemmer, S., Krauss, J., Lange, M., Müller, J., Overmann, J., Pašali, E., Penone, C., Perovic, D.J., Purschke, O., Schall, P., Socher, S.A., Sonnemann, I., Tschapka, M., Tscharntke, T., Türke, M., Venter, P.C., Weiner, C.N., Werner, M., Wolters, V., Wurst, S., Westphal, C., Fischer, M., Weisser, W.W., Allan, E., 2016. Land-use intensification causes multitrophic homogenization of grassland communities. Nature 540, 266–269. https://doi.org/10.1038/nature20575.

Grant, K., Siegmund, R., Wagner, M., Hartmann, S., 2015. Satellite-based assessment of grassland yields. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. - ISPRS Arch. 40, 15–18. https://doi.org/10.5194/isprsarchives-XL-7-W3-15-2015.

Graves, A., 2012. Supervised Sequence Labelling with Recurrent Neural Networks, Studies in Computational Intelligence. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-24797-2.

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 Sens. Environ. 238, 111124. https://doi.org/10.1016/j.rse.2019.03.017.

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 Sens. 7, 6107–6132. https://doi.org/10.3390/rs70506107.

Hall-Beyer, M., 2017. Practical guidelines for choosing GLCM textures to use in landscape classification tasks over a range of moderate spatial scales. Int. J. Remote Sens. 38, 1312–1338. https://doi.org/10.1080/01431161.2016.1278314.

Hallmann, C.A., Sorg, M., Jongejans, E., Siepel, H., Hofland, N., Schwan, H., Stenmans, W., Müller, A., Sumser, H., Hörren, T., Goulson, D., De Kroon, H., 2017. More than 75 percent decline over 27 years in total flying insect biomass in protected areas. PLoS One 12. https://doi.org/10.1371/journal.pone.0185809.

Haralick, R.M., Dinstein, I., Shanmugam, K., 1973. Textural features for image classification. IEEE Trans. Syst. Man Cybern. SMC-3, 610–621. https://doi.org/10.1109/TSMC.1973.4309314.

Holtgrave, A.-K., Röder, N., Ackermann, A., Erasmi, S., Kleinschmit, B., 2020. Comparing Sentinel-1 and -2 data and indices for agricultural land use monitoring. Remote Sens. 12, 2919. https://doi.org/10.3390/rs12182919.

Hörtnagl, L., Barthel, M., Buchmann, N., Eugster, W., Butterbach-Bahl, K., Díaz-Pinés, E., Zeeman, M., Klumpp, K., Kiese, R., Bahn, M., Hammerle, A., Lu, H., Ladreiter-Knauss, T., Burri, S., Merbold, L., 2018. Greenhouse gas fluxes over managed grasslands in Central Europe. Glob. Chang. Biol. 24, 1843–1872. https://doi.org/10.1111/gcb.14079.

IPBES, 2019. In: Díaz, S., Settele, J., Brondízio, E.S., Ngo, H.T., Guèze, M., Aga, J. (Eds.), Summary for Policymakers of the Global Assessment Report on Biodiversity and Ecosystem Services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. https://doi.org/10.1111/padr.12283. Population and Development Review. Bonn, Germany.

Jones, S.K., Rees, R.M., Skiba, U.M., Ball, B.C., 2005. Greenhouse gas emissions from a managed grassland. In: Global and Planetary Change. Elsevier, pp. 201–211. https://doi.org/10.1016/j.gloplacha.2004.10.011.

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

Kavats, O., Khramov, D., Sergieieva, K., Vasyliev, V., 2019. Monitoring harvesting by time series of Sentinel-1 SAR data. Remote Sens. 11, 1–16. https://doi.org/10.3390/rs11212496.

Khatami, R., Mountrakis, G., Stehman, S.V., 2016. A meta-analysis of remote sensing research on supervised pixel-based land-cover image classification processes: general guidelines for practitioners and future research. Remote Sens. Environ. 177, 89–100. https://doi.org/10.1016/j.rse.2016.02.028.

Kingma, D.P., Ba, J.L., 2015. Adam: A method for stochastic optimization. In: 3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc, pp. 1–15.

