Remote Sensing of Environment 342 (2026) 115466

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Mowing detection intercomparison exercise (MODCiX) – Evaluation of grassland mowing detection algorithms across Europe

Marcel Schwieder [a,b,*], Felix Lobert [a,b], Dominique Weber [c], Sophie Reinermann [d,e], Sarah Asam [e], Filippo Sarvia [f], Samuele De Petris [f], Enrico Borgogno-Mondino [f], Arnab Muhuri [g], Natascha Oppelt [g], Clement Atzberger [h], Iason Tsardanidis [i], Charalampos Kontoes [i], François Godechal [j], Cozmin Lucau-Danila [j], Viviane Planchon [j], Anatol Garioud [k], Célestin Huet [k], Silvia Valero [u], Clément Mallet [v], Julien Morel [l,m], Mattia Rossi [l], Francesco Vuolo [n], Aleksandar Dujakovic [n], Andreas Schaumberger [o], Andreas Klingler [o], Ann-Kathrin Holtgrave [p,q], Zander Venter [s], Ruth Sonnenschein [t], Mathilde De Vroey [w], Julien Radoux [w], Oliver Buck [x], Anna Katharina Franke [y], Uta Schumacher [z], Andreas Ostrowski [aa], Patrick Hostert [b,r], Stefan Erasmi [a]

[a] Thünen Earth Observation (ThEO), Thünen Institute of Farm Economics, Bundesallee 63, 38116 Braunschweig, Germany
[b] Geography Department, Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany
[c] Swiss Federal Research Institute WSL, Zürcherstrasse 111, 8903 Birmensdorf, Switzerland
[d] University of Würzburg, Institute of Geography and Geology, Department of Remote Sensing, Am Hubland, 97074 Würzburg, Germany
[e] German Remote Sensing Data Center (DFD), Earth Observation Center (EOC), German Aerospace Center (DLR), Weßling, Germany
[f] Department of Agricultural, Forest and Food Sciences, University of Turin, L.go Braccini 2, 10095 Grugliasco, Italy
[g] Earth Observation and Modelling (EoM), Geographisches Institut, Christian-Albrechts-Universität zu Kiel, 24118 Kiel, Germany
[h] Climate Labs, New York, United States
[i] Operational Unit BEYOND Centre for EO Research and Satellite Remote Sensing, Institute for Astronomy, Astrophysics Space Applications and Remote Sensing (IAASARS), National Observatory of Athens (NOA), Athens, Greece
[j] Walloon Agricultural Research Centre, Agriculture Territory and Technologies Integration Unit, 5030 Gembloux, Belgium
[k] Institut national de l'information géographique et forestière (IGN), France
[l] European Commission, Joint Research Centre (JRC), Ispra, Italy
[m] Department of Crop Production Ecology, Swedish University of Agricultural Sciences, Umeå 90183, Sweden
[n] BOKU University, Institute of Geomatics, Department of Ecosystem Management, Climate and Biodiversity, Peter-Jordan-Straße 82, Vienna 1190, Austria
[o] Agricultural Research and Education Center Raumberg-Gumpenstein, 8952 Irdning-Donnersbachtal, Austria
[p] LUP – Luftbild Umwelt Planung GmbH, Potsdam, Germany
[q] Geoinformation in Environmental Planning, Technische Universität Berlin, Germany
[r] Integrative Research Institute of Transformations of Human-Environment Systems (IRI THESys), Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany
[s] Norwegian Institute for Nature Research – NINA, Sognsveien 68, 0855 Oslo, Norway
[t] Eurac Research, Institute for Earth Observation, Bolzano, Italy
[u] CESBIO, CNES/CNRS/INRAE/IRD/UT3 – Université de Toulouse, France
[v] Univ Gustave Eiffel, ENSG, IGN, LASTIG, France
[w] Earth and Life Institute, Université catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium
[x] EFTAS Fernerkundung Technologietransfer GmbH, Oststraße 2, 48145, Münster, Germany
[y] TUM School of Life Sciences, Technische Universität München, 85354 Freising, Germany
[z] Georg-August-Universität Göttingen, Germany
[aa] Friedrich Schiller University Jena, 07743 Jena, Germany

* Corresponding author at: Thünen Earth Observation (ThEO), Thünen Institute of Farm Economics, Bundesallee 63, 38116 Braunschweig, Germany.
E-mail address: marcel.schwieder@thuenen.de (M. Schwieder).

https://doi.org/10.1016/j.rse.2026.115466
Received 12 September 2025; Received in revised form 28 April 2026; Accepted 3 May 2026
Available online 13 May 2026

A R T I C L E I N F O

Edited by: Marie Weiss

Keywords:
Agriculture
Land use intensity
Grassland
Evaluation
Landsat
Sentinel-2
Sentinel-1
Europe
Common agricultural policy

A B S T R A C T

Grasslands deliver a variety of ecosystem services, such as the provision of biomass, carbon sequestration or water retention and with that play a key role for climate change mitigation and preservation of biological diversity. The intensity of grassland management directly impacts these ecosystem services and functions. However, spatial information on grassland use intensity is scarce. Remote sensing time series from optical and/or SAR sensors help to overcome this data scarcity, as they enable to derive dates and frequency of mowing events as a proxy of grassland use intensity. A growing number of published algorithms relate abrupt changes in remote sensing time series to grassland management activities either by defining threshold-based rules or by making use of machine/deep learning techniques. So far, the different algorithms have not been compared and, due to a lack of suitable reference data, have usually not been tested for spatial and temporal transferability. We present the results of a comparison exercise based on an unprecedented set of independent reference data, containing information on more than 5000 grassland mowing dates that were compiled from eight European countries over a five-year period. We analyzed the performance of ten mowing detection algorithms across different geographic regions, years, mowing intensity, levels of reference data quality, and method and satellite sensor domains. The overall results show that when using all available reference data, F1 scores ranged from 0.55 to 0.74, and from 0.56 to 0.71 when only the highest quality reference data were considered. This decision, however, reduced the number of reference events to around 1500 with a regional bias to Austria, Germany, and Switzerland. We found that algorithm performance varies across space and time and that overall, the highest accuracies were achieved by machine learning based algorithms, although not substantially outperforming rule-based algorithms. The results did not confirm a consistently positive influence of the combined use of optical and SAR data in the prediction of mowing events, but we observed variations in algorithm performance, towards the lower and higher ends of grassland use intensity. Despite testing a variety of algorithms from different method and sensor domains, we observed a general upper limit of model performance and could not identify one single algorithm that performed best in all cases. The results of this comparison exercise can guide practitioners to choose approaches and input data that are most suitable for their specific use-case. The comparison exercise also highlights the importance of consistent, representative, and reliable reference data. We therefore recommend maintaining and extending this baseline dataset for the evaluation of upcoming algorithms, Earth Observation missions, and derived products for comprehensive monitoring of grassland use intensity.

1. Introduction

Covering an area of 30.1 million km^2, grasslands account for approximately 22.8% of the global land surface and are therefore one of the world's most important ecosystems (MacDougall et al., 2026). As part of a multifunctional agriculture, grasslands provide important ecosystem services such as carbon storage, erosion control, and water flow regulation (Bengtsson et al., 2019; Schils et al., 2022). Moreover, grasslands provide habitats for a wide range of animal and plant species, stressing their role in the context of biodiversity (Gossner et al., 2016; Habel et al., 2013; Petermann and Buzhdygan, 2021). One of the most prominent grassland ecosystem services is the provision of biomass, though, which is used as animal fodder.

Land use change, e.g., the conversion of grassland to cropland and the intensification of grassland use threatens the above-mentioned functions and the resilience of grasslands (Chang et al., 2021; Habel et al., 2013; Knozowski et al., 2023; Neyret et al., 2021). The high multifunctionality of grasslands in agricultural landscapes brings them into the focus of agricultural policies and nature conservation strategies. In Europe, for example, grassland, which is mainly used for fodder production or grazing, covers around 30% of the agricultural area, which makes them an essential component of the agricultural landscape (Eurostat, 2020) and an integral part of the Common Agricultural Policy (CAP). In general, the CAP requires farmers receiving subsidies to comply with EU standards for good agricultural and environmental conditions (GAEC), while additional voluntary eco-schemes promote sustainable grassland management by, for example, providing incentives to reduce mowing frequency, delay the first cut of the season, and promote extensive grazing. Furthermore, acknowledges the EU Biodiversity Strategy for 2030 the importance of grasslands for biodiversity conservation in agricultural landscapes (European Commission, 2020), while the EU Protein Strategy emphasizes the role of grasslands in providing protein for animal feed and calls for a diversified and sustainable approach to protein production.

Despite the importance of grassland use intensity indicators for monitoring and impact assessment of biodiversity, climate protection, or agri-environmental measures within the CAP, such indicators are not widely available as they are usually not reported in agricultural censuses. Satellite remote sensing-based algorithms have been shown to be a valuable way to fill this data gap, which is underlined by a steep increase in remote sensing-based developments in the past few years that focus on deriving spatially explicit information on grassland use intensity (Reinermann et al., 2020). Here, the basic rationale is that the extraction of grassland biomass, mainly by mowing, can be detected in time series of remote sensing data, as it causes a substantial alteration of the received vegetation signal. These algorithms are based either on rule-based concepts in which thresholds are utilized to identify abrupt changes in time series of remote sensing-based vegetation indices that are related to grassland management activities (e.g., Griffiths et al., 2020; Reinermann et al., 2023), originate from the field of supervised machine or deep learning (e.g., Holtgrave et al., 2023; Taravat et al., 2019) or are a combination of both (Watzig et al., 2023). Many of the algorithms use time series of either optical (e.g., Kolecka et al., 2018; Stumpf et al., 2020) or Synthetic Aperture Radar (SAR) data (e.g., De Vroey et al., 2021; Voormansik et al., 2016), while some rely on data from both data domains (e.g., Komisarenko et al., 2022; Lobert et al., 2021; Standardi et al., 2019). In terms of spatial extent, most of the published studies have a rather local focus on a specific region for which different types of field data was available for training and validation, but there is a recent trend in upscaling the algorithms in space and time by producing multi-annual maps on a national level (e.g., Griffiths et al., 2020; Reinermann et al., 2023; Schwieder et al., 2022; Stumpf et al., 2020; Weber et al., 2024). Each algorithm describes an innovation that aims to increase predictive accuracy and robustness compared to previous algorithms. However, reliable statements about the advantages and disadvantages can only be made when algorithms are compared across different environmental conditions using a common set of standardized input features and coherent reference data.

To enable insightful comparative studies, methods, data, and case studies on mowing detection from satellite data time series developed in recent years were compiled in this study labelled the “Mowing Detection intercomparison Exercise (MODCiX)”. MODCiX is as a joint effort of numerous working groups researching this topic in Europe, aiming to improve area-wide monitoring of grassland use intensity. In total, this unique reference dataset compiled around 5600 mowing events between 2017 and 2021 on more than 3000 parcels, representative of various grassland areas and types in Europe with varying use intensities as well as different climatic and environmental conditions. Based on these reference data and a common set of openly available satellite data (from Sentinel-1, Sentinel-2 and the Landsat series), we evaluated ten mowing detection algorithms to estimate grassland use intensity from time series of remote sensing data. This setting allowed us to identify the potential and limitations of the different algorithms based on a common set of independent reference data under a standardized validation framework. More precisely, the overall objective of this study is to identify the most robust algorithms and the related data and methods domain, which we assessed by answering the following research questions:

1 a) How well do the different mowing detection algorithms perform when evaluated based on a common set of input and reference data?

b) What differences can be observed between algorithms from different method and input data domains?

2) How well do they generalize when applied across diverse regions and multiple years?

3) How do environmental conditions and management intensities influence their performance?

2. Data and methods

MODCiX aimed to evaluate and compare the performance of different algorithms based on a common set of input and reference data. For this purpose, participating research groups shared their available reference data on grassland management in a disclosed data environment (Fig. 1, Table 1). There, the data were harmonized, anonymized, and split into training and validation data. In the next step, time series of optical (Landsat 7–9, Sentinel-2) and SAR (Sentinel-1) data were centrally preprocessed and shared with the groups to provide a common data basis.

2.1. Reference data

The reference data were acquired from different sources such as farmer reports, research facilities, field observations, satellite or webcam time series interpretation and consist of geolocations (points or polygons) for which dates of mowing events were reported (Table 1). All data were reprojected to a common coordinate reference system (EPSG:3035), column names and date formats were standardized and a unique ID as well as regional and temporal identifier attributes were assigned. Sometimes mowing events were, due to the timing of field work, reported as time periods in which a grassland was mown, rather

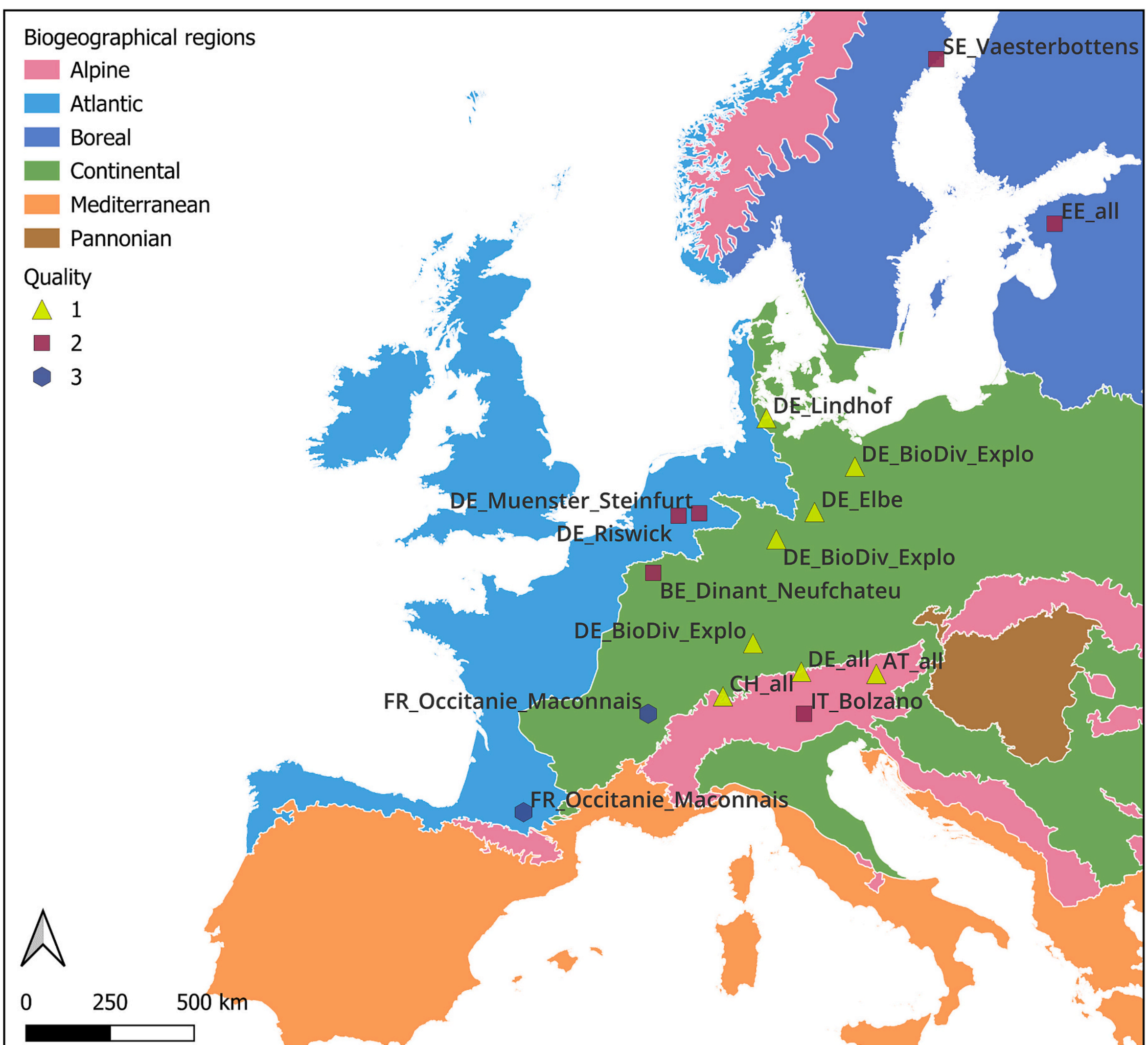

Fig. 1. Spatial distribution of study sites for which reference data were available across the biogeographical regions in Europe as defined by EEA (European Environment Agency, 2016). The different colour of the points indicate the assigned quality level of the reference data.

Table 1
Distribution of reference data by country (biogeographic regions in brackets), region, year and quality label and the amount of available parcels per year. In the column Reported cuts the range of reported mowing events for the grasslands in the specific region is shown, indicating an intensive or extensive grassland use. The data set names refer to the sites indicated in Fig. 1.

Country	Dataset	Region	Year	No. parcels	Reported cuts	Quality
Austria (Alpine)	AT_all	Several regions across Austria	2021	113	1–6	1 (Farmers' reports)
Belgium (Continental)	BE_Dinant_Neufchateau	Dinant / Neufchateu	2019	219	0–2	2 (Field surveys)
Estonia (Boreal)	EE_all	Several regions across Estonia	2020	34	1–4	2 (Farmers' reports)
			2021	34	1–4	
France (Atlantic/ Continental)	FR_Occitanie_Maconnais	Gers / Ariege / Haute-Garonne / Jura / Ain / Saône-et-Loire	2017	1033	1–4	3 (Time series interpretation)
Germany (Continental)	DE_BioDiv_Explo (Vogt et al., 2025)	Schorfheide-Chorin / Hainich-Dün / Schwäbische Alb	2017	98	1–5	1 (Farmers' reports)
			2018	90	1–4	
			2019	99	1–4	
			2020	91	1–3	
			2021	104	1–4	
Germany (Continental)	DE_all	Several regions across Southern Germany	2019	110	1–5	1 (Webcam interpretation)
Germany (Continental)	DE_Lindhof	Lindhof	2018	7	1–3	1 (Farmers' reports)
Germany (Atlantic)	DE_Muenster_Steinfurt	Münster /Steinfurt	2021	86	1–5	2 (Field surveys)
Germany (Atlantic)	DE_Riswick	Riswick	2017	6	3–4	2 (Farmers' reports)
			2018	7	1–4	
			2019	2	4	
Germany (Continental)	DE_Elbe	Elbe	2020	17	1–3	1 (Farmers' reports)
Italy (Alpine)	IT_Bolzano	Bolzano	2017	19	1–4	2 (Webcam interpretation)
			2018	164	1–4	
			2019	156	1–4	
			2020	199	1–5	
Sweden (Boreal)	SE_Vaesterbottens	Västerbotten	2018	11	1–2	2 (Field surveys)
			2019	8	1–2	
Switzerland (Alpine)	CH_all (Weber et al., 2023)	Several regions across Switzerland	2020	30	0–6	1 (Webcam interpretation)
			2021	32	0–5	

than an exact date. In these cases, the central date of the interval was chosen as mowing date. We excluded all mowing events from the reference data that were not within the defined mowing period (15th of March to 27th of October). Parcels for which mowing events were reported at too short intervals ($<$ 15 days) were also excluded from the analysis for the respective year. This quality control measure was carried out to exclude implausible reference data since grass regrowth in Europe usually does not justify two mowing events within two weeks. We assigned a square buffer of 30 m around the reported geolocation to all reference data that were not shared with an explicit parcel boundary (polygon).