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 Sens. 10 https://doi.org/10.3390/rs10081221.

Lobert, F., Holtgrave, A.-K., Schwieder, M., Pause, M., Gocht, A., Vogt, J., Erasmi, S., 2021. Detection of mowing events from combined Sentinel-1, Sentinel-2, and Landsat 8 time series with machine learning. In: Grassl Sci Europe, 26, pp. 123–125.

Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., Johnson, B.A., 2019. Deep learning in remote sensing applications: a meta-analysis and review. ISPRS J. Photogramm. Remote Sens. 152, 166–177. https://doi.org/10.1016/j.isprsjprs.2019.04.015.

Moreira, A., Prats-Iraola, P., Younis, M., Krieger, G., Hajnsek, I., Papathanassiou, K.P., 2013. A tutorial on synthetic aperture radar. IEEE Geosci. Remote Sens. Mag. 1, 6–43. https://doi.org/10.1109/MGRS.2013.2248301.

Nasirzadehdizaji, R., Cakir, Z., Balik Sanli, F., Abdikan, S., Pepe, A., Calò, F., 2021. Sentinel-1 interferometric coherence and backscattering analysis for crop monitoring. Comput. Electron. Agric. 185, 106118. https://doi.org/10.1016/j.compag.2021.106118.

Pelletier, C., Webb, G.I., Petitjean, F., 2019. Temporal convolutional neural network for the classification of satellite image time series. Remote Sens. 11, 1–22. https://doi.org/10.3390/rs11050523.

Poeplau, C., Jacobs, A., Don, A., Vos, C., Schneider, F., Wittnebel, M., Tiemeyer, B., Heidkamp, A., Prietz, R., Flessa, H., 2020. Stocks of organic carbon in German agricultural soils—key results of the first comprehensive inventory. J. Plant Nutr. Soil Sci. 183, 665–681. https://doi.org/10.1002/jpln.202000113.

R Core Team, 2021. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL. https://www.R-project.org/.

Reinermann, S., Asam, S., Kuenzer, C., 2020. Remote sensing of grassland production and management—a review. Remote Sens. 12, 1949. https://doi.org/10.3390/rs12121949.

Roy, D., Hung, H., Houborg, R., Martins, V., 2021. A global analysis of the temporal availability of PlanetScope high spatial resolution multi-spectral imagery. Remote. Sens. Environ. 264, 112586. https://doi.org/10.1016/j.rse.2021.112586.

Sakamoto, T., Yokozawa, M., Toritani, H., Shibayama, M., Ishitsuka, N., Ohno, H., 2005. A crop phenology detection method using time-series MODIS data. Remote Sens. Environ. 96, 366–374. https://doi.org/10.1016/j.rse.2005.03.008.

Sándor, R., Ehrhardt, F., Brilli, L., Carozzi, M., Recous, S., Smith, P., Snow, V., Soussana, J.F., Dorich, C.D., Fuchs, K., Fitton, N., Gongadze, K., Klumpp, K., Liebig, M., Martin, R., Merbold, L., Newton, P.C.D., Rees, R.M., Rolinski, S., Bellocchi, G., 2018. The use of biogeochemical models to evaluate mitigation of greenhouse gas emissions from managed grasslands. Sci. Total Environ. 642, 292–306. https://doi.org/10.1016/j.scitotenv.2018.06.020.

Savitzky, A., Golay, M.J.E., 1964. Smoothing and differentiation of data by simplified least squares procedures. Anal. Chem. 36, 1627–1639. https://doi.org/10.1021/ac60214a047.

Schlund, M., Erasmi, S., 2020. Sentinel-1 time series data for monitoring the phenology of winter wheat. Remote Sens. Environ. 246, 111814. https://doi.org/10.1016/j.rse.2020.111814.

Schlund, M., Lobert, F., Erasmi, S., 2021. Potential of Sentinel-1 time series data for the estimation of season length in winter wheat phenology. 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, pp. 5917–5920. https://doi.org/10.1109/igarss47720.2021.9554454.