The reference data were initially acquired in different ways, e.g., by interpreting webcam imagery and satellite data time series or collecting data from farmers and thus came with different confounding factors and uncertainties. We accounted for these uncertainties by assigning quality labels to the reference data, which ranged from 1 (very certain) to 3 (confounding factors) based on the initial way of data acquisition and reported issues (Table 1). We rated farmers' / grassland research facility reports as well as webcam image interpretation very high. Field data collections in which observed events were reported in intervals (due to the field campaign design) or had data gaps, as well as interpretations of confined webcam time series (e.g., by fog, clouds) were rated as quality level 2. Quality level 3 was assigned to satellite time series interpretations as these data are highly influenced by the data density of clear sky observation time series. Finally, each reference parcel was classified into a class of grassland use intensity (GLUI). Here we differentiated between extensive grassland use (1–2 mowing events; GLUI 1), semi-intensive grassland use (3–4 mowing events; GLUI 2) and intensive grassland use (more than 4 mowing events; GLUI 3).

To allow the groups to train or refine their algorithms on a common data base, the harmonized reference data were split into training and validation data using a random sampling strategy that was stratified across years, regions, quality labels, and GLUI. A share of 30% of the data were provided for training purposes including the geometry as well as the mowing date information. The leftover 70% of the data remained unseen by all participants. This unbalanced split into training and validation data was chosen to have a stronger focus on model evaluation than on refinement of the individual algorithms. For the final prediction, only the geometries of the validation data without any information on the mowing dates were shared.

2.2. Satellite data

2.2.1. Optical

Analysis ready time series of optical satellite data were created using the Framework for Operational Radiometric Correction for Environmental monitoring (FORCE; (Frantz, 2019), which enables data download, preprocessing, tiling and data harmonization of different sensor types. For the analysis, all available satellite data for the regions with reference data (Table 1) of the Landsat 7, 8, and 9 as well as Sentinel-2 A and B (S2) missions with a cloud cover of less than 70% were downloaded and organized in a 30 km $\times$ 30 km tiling scheme. Revisit frequencies of the sensors range between 5 days (S2) and 16 days (Landsat). All data were preprocessed to Level 2, which includes the correction of atmospheric and topographic effects as well as cloud masking using the FORCE Level 2 Processing System. Spectral harmonization between the sensors was not carried out. In a final step, analysis ready time series stacks were generated for each tile which included all available observations for all common bands of the different missions (i.e., blue, green, red, near infrared, shortwave infrared 1 and 2) as well as the Normalized Vegetation Index (NDVI, Tucker (1979)) and the Enhanced Vegetation Index (EVI, Huete et al. (2002)).

2.2.2. Synthetic aperture radar

SAR data from the Sentinel-1 (S1) A and B constellation provides imagery in VV and VH polarization. Between the launch of S1B in 2016 and its failure in late 2021, the S1 constellation collected data in a 6-day interval between acquisitions from the same orbit over Europe. We derived 6-day interferometric coherence from Single Look Complex (SLC) data and the gamma naught backscatter coefficient (γ^0) from Ground Range Detected (GRD) data. For the GRD scenes, calibration was applied to convert digital numbers to physically meaningful radar

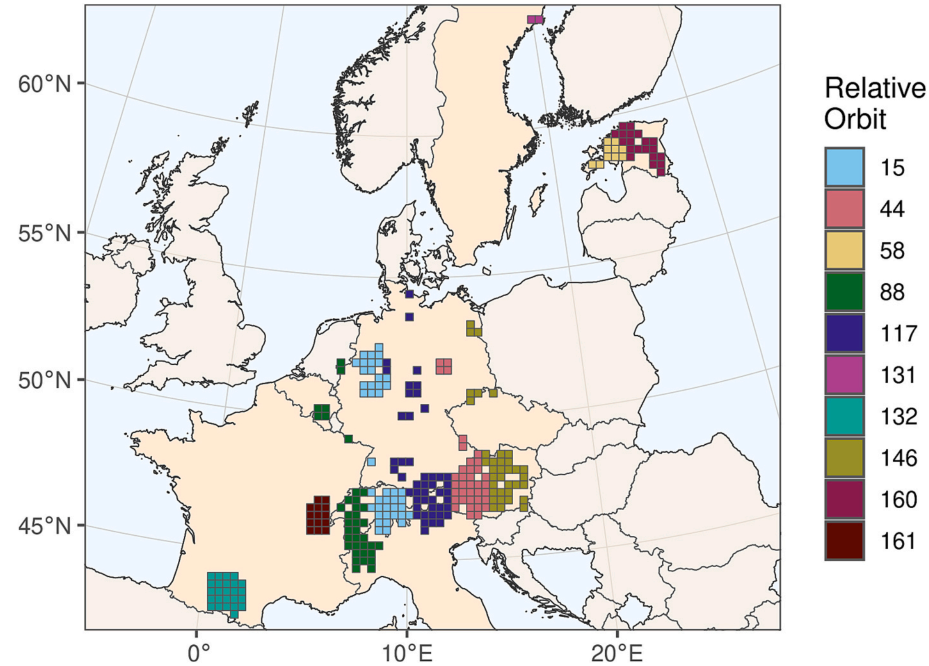

Fig. 2. Map of S1 relative orbits used for the different study areas.

backscatter, followed by radiometric flattening, to correct for incidence angle effects and terrain induced illumination, yielding γ^0 (Small, 2011). We selected a single relative orbit for each tile that covers all respective parcels (Fig. 2). This selection ensured a consistent acquisition geometry for the time series. We chose only ascending orbits (afternoon local acquisition time) to minimize the impact of morning dew on the SAR acquisitions. All available scenes from March to October were accessed through the Copernicus Data and Exploitation Platform - Deutschland (CODE-DE; Benz et al., 2020) and the Copernicus data and information access service CREODIAS. The data were processed using the Sentinel Application Platform (SNAP) and the built-in Graph Processing Tool (GPT). A detailed description can be found in Lobert et al. (2021).

To process γ^0, GRD data were first filtered for border and thermal noise, followed by calibration and radiometric flattening. A 3 × 3 boxcar filter was applied afterwards to reduce speckle before the images were terrain corrected and resampled to a 10 m spatial resolution that matched the optical data. Here, we used the Copernicus digital elevation model (DEM) GLO-30 (Copernicus, 2023). Finally, the γ^0 values were scaled to decibel range (dB).

Interferometric coherence was derived by first updating exact orbit positions and co-registering each 6-day image pair using back-geocoding. We then calculated the coherence in a two pixel (azimuth) by ten pixel (range) window to obtain an approximately square pixel size. Finally, the coherence images were terrain corrected and resampled to a spatial resolution of 10 m based on the GLO-30.

2.3. Algorithms

In this exercise, a total of ten algorithms were compared, which were developed and published at different institutions (Table 2). Additionally, two algorithms (NINA_ML and TICNN) were tested with different input data (optical only and optical + SAR), so that we in total compared twelve predictions. To ease interpretation, the algorithms were assigned to different groups based on the technical concept (ML: machine / deep learning, RBA: rule-based) and input data domain i.e., optical (OPT) and SAR. A brief description of each algorithm and the settings used for this comparison exercise can be found in the supplementary material.

2.4. Evaluation

All participating groups handed in their predictions of mowing dates for each reference geometry. The predictions were then related to the validation data based on the unique MODCiX IDs. All mowing dates that were reported in the reference data were compared to the predicted dates of mowing events, to identify true and false positives as well as false negative predictions. A predicted mowing event was considered correct if it was within a time frame of 12 days before or after the reported reference date, which matches the interval of two Sentinel-1

Table 2
Overview of compared algorithms in MODCiX in alphabetical order.

Institution	Acronym	Concept	Input	Spatial unit	Reference
Universität für Bodenkultur Wien	BOKU	RBA	Optical	Parcel	(Watzig et al., 2023)
Walloon Agricultural Research Centre	CRA	RBA	Optical	Parcel	Unpublished
German Aerospace Center	DLR	RBA	Optical	Pixel	(Reinermann et al., 2023)
EURAC Research	EURAC	RBA	Optical	Pixel	Unpublished
National Institute of Geographic and Forest Information	IGN	ML	Optical + SAR	Parcel	(Garioud et al., 2021)
Norwegian Institute for Nature Research	NINA_RBA	RBA	Optical	Pixel	Unpublished
	NINA_ML	ML	Optical + SAR	Pixel	Unpublished
	NINA_ML_OPT	ML	Optical	Pixel	Unpublished
National Observatory of Athens	NOA	ML	Optical + SAR	Pixel	(Tsardanidis et al., 2025)
Thünen Institute	TICNN	ML	Optical + SAR	Parcel	(Lobert et al., 2021)
	TICNN_OPT	ML	Optical	Parcel	(Lobert et al., 2021)
	TIRBA	RBA	Optical	Pixel	(Schwieder et al., 2022)

acquisitions and is comparable to the tolerance values used in Lobert et al. (2021) and Schwieder et al. (2022). Based on true and false predictions, the performance measures recall (producer's accuracy; Eq. 1), precision (user's accuracy; Eq. 2) and F1-Score (Eq. 3) were calculated per region, year, quality label and level of GLUI. Additionally, the mean absolute error (MAE) and the bias were derived for the number of predicted mowing events and the related dates. The latter statistic allows to assess the temporal offset between predictions and the corresponding reference events.

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \tag{1}$$

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \tag{2}$$

$$F1\ score = 2*\frac{Precision*Recall}{Precision + Recall} \tag{3}$$

3. Results

Since the results of the various combinations of input data, reference data quality methods used, years, regions, and usage intensities cannot all be shown in detail here, we also refer to the interactive web application (https://modcix.thuenen.de/), which allows for a differentiated visualization of all results. Here we focused on the general algorithm performances and selected exemplary subsets of the results.

3.1. Overall performance evaluation

The evaluation using reference data of all quality levels, regions, years and levels of GLUI revealed that all algorithms were within an F1-Score range from 0.55 to 0.74, while the difference between precision (overestimation of mowing events) and recall (omission of actual mowing events) varied among the algorithms (Fig. 3 A). When only reference data with quality label 1 were considered, the performance ranking slightly shifted and the F1-Scores ranged from 0.56 to 0.71 (Fig. 3 B). To ensure a meaningful analysis, only data of the best quality (level == 1) is shown further, unless otherwise specified. However, it is noteworthy that this subset of data substantially reduced the number of reference mowing events from 5324 to 1559 (71%) and biased the analysis to a regional focus on Germany, Switzerland, and Austria, with higher management intensities than average.

In this overall evaluation, the algorithms that combined optical and SAR data showed the highest F1-Scores (NOA, TICNN). However, some approaches that only used optical data work comparably well (TICN-N_OPT, NINA_RBA). Regarding the technical concept, the ML approaches achieved the overall highest F1-Scores (TICNN, NOA) but also here, most RBA procedures yielded only slightly worse results. Regarding the balance between precision and recall, we observed very narrow ranges for the DL algorithms NOA and TICNN but also for TIRBA. When evaluating the algorithms for each year individually, the rank patterns remained similar. However, in the year 2018, which was also the year with the least number of reported mowing events, all algorithms result in lower F1-Scores (Fig. 4).

3.2. Evaluation of regional and land use intensity subsets

On a regional scale, when considering highest quality reference data for all years and grassland use intensity levels, F1-Scores ranged from 0 to 0.89 and differences between the individual regions can be observed (Fig. 5). It became evident that the filtered (high quality) reported reference events are not evenly distributed across regions as they ranged from 9 (DE_Lindhof) up to 720 (DE_Elbe). In regions with many reference events, the differences in accuracy metrics resembled the patterns observed for all regions together (Fig. 3 A and B), with somewhat higher accuracies and less variations between precision and recall for ML algorithms that made use of deep learning such as IGN, NOA, and TICNN. It was noticeable that the inclusion of SAR data in the latter did not consistently lead to higher accuracies. However, in the Elbe region we observed a strong positive influence of the inclusion of SAR data. Overall, accuracy values showed higher variation in regions for which only a few reference data were available. In the test sites in Lindhof, Elbe, and Vaesterbottens we observed low precision metrics for most of the algorithms, indicating the overprediction of mowing events. It is

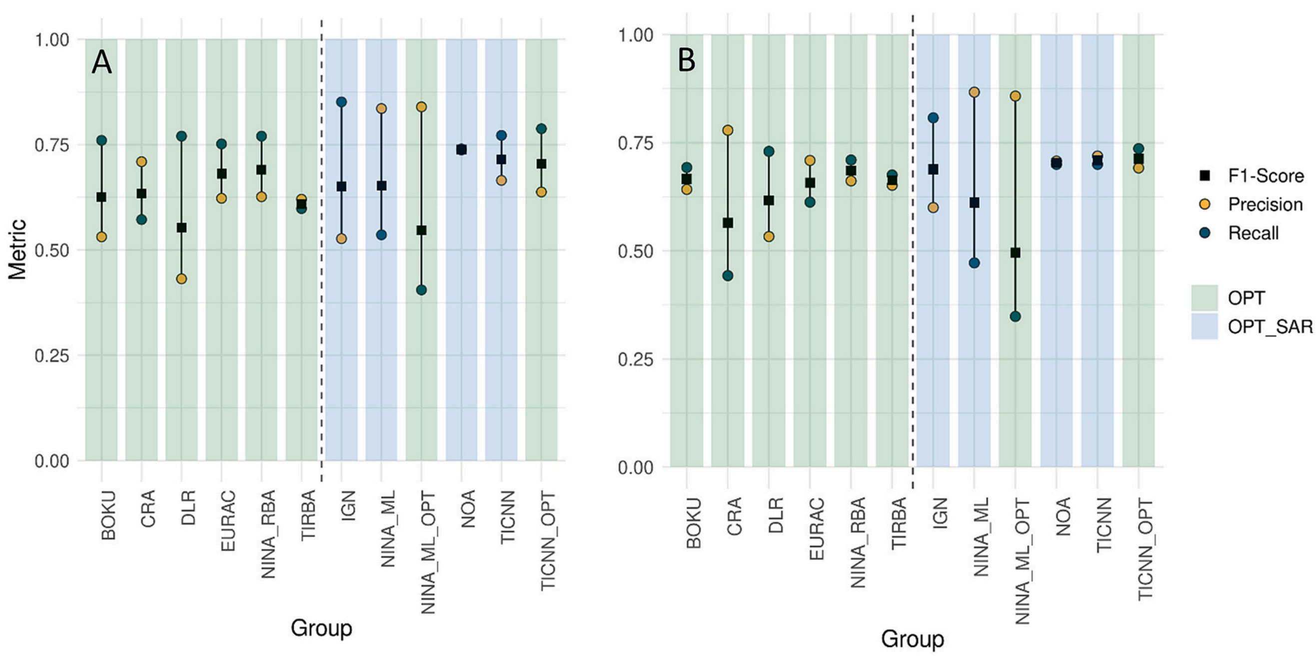

Fig. 3. Accuracy metrics considering all reference data (A) and only the best quality label (B). In each plot rule-based algorithms are on the left of the dashed line, the ones using machine learning on the right. The data domain is shaded in green for algorithms using optical (OPT) only and in blue when optical and SAR data (OPT_SAR) are used. Results can also be explored interactively on this webpage: https://modcix.thuenen.de/. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

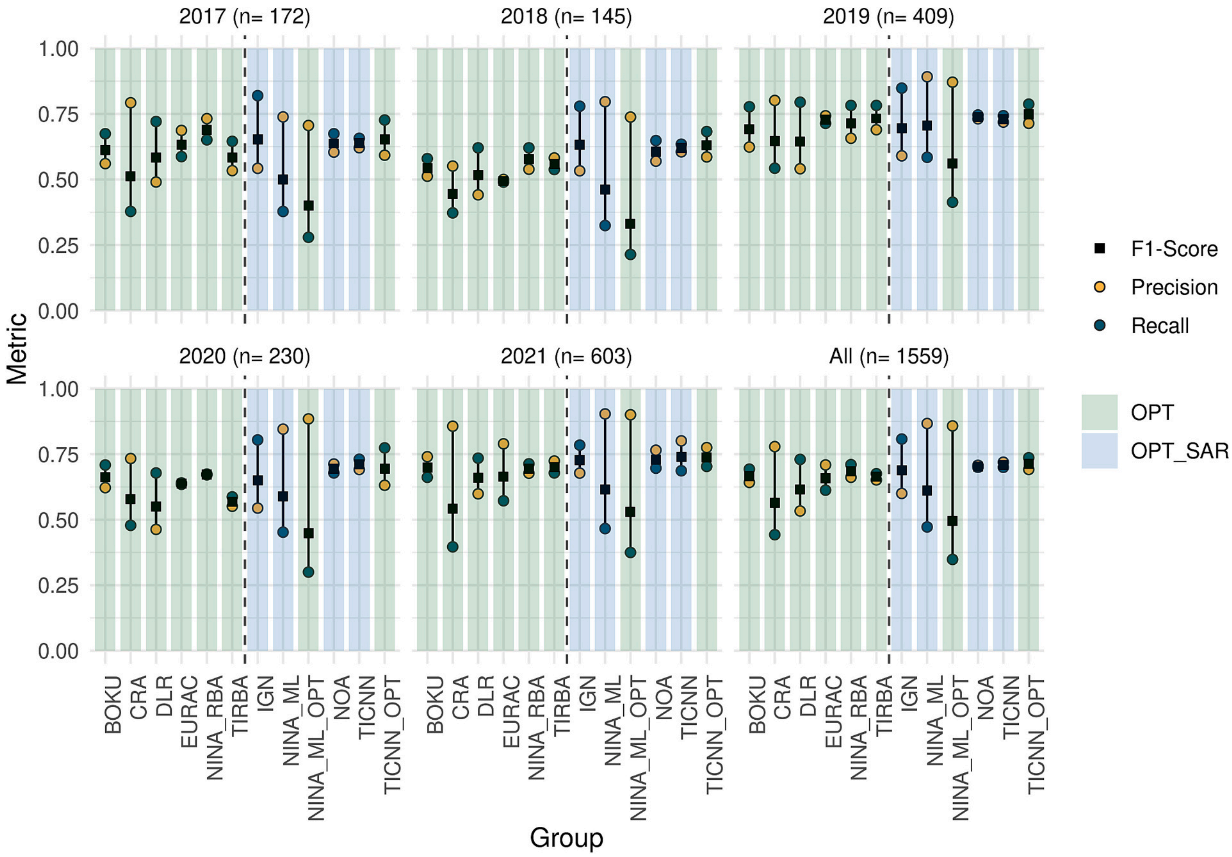

Fig. 4. Accuracy metrics per year considering reference data for all regions, levels of grassland use intensity, and the best quality label. In each plot rule-based algorithms are on the left of the dashed line, the ones using machine learning on the right. The data domain is shaded in green for algorithms using optical (OPT) only and in blue when optical and SAR data (OPT_SAR) are used. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

worth noting that in these regions, the accuracy of ML algorithms also showed high variation.

Considering all years and regions with extensive, moderately intensive, or intensive grassland use the F1-Scores ranged between 0.42 and 0.68 (GLUI 1), 0.54 and 0.81 (GLUI 2) and 0.47 and 0.81 (GLUI 3). The results showed that most algorithms performed more reliable in the moderately and intensively used grasslands and had a higher error of commission in the extensively managed regions compared to the moderately and intensively used regions in which precision was higher than recall for all algorithms (Fig. 6).

3.3. Timing and number of mowing events

The actual number of predicted mowing events deviated for all algorithms, except for CRA and NINA_ML_OPT, by an average of around one predicted event. There was no clear bias pattern observable, neither between methods nor input data used. The results showed that NINA_RBA, TIRBA, NOA, TICNN, and TICNN_OPT had a bias close to zero, indicating that these algorithms did on average not systematically under- or overpredict, considering reference of highest quality for all levels of grassland use intensity (Fig. 7 A). The prediction dates of the mowing events aligned generally well with the distribution of reference dates. We found that most of the RBA algorithms, apart from DLR and NINA_RBA, had a positive bias of about 2–3 days, while the ML algorithms, except for NINA_ML, had a smaller negative bias of about 1–2 days (Fig. 7 B). The mean absolute error ranged from 3.14 to 4.13 days.

4. Discussion

We compared the performance of ten mowing detection algorithms (twelve with input data variations) across five years and various grassland regions in Europe. All predictions were made based on a common set of satellite input data that were available for all participants, to minimize effects of data preprocessing on the generalizability assessment. The exercise showed that in general all algorithms were transferable but with difference in performance. The observed differences were related to two general boundary conditions: (i) land use and environmental factors, and (ii) data and algorithms. Both are to a certain degree interrelated, because, e.g., preprocessing or development and training of the algorithms have been parameterized to match the conditions of a given test site or region.

4.1. Differences in performance between method and data domains

Overall, the differences in accuracy metrics between the different tested methods and data domains did not show clear patterns that would allow a decision to be made in favor of one group (ML vs. RBA; OPT vs. OPT+SAR) or the other. However, in most comparisons (e.g., by year, intensity, region), the highest F1 scores came from the group of ML models, with the deep learning-based approaches NOA and TICNN standing out in particular. Still, the difference to other RBA or ML models was usually small. Moreover, the ML algorithms mostly showed narrower ranges between precision and recall compared to the RBA. Exceptions where can be inferred. In the regions Riswick and Lindhof, with only a few reported mowing events, ML algorithms clearly outperformed RBA at the cost of lower recall. In the 2017 RBA slightly

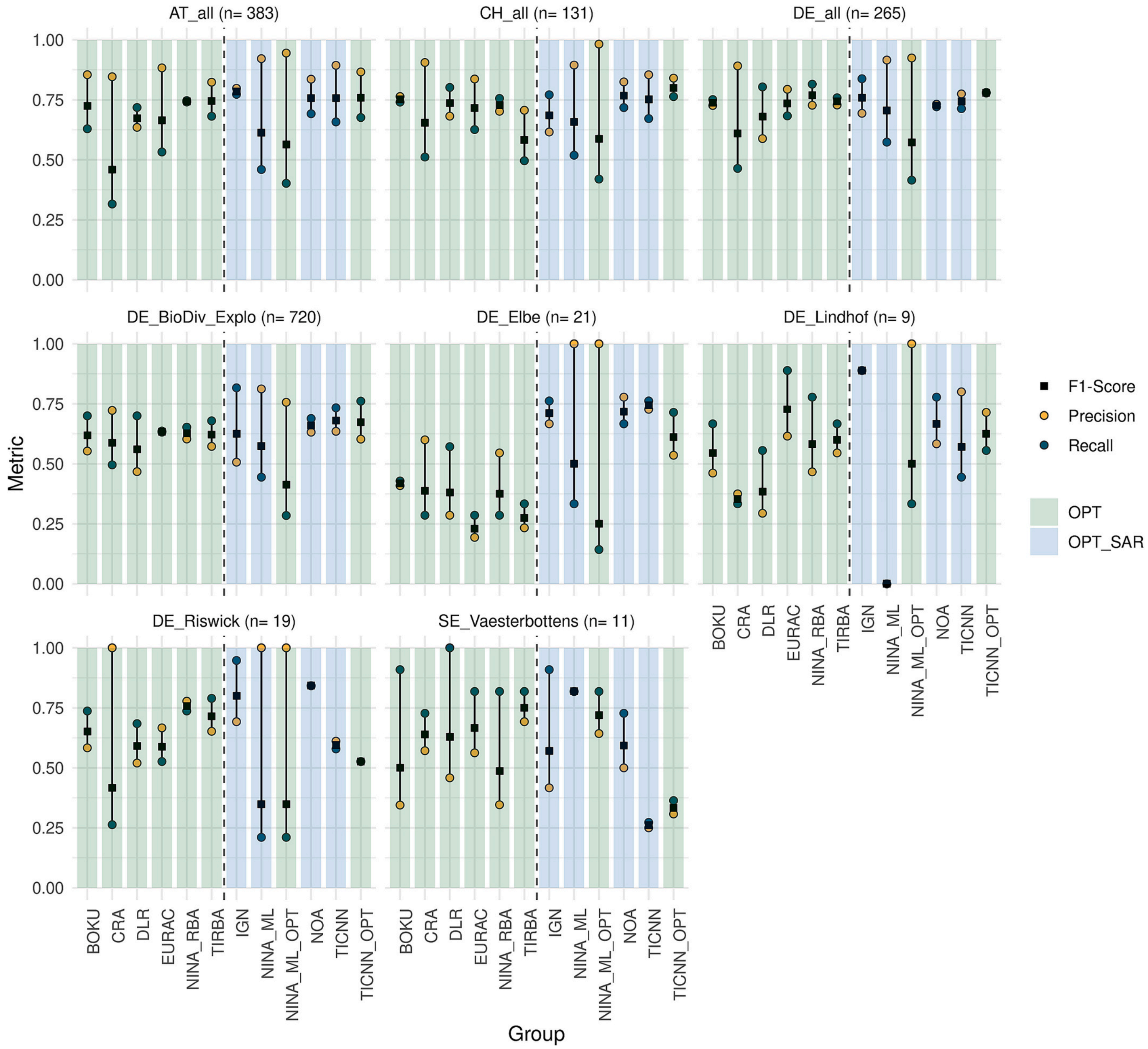

Fig. 5. Accuracy metrics per region considering reference data for all years, grassland use intensity levels, and the highest quality label. In each plot rule-based algorithms are on the left of the dashed line, the ones using machine learning on the right. The data domain is shaded in green for algorithms using optical (OPT) only and in blue when optical and SAR data (OPT_SAR) are used. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

outperformed ML methods.

Even though the comparison is not straightforward, we found that on average the ML performance metrics did not reach the level of the respective original publications. For example, Tsardanidis et al. (2025) and Lobert et al. (2021) reported F1-Scores of up to 0.84 but did overall not exceed 0.74 and 0.71 in this study (quality label 1, GLUI 1–3). The tested RBA approaches were on the other hand comparable to and even exceeded the results reported in the related publications by Reinermann et al. (2023), Watzig et al. (2023) and Schwieder et al. (2022), underlining their overall generalizability to other years and regions.

The joint use of SAR and optical data is straightforward within machine learning frameworks, whereas rule-based approaches combining both data domains are rare. As only rule-based algorithms that rely on optical alone were available for comparison, the effects of the method and data domain cannot be fully disentangled. We therefore included two optical only machine learning model setups, NINA_ML_OPT and TICNN_OPT, which differed from their original setup by excluding SAR inputs. The results revealed a model-dependent effect of removing SAR data. While NINA_ML_OPT showed a consistent and notable decrease in performance, TICNN_OPT using optical data only showed a slight increase in accuracy. This unambiguous behavior is consistent with previous studies reporting mixed findings regarding the benefit of including SAR data. Several studies highlight improved performance or transferability when combining SAR and optical observations (Lobert et al., 2021; Holtgrave et al., 2023; De Vroey et al., 2021, 2022); whereas Reinermann et al. (2022) observed limited or even negative effects due to increased false detections. A plausible explanation for the limited contribution of SAR data in this comparison exercise is that mowing events induce strong and rapid changes in vegetation greenness and canopy structure, which are directly captured by dense optical time

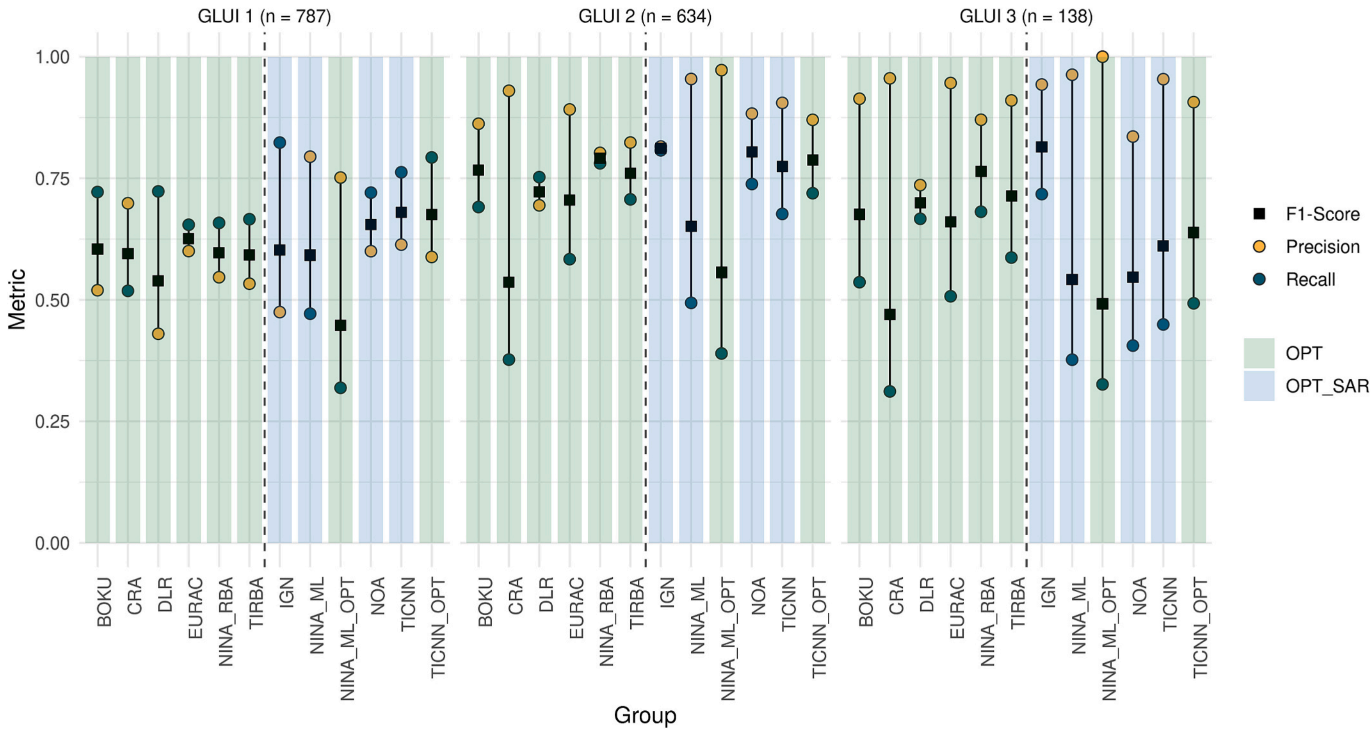

Fig. 6. Accuracy metrics per grassland use intensity (GLUI) class considering reference data for all years, regions, and the highest quality label. In each plot rule-based algorithms are on the left of the dashed line, the ones using machine learning on the right. The data domain is shaded in green for algorithms using optical (OPT) only and in blue when optical and SAR data (OPT_SAR) are used. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

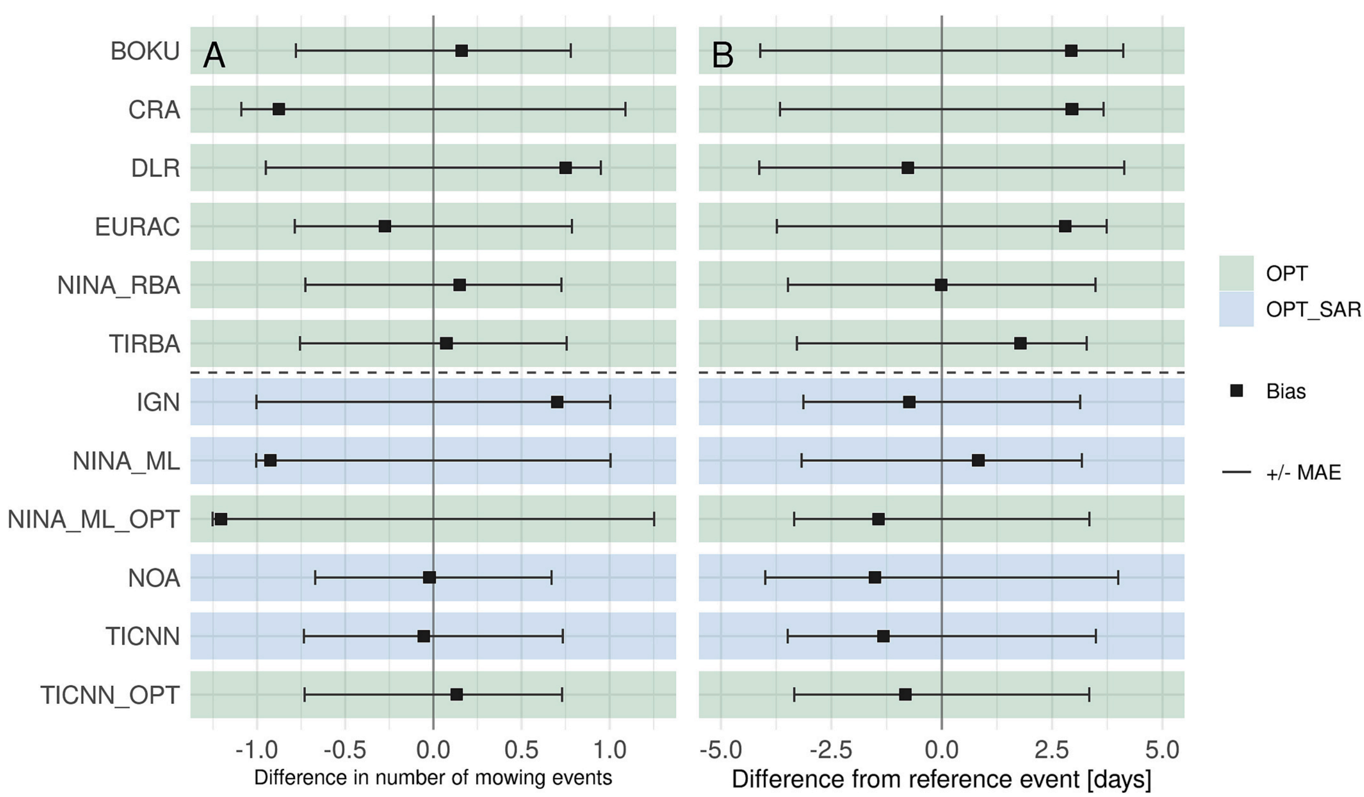

Fig. 7. Difference in predicted number of mowing events (A) and positive / negative MAE values (related to zero) and bias for the difference in days from the true positive predictions (B). Reference data for all years, regions, and the highest quality label are considered. In each plot rule-based algorithms are on top of the dashed line, the ones using machine learning at the bottom. The data domain is shaded in green for algorithms using optical (OPT) only and in blue when optical and SAR data (OPT_SAR) are used. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

series, while SAR-base signals respond more indirectly and are additionally influenced by soil moisture, surface roughness, and observation geometry including local terrain slope and aspect. Although we applied radiometric terrain correction as proposed by Small (2011), uncertainties arising from slope and aspect differences cannot be ruled out. Moreover, differences in spatial resolution, acquisition geometry, and temporal sampling between SAR and optical sensors may introduce additional variability when both data domains are jointly used in machine learning-based models, thereby reducing the effective information gain from SAR features. Our findings therefore suggest that the added value of SAR data is strongly context dependent. SAR inputs are likely most beneficial in situations where optical time series are severely affected by cloud cover or where optical information alone is insufficient to capture management events; whereas, under conditions of dense and high-quality optical observations, gains from additional sensor types like SAR may be limited. These results should not be interpreted as SAR data being generally uninformative but rather underline the importance of context specific sensor selection and consideration of model capacity and feature redundancy.

4.2. Performance differences across levels of grassland use intensity and regions

Overall, our results suggest that all algorithms generalized well, however, the results differed when the predictions were compared between different grassland use intensities. We observed that predictions in rather extensively used grassland areas (GLUI 1) led to lower F1-Scores with higher recall values, but considerably reduced precision values compared to intensively used grassland areas (GLUI 3). For those intensively used areas, we observed overall higher F1-Scores and, in particular, high precision values across all approaches. This pattern of accuracy metrics suggest that all algorithms only slightly overestimate mowing activity, but on the other hand miss relevant mowing events, as indicated by a reduced recall compared to GLUI 1. We observed this pattern in most regions and for both deep learning and rule-based algorithms. When considering all regions and years together, rule-based algorithms showed overall better performances in intensively used grassland areas than most deep learning algorithms, except for the IGN algorithm that outperformed all other algorithms (GLUI 3). One explanation for this finding could be the combined approach of the IGN method that uses deep learning to integrate optical and SAR data before applying decision rules to identify mowing events. This approach allows for an optimization of a dense satellite time series, which could be particularly advantageous for detecting mowing events that occur in quick succession on intensively used grassland areas. The overall higher accuracy of optical RBA approaches in intensively managed regions also supports the assumption that SAR data introduces noise in the model that negatively affects accuracy (Reinermann et al., 2022).

The higher error of commission (lower precision) in extensively used grasslands might be due to differences in specific grassland traits such as plant vigour, species composition and less distinct biomass removal events in comparison to intensively used grassland areas. These differences in grassland traits might lead to less pronounced spectral-temporal patterns, which makes the model-based differentiation of mowing events more complicated. Holtgrave et al. (2023) and Lobert et al. (2021) reported similar patterns of overestimation of mowing events on extensively used meadows and both accounted it to the longer grass on less-cut grasslands being more prone to weather events (hail, storm, drought) potentially resembling mowing events and non-reported grazing events on mixed-use grasslands. De Vroey et al. (2023) have shown that it is possible to differentiate both grassland types on a regional scale using remote sensing. Even though the differentiation of pastures and mown grasslands was not subject to this work and not feasible based on the available reference data, we highlight its importance for a holistic impact assessment of grassland management on which future research should focus.

We evaluated the algorithm performance across a broad environmental gradient in Europe, covering diverse regions. For example, our study included Eastern France with its varied topography, river valleys, hills, and plateaus or Southern France, with the Pyrenees mountains. But also, regions with maritime temperate climate like the Wallonian grasslands in Belgium, alpine areas in Austria and Switzerland (where most algorithms performed particularly well), and northern regions like Estonia and Sweden were included. The observed accuracies were mostly consistent between the dry and warm conditions of Southern France and the cooler Nordic climates, which suggests a general robustness of the algorithms to climatic or topographic gradients. Covering such large areas was only possible, however, with reference data of all quality levels. Yet parts of Europe with unique biogeographical regions and substantial shares of agricultural grasslands such as the United Kingdom, the Pannonian Basin, and the Mediterranean could not be included in this study at all.

4.3. The role of reference data

The basis for the mowing detection comparison exercise was the compilation of a unique reference dataset on mowing events. Differences in data formats, temporal resolution, and spatial context required harmonization and methodological decisions. For instance, reference events from webcam interpretation were linked to parcels, which introduces spatial uncertainties (Reinermann et al., 2023; Weber et al., 2024). Moreover, there are inconsistencies in the temporal reporting schemes. Some sources reported mowing periods rather than specific dates, which required to average intervals to a single date, which adds temporal uncertainty (De Vroey et al., 2023) and can influence the performance metrics. Although ideally all reference data records should be complete and correct, events can be missing due to underreporting of mowing events or, e.g., partial mowing of meadows (Watzig et al., 2023). These deficiencies in the reference data led to patterns where all algorithms showed higher recall but lower precision as can be seen for example for Västerbottens, Sweden (see Fig. 5). To account for those issues, we assigned quality labels to the reference data depending on their assumed reliability. However, even highest quality reference data still cannot guarantee absolute accuracy due to inherent limitations in the methods. For instance, the manual interpretation of satellite-based time series is directly influenced by the availability of clear sky observations and is subjectively dependent on the decision of the image interpreter. Field surveys or farmer reports may be affected by reporting biases or subjective interpretations of what constitutes a mowing event. In practice, several factors can have an impact on reporting on a field scale and also on the remote sensing signal, e.g., partial harvesting, surrounding field work like ploughing or haymaking or grass laying down to dry for. Given the uncertainties in the reference data, the actual accuracies of the algorithms could even be higher than reported here.