Shang, J., Liu, J., Poncos, V., Geng, X., Qian, B., Chen, Q., Dong, T., Macdonald, D., Martin, T., Kovacs, J., Walters, D., 2020. Detection of crop seeding and harvest through analysis of time-series Sentinel-1 interferometric SAR data. Remote Sens. 12, 1–18. https://doi.org/10.3390/rs12101551.

Siegmund, R., Grant, K., Wagner, M., Hartmann, S., 2016. Satellite-based monitoring of grassland: assessment of harvest dates and frequency using SAR. Remote Sens. Agric. Ecosyst. Hydrol. XVIII 9998, 999803. https://doi.org/10.1117/12.2240947.

Small, D., 2011. Flattening gamma: radiometric terrain correction for SAR imagery. IEEE Trans. Geosci. Remote Sens. 49, 3081–3093. https://doi.org/10.1109/TGRS.2011.2120616.

Sokolova, M., Lapalme, G., 2009. A systematic analysis of performance measures for classification tasks. Inf. Process. Manag. 45, 427–437. https://doi.org/10.1016/j.ipm.2009.03.002.

Stendardi, L., Karlsen, S.R., Niedrist, G., Gerdol, R., Zebisch, M., Rossi, M., Notarnicola, C., 2019. Exploiting time series of Sentinel-1 and Sentinel-2 imagery to detect meadow phenology in mountain regions. Remote Sens. 11, 1–24. https://doi.org/10.3390/rs11050542.

Tamm, T., Zalite, K., Voormansik, K., Talgre, L., 2016. Relating Sentinel-1 interferometric coherence to mowing events on grasslands. Remote Sens. 8, 1–19. https://doi.org/10.3390/rs8100802.

Taravat, A., Wagner, M., Oppelt, N., 2019. Automatic grassland cutting status detection in the context of spatiotemporal Sentinel-1 imagery analysis and artificial neural networks. Remote Sens. 11, 711. https://doi.org/10.3390/rs11060711.

Tetteh, G.O., Gocht, A., Conrad, C., 2020. Optimal parameters for delineating agricultural parcels from satellite images based on supervised Bayesian optimization. Comput. Electron. Agric. 178, 105696. https://doi.org/10.1016/j.compag.2020.105696.

Torres, R., Snoeij, P., Geudtner, D., Bibby, D., Davidson, M., Attema, E., Potin, P., Rommen, B.Ö., Floury, N., Brown, M., Traver, I.N., Deghaye, P., Duesmann, B., Rosich, B., Miranda, N., Bruno, C., L'Abbate, M., Croci, R., Pietropaolo, A., Huchler, M., Rostan, F., 2012. GMES Sentinel-1 mission. Remote Sens. Environ. 120, 9–24. https://doi.org/10.1016/j.rse.2011.05.028.

Tscharntke, T., Klein, A.M., Kruess, A., Steffan-Dewenter, I., Thies, C., 2005. Landscape perspectives on agricultural intensification and biodiversity - ecosystem service management. Ecol. Lett. 8, 857–874. https://doi.org/10.1111/j.1461-0248.2005.00782.x.

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

Vogt, J., Klaus, V., Both, S., Fürstenau, C., Gockel, S., Gossner, M., Heinze, J., Hemp, A., Hölzel, N., Jung, K., Kleinebecker, T., Lauterbach, R., Lorenzen, K., Ostrowski, A., Otto, N., Prati, D., Renner, S., Schumacher, U., Seibold, S., Simons, N., Steitz, I., Teuscher, M., Thiele, J., Weithmann, S., Wells, K., Wiesner, K., Ayasse, M., Blüthgen, N., Fischer, M., Weisser, W., 2019. Eleven years' data of grassland management in Germany. Biodivers. Data J. 7, e36387 https://doi.org/10.3897/bdj.7.e36387.

Voormansik, K., Jagdhuber, T., Zalite, K., Noorma, M., Hajnsek, I., 2016. Observations of cutting practices in agricultural grasslands using polarimetric SAR. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 9, 1382–1396. https://doi.org/10.1109/JSTARS.2015.2503773.