By using only highest quality reference data for the main evaluation, we were forced to reduce the number of usable parcels by 71% and restricted the analysis to fewer regions and years. Thus, it is critical to keep on collecting more high-quality reference data and making them available to the community following FAIR (Findability, Accessibility, Interoperability, and Reuse) data sharing standards (Wilkinson et al., 2016). Future work should aim to collect further high-quality reference data, for which citizen science approaches and extended calls for data collection could develop an untapped potential. Such data sets do not only allow a more generalized training of newly developed models but also refining existing approaches and independently evaluating area-wide national (Reinermann et al., 2023; Schwieder et al., 2022; Weber et al., 2024) and transnational (e.g., Copernicus Land Monitoring Service, 2024) grassland use intensity products that become more frequently available.

Furthermore, while splits with 70 to 80% of the available samples for training are common (Joseph, 2022), we only used 30% per region provided in this study to enable a meaningful evaluation with sufficient samples. More training data will therefore be beneficial for future optimization and probably even further increase the generalizability of the ML models.

4.4. Future directions

4.4.1. Data availability

Our results showed that several remote sensing-based algorithms for the detection of mowing events exist that can be applied under different environmental and management conditions. The developments of the recent years were driven by the availability of dense time series of satellite data and technical advancements in models and computational performance (Reinermann et al., 2020) as well as by the increasing demand of area-wide management information for policy monitoring and impact assessment. The current exceptional constellation of even three Sentinel-2 and two Landsat satellites in orbit, provides unprecedented availability of dense time series with potentially global coverage. The accuracy and spatial detail of mowing event detection might even advance, when data from planned future satellites, such as Sentinel-2 NG and Landsat Next, with enhanced spectral, temporal, and spatial resolution, are available. Especially, the combined use of data from different satellite sensors and systems might help to overcome temporal mismatches between actual management activity and overflight intervals, as the availability of clear sky observations varies throughout the season and across regions (Lewińska et al., 2024). Our study did not compare alternative pre-processing routines and analysis ready datasets. Although analysis ready data processed with FORCE (Frantz, 2019) appear to be effective for mowing detection, it is worth exploring whether other products with alternative cloud masking and sensor harmonization like HLS (Masek et al., 2021) are comparable for mowing detection. In parallel with optical data developments, recently launched SAR missions like NISAR (Kellogg et al., 2020) will provide L-band (global) and S-band (South Asia only) data. Those longer wavelengths come with increased canopy penetration ability and may provide complementary information to the C-band data commonly used for grassland monitoring (El Hajj et al., 2019).

4.4.2. Spatial detail

An improved spatial resolution of optical and SAR products will likely minimize the influence of mixed-pixel effects at parcel boundaries and enable to better distinguish heterogenous management within individual parcels. This effect is particularly relevant for all pixel-based algorithms that have the advantage of being applicable everywhere and are not restricted to regions where parcel boundaries are available. In this study, we aggregated all predictions to the provided or defined spatial units, assuming that they were homogeneously managed. As this assumption does not always hold true, it should be considered to approximate spatial boundaries of homogeneously managed fields, e.g., by segmentation of remote sensing data (Tetteh et al., 2021; Tetteh et al., 2023).

4.4.3. Deep learning advances

The increasing availability and rapid advancement of deep learning applications will most likely also lead to new developments in the grassland focused remote sensing community. Recent developments in semi-supervised machine learning indicate that pre-trained foundation models may have the potential to increase our capacity to assess grassland use intensity. Earth observation foundation models are transforming how we approach mapping tasks, as these models learn from massive, unlabeled datasets through self-supervised methods, capturing the multi-sensor and multi-temporal patterns inherent in the Earth's surface dynamics (e.g., Danish et al., 2025; Li et al., 2026; Szwarcman et al., 2026; Xiao et al., 2025). Due to the usual scarcity of reference data, we identify the detection of mowing events as an interesting use case to test the applicability of foundation models.

4.4.4. Applications and uncertainties

Besides these technical improvements, it is noteworthy that satellite-based grassland use intensity metrics are of high relevance for follow-up studies. Spatially explicit information on grassland use intensity has for example been used to assess the impact of grassland management on biodiversity using ecological indicators (Weber et al., 2024), specific taxa such as butterflies (Kasiske et al., 2024), butterflies and grasshoppers (Fumy et al., 2023), and wild bees (Hellwig et al., 2022). Such data further allow to integrate approximated grassland management schemes in greenhouse gas emission models (Rösemann et al., 2025) and use it for grassland yield estimations (Klingler et al., 2025; Reinermann et al., 2025). However, for such downstream applications, among others, the provision of a spatially explicit uncertainty layer would be beneficial to account for the uncertainties of mowing event predictions. Even though some approaches provide approximate estimates of prediction uncertainty based on time series characteristics (Reinermann et al., 2022; Schwieder et al., 2022), we recommend improving such information and making it a standard. Recent advances in probabilistic machine learning, such as conformal prediction (Singh et al., 2024) make pixel-level uncertainty estimation possible.

5. Conclusion

The Mowing Detection intercomparison Exercise enabled an in-depth analysis of the performance of ten independent algorithms, which were developed to identify mowing events in time series of satellite data. The algorithms were evaluated based on an unprecedented set of harmonized grassland mowing event reference data from various regions in eight different European countries and a uniform set of satellite data. While all algorithms worked generally well when applied to independent data (e.g., from regions and years for which they were not specifically developed), the most consistent performances were reached by algorithms from the domain of machine learning, with narrower ranges between precision and recall compared to rule-based approaches. Regarding the data domain, the results do not confirm a consistently positive influence of the combined use of optical and SAR data in the prediction of mowing events, as patterns were indifferent for the tested subsets (e.g., grassland use intensity classes, regions). Overall, we observed an upper limit to algorithm performance despite the usage of diverse, innovative, and state-of-the-art methods. This finding might indicate a limitation in the current freely available global optical and radar Earth observation systems for the detection of events with a narrow time window, such as grassland mowing. This limitation may be overcome with the continuation and further expansion of existing missions (e.g., Sentinel-2 Next Generation, Landsat-Next) and the inclusion of future Earth Observation missions for the generation of multi-constellation, high spectral and temporal resolution satellite data archiving and processing facilities. The comparison exercise highlights the critical need for a representative set of high-quality reference data, which accounts for regional environmental differences in the training process. Only such data enable the independent validation of increasingly frequently available area-wide regional, national, or transnational products. With MODCiX we set the starting point for an international collaboration that fosters the collection and FAIR provision of independent benchmark data on grassland use intensity. We encourage further contributions and the expansion of the dataset to include data on grassland use and management, i.e., grazing intensity, fertilization, and parameters such as biomass/yield and quality (e.g., protein content). While this comparison exercise had a focus on the estimation of mowing frequencies in meadows, we stress the need to include pastures and grazed grasslands for a holistic assessment of grassland use intensity.

CRediT authorship contribution statement

Marcel Schwieder: Writing – review & editing, Writing – original draft, Visualization, Validation, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Felix Lobert:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Dominique Weber:** Writing – review & editing, Data curation. **Sophie Reinermann:** Writing – review & editing, Data curation. **Sarah Asam:** Writing – review & editing. **Filippo Sarvia:** Writing – review & editing. **Samuele De Petris:** Writing – review & editing. **Enrico Borgogno-Mondino:** Writing – review & editing. **Arnab Muhuri:** Writing – review & editing. **Natascha Oppelt:** Writing – review & editing. **Clement Atzberger:** Writing – review & editing. **Iason Tsardanidis:** Writing – review & editing. **Charalampos Kontoes:** Writing – review & editing. **François Godechal:** Writing – review & editing. **Cozmin Lucau-Danila:** Writing – review & editing. **Viviane Planchon:** Writing – review & editing. **Anatol Garioud:** Writing – review & editing, Data curation. **Célestin Huet:** Writing – review & editing. **Silvia Valero:** Writing – review & editing. **Clément Mallet:** Writing – review & editing. **Julien Morel:** Writing – review & editing, Data curation. **Mattia Rossi:** Writing – review & editing. **Francesco Vuolo:** Writing – review & editing. **Aleksandar Dujakovic:** Writing – review & editing. **Andreas Schaumberger:** Writing – review & editing. **Andreas Klingler:** Writing – review & editing, Data curation. **Ann-Kathrin Holtgrave:** Writing – review & editing, Data curation. **Zander Venter:** Writing – review & editing. **Ruth Sonnenschein:** Writing – review & editing, Data curation. **Mathilde de Vroey:** Writing – review & editing, Data curation. **Julien Radoux:** Writing – review & editing, Data curation. **Oliver Buck:** Data curation. **Anna Katharina Franke:** Data curation. **Uta Schuhmacher:** Data curation. **Andreas Ostrowski:** Data curation. **Patrick Hostert:** Writing – review & editing. **Stefan Erasmi:** Writing – review & editing, Supervision.

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.

Data availability

We encourage all interested parties to participate in this comparison exercise by testing their own algorithms against parts of the benchmark data set used, for which open licenses apply. Please use the interactive Shiny-App https://modcix.thuenen.de to explore the results of this paper and to upload your own mowing event predictions for independent validation. Reference data can be accessed through this Zenodo repository: https://doi.org/10.5281/zenodo.18834294 and used for model optimization and predictions. An R-Notebook with code examples used for the comparison can be accessed through this GitHub repository: https://github.com/geo-masc/modcix. This work includes data elaborated by several projects of the Biodiversity Exploratories program (DFG Priority Program 1374). The related datasets are listed in the references section.

Acknowledgements

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. We thank the field operators from SITES Röbäcksdalen Field Station for collecting the Swedish dataset. NINA's contribution was facillitated by the GreeNet project which is funded by Biodiversa+. We thank Kaupo Voormansik and Liza Vabištševitš for sharing reference data for Estonia as well as Emilie Beriaux and Alban Jago for supporting the project while being affiliated to the Walloon Agricultural Research Centre. We further acknowledge Annika Ludwig for her valuable assistance in interpreting webcam images in South Tyrol. We also thank the Autonomous Province of Bozen/Bolzano - South Tyrol for funding the international mobility programme for scientists (Decreto 20176/2022), which supported this research within the GRITA-EO project. This publication was prepared within the framework of MonViA 'Monitoring of Biodiversity in Agricultural Landscapes' (www.agrarmonitoring-monvia.de/en). MonViA is funded by the German Federal Ministry of Agriculture, Food and Regional Identity (BMLEH).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.rse.2026.115466.

References

Bengtsson, J., Bullock, J.M., Egoh, B., Everson, C., Everson, T., O'Connor, T., O'Farrell, P. J., Smith, H.G., Lindborg, R., 2019. Grasslands—more important for ecosystem services than you might think. Ecosphere 10. https://doi.org/10.1002/ecs2.2582.

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

Chang, J., Ciais, P., Gasser, T., Smith, P., Herrero, M., Havlik, P., Obersteiner, M., Guenet, B., Goll, D.S., Li, W., Naipal, V., Peng, S., Qiu, C., Tian, H., Viovy, N., Yue, C., Zhu, D., 2021. Climate warming from managed grasslands cancels the cooling effect of carbon sinks in sparsely grazed and natural grasslands. Nat. Commun. 12, 118. https://doi.org/10.1038/s41467-020-20406-7.

Copernicus, 2023. Copernicus Digital Elevation Model GLO-30. https://doi.org/10.5270/ESA-c5d3d65.

Danish, M.S., Munir, M.A., Shah, S.R.A., Khan, M.H., Anwer, R.M., Laaksonen, J., Khan, F.S., Khan, S., 2025. TerraFM: a scalable foundation model for unified multisensor. Earth Observ. arXiv 2506.06281 https://arxiv.org/abs/2506.06281.

De Vroey, M., Radoux, J., Defourny, P., 2021. Grassland mowing detection using Sentinel-1 time series: potential and limitations. Remote Sens 13, 348. https://doi.org/10.3390/rs1303034.

De Vroey, M., de Vendictis, L., Zavagli, M., Bontemps, S., Heymans, D., Radoux, J., Koetz, B., Defourny, P., 2022. Mowing detection using Sentinel-1 and Sentinel-2 time series for large scale grassland monitoring. Remote Sens. Environ. 280, 113145. https://doi.org/10.1016/j.rse.2022.113145.

De Vroey, M., Radoux, J., Defourny, P., 2023. Classifying sub-parcel grassland management practices by optical and microwave remote sensing. Remote Sens 15, 181. https://doi.org/10.3390/rs1501018.

El Hajj, M., Baghdadi, N., Bazzi, H., Zribi, M., 2019. Penetration analysis of SAR signals in the C and L bands for wheat, maize, and grasslands. Remote Sens 11 (1), 31. https://doi.org/10.3390/rs11010031.

European Environment Agency, 2016. Biogeographical Regions, Europe 2016, ver. 1.

Eurostat, 2020. Share of Main Land Types in Utilised Agricultural Area (UAA) by NUTS 2 Regions. Eurostat, Luxembourg.

Frantz, D., 2019. FORCE—Landsat + Sentinel-2 analysis ready data and beyond. Remote Sens 11, 1124. https://doi.org/10.3390/rs11091124.

Fumy, F., Schwarz, C., Fartmann, T., 2023. Intensity of grassland management and landscape heterogeneity determine species richness of insects in fragmented hay meadows. Global Ecol. Conserv. 47, e02672. https://doi.org/10.1016/j.gecco.2023.e02672.

Garioud, A., Valero, S., Giordano, S., Mallet, C., 2021. Recurrent-based regression of sentinel time series for continuous vegetation monitoring. Remote Sens. Environ. 263. https://doi.org/10.1016/j.rse.2021.112419.

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., Pašalić, E., Penone, C., Perović, 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.

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.
Habel, J.C., Dengler, J., Janišová, M., Török, P., Wellstein, C., Wiezik, M., 2013. European grassland ecosystems: threatened hotspots of biodiversity. Biodivers. Conserv. 22, 2131–2138. https://doi.org/10.1007/s10531-013-0537-x.
Hellwig, N., Schubert, L.F., Kirmer, A., Tischew, S., Dieker, P., 2022. Effects of wildflower strips, landscape structure and agricultural practices on wild bee assemblages – a matter of data resolution and spatial scale? Agric. Ecosyst. Environ. 326, 107764. https://doi.org/10.1016/j.agee.2021.107764.
Holtgrave, A.-K., Lobert, F., Erasmi, S., Röder, N., Kleinschmit, B., 2023. Grassland mowing event detection using combined optical, SAR, and weather time series. Remote Sens. Environ. 295, 113680. https://doi.org/10.1016/j.rse.2023.113680.
Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., Ferreira, L.G., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sens. Environ. 83, 195–213. https://doi.org/10.1016/S0034-4257(02)00096-2.
Service, Copernicus Land Monitoring, 2024. Grassland mowing events 2017 - present (raster 10m), europe, yearly, Nov 2024. European Environment Agency. https://doi.org/10.2909/114e8cae-1cd7-4adc-8c5f-a04863fc6af9.
Joseph, V.R., 2022. Optimal ratio for data splitting. Stat. Anal. Data Min: ASA Data Sci. J. 15, 531–538. https://doi.org/10.1002/sam.11583.
Kasiske, T., Dauber, J., Dieker, P., Harpke, A., Klimek, S., Kühn, E., Levers, C., Schwieder, M., Settele, J., Musche, M., 2024. Assessing landscape-level effects of permanent grassland management and landscape configuration on open-land butterflies based on national monitoring data. Biodiversity Conserv 33 (8), 2381–2404. https://doi.org/10.1007/s10531-024-02861-6.
Kellogg, K., Hoffman, P., Standley, S., Shaffer, S., Rosen, P., Edelstein, W., Dunn, C., Baker, C., Barela, P., Shen, Y., Guerrero, A.M., Xaypraseuth, P., Sagi, V.R., Sreekantha, C.V., Harinath, N., Kumar, R., Bhan, R., Sarma, C.V.H.S.. NASA-ISRO Synthetic Aperture Radar (NISAR) Mission. https://doi.org/10.1109/AERO47225.2020.9172638.
Klingler, A., Dujakovic, A., Vuolo, F., Mayer, K., Gaier, L., Pötsch, E.M., Schaumberger, A., 2025. Prediction of grassland yield in Austria: A machine learning approach based on satellite, weather, and extensive in situ data. Eur. J. Agron. 170, 127701. https://doi.org/10.1016/j.eja.2025.127701.
Knozowski, P., Nowakowski, J.J., Stawicka, A.M., Gorski, A., Dulisz, B., 2023. Effect of nature protection and management of grassland on biodiversity - case from big flooded river valley (NE Poland). Sci. Total Environ. 898, 165280. https://doi.org/10.1016/j.scitotenv.2023.165280.
Kolecka, N., Ginzler, C., Pazur, R., Price, B., Verburg, P., 2018. Regional scale mapping of grassland mowing frequency with Sentinel-2 time series. Remote Sens 10, 1221. https://doi.org/10.3390/rs10081221.
Komisarenko, V., Voormansik, K., Elshawi, R., Sakr, S., 2022. Exploiting time series of Sentinel-1 and Sentinel-2 to detect grassland mowing events using deep learning with reject region. Sci. Rep. 12, 983. https://doi.org/10.1038/s41598-022-04932-6.
Lewińska, K.E., Frantz, D., Leser, U., Hostert, P., 2024. Usable observations over Europe: evaluation of compositing windows for Landsat and Sentinel-2 time series. Eur. J. Remote Sens. 57, 2372855. https://doi.org/10.1080/22797254.2024.2372855.
Li, X., Li, C., Ghamisi, P., Hong, D., Benediktsson, J.A., Chanussot, J., 2026. FlexiMo: A flexible remote sensing foundation model. IEEE Trans. Geosci. Remote Sens. 64, 1–16. https://doi.org/10.1109/TGRS.2026.3656362.
Lobert, F., Holtgrave, A.-K., Schwieder, M., Pause, M., Vogt, J., Gocht, A., Erasmi, S., 2021. Mowing event detection in permanent grasslands: systematic evaluation of input features from Sentinel-1, Sentinel-2. and Landsat 8 time series. Remote Sens. Environ. 267, 112751. https://doi.org/10.1016/j.rse.2021.112751.
European Commission, D.-G.f.E, 2020. EU biodiversity strategy for 2030: Bringing nature back into our lives. European Commission, Brussels, Belgium, p. 13. https://doi.org/10.2779/677548.
MacDougall, A.S., Vanzant, B., Sulik, J., Bagchi, S., Naidu, D., Muraina, T.O., Seabloom, E.W., Borer, E.T., Wilfahrt, P., Slette, I., Hierro, J.L., Pearson, D.E., Abedi, M., Akasaka, M., Alberti, J., Aleksanyan, A., Amisu, A.A., Anderson, T.M., Arnillas, C.A., Ayer, M., Bakker, J.D., Basant, S., Basto, S., Biederman, L., Bloodworth, K.J., Boscutti, F., Boughton, E.H., Bruschetti, C.M., Buckley, H.L., Buckley, Y.M., Bugalho, M.N., Caldeira, M.C., Campetella, G., Cannone, N., Carbognani, M., Carbutt, C., Carniello, M.A., Cervellini, M., Chaudhary, T., Chen, Q., Clark, A.T., Cousins, S., Dalle Fratte, M., Day, N.J., Deák, B., Dietrich, J., Dixon, A., Eisenhauer, N., Elgersma, K.J., Eren, O., Eskelinen, A., Estrada, C., Fay, P.A., Fayvush, G