Voormansik, K., Zalite, K., Sünter, I., Tamm, T., Koppel, K., Verro, T., Brauns, A., Jakovels, D., Praks, J., 2020. Separability of mowing and ploughing events on short temporal baseline Sentinel-1 coherence time series. Remote Sens. 12, 3784. https://doi.org/10.3390/rs12223784.

Vreugdenhil, M., Wagner, W., Bauer-Marschallinger, B., Pfeil, I., Teubner, I., Rüdiger, C., Strauss, P., 2018. Sensitivity of Sentinel-1 backscatter to vegetation dynamics: an Austrian case study. Remote Sens. 10, 1–19. https://doi.org/10.3390/rs10091396.

Vrieling, A., Meroni, M., Darvishzadeh, R., Skidmore, A.K., Wang, T., Zurita-Milla, R., Oosterbeek, K., O'Connor, B., Paganini, M., 2018. Vegetation phenology from Sentinel-2 and field cameras for a Dutch barrier island. Remote Sens. Environ. 215, 517–529. https://doi.org/10.1016/j.rse.2018.03.014.

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

Wang, Z., Yan, W., Oates, T., 2017. Time series classification from scratch with deep neural networks: A strong baseline. In: Proc. Int. Jt. Conf. Neural Networks 2017-May, pp. 1578–1585. https://doi.org/10.1109/IJCNN.2017.7966039.

Weiner, C.N., Werner, M., Linsenmair, K.E., Blüthgen, N., 2011. Land use intensity in grasslands: changes in biodiversity, species composition and specialisation in flower visitor networks. Basic Appl. Ecol. 12, 292–299. https://doi.org/10.1016/j.baae.2010.08.006.

Yu, Le, Liang, L., Wang, J., Zhao, Y., Cheng, Q., Hu, L., Liu, S., Yu, Liang, Wang, X., Zhu, P., Li, Xueyan, Xu, Y., Li, C., Fu, W., Li, Xuecao, Li, W., Liu, C., Cong, N., Zhang, H., Sun, F., Bi, X., Xin, Q., Li, D., Yan, D., Zhu, Z., Goodchild, M.F., Gong, P., 2014. Meta-discoveries from a synthesis of satellite-based land-cover mapping research. Int. J. Remote Sens. 35, 4573–4588. https://doi.org/10.1080/01431161.2014.930206.

Zalite, K., Voormansik, K., Praks, J., Antropov, O., Noorma, M., 2014. Towards detecting mowing of agricultural grasslands from multi-temporal COSMO-SkyMed data. In: International Geoscience and Remote Sensing Symposium (IGARSS). Institute of Electrical and Electronics Engineers Inc, pp. 5076–5079. https://doi.org/10.1109/IGARSS.2014.6947638.

Zalite, K., Antropov, O., Praks, J., Voormansik, K., Noorma, M., 2016. Monitoring of agricultural grasslands with time series of X-band repeat-pass interferometric SAR. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 9, 3687–3697. https://doi.org/10.1109/JSTARS.2015.2478120.

Zeng, L., Wardlow, B.D., Xiang, D., Hu, S., Li, D., 2020. A review of vegetation phenological metrics extraction using time-series, multispectral satellite data. Remote Sens. Environ. 237, 111511. https://doi.org/10.1016/j.rse.2019.111511.

Zhong, L., Hu, L., Zhou, H., 2019. Deep learning based multi-temporal crop classification. Remote Sens. Environ. 221, 430–443. https://doi.org/10.1016/j.rse.2018.11.032.

Zhu, Z., Woodcock, C.E., 2012. Object-based cloud and cloud shadow detection in Landsat imagery. Remote Sens. Environ. 118, 83–94. https://doi.org/10.1016/j.rse.2011.10.028.

Zhu, Z., Wang, S., Woodcock, C.E., 2015. Improvement and expansion of the Fmask algorithm: cloud, cloud shadow, and snow detection for Landsats 4-7, 8, and sentinel 2 images. Remote Sens. Environ. 159, 269–277. https://doi.org/10.1016/j.rse.2014.12.014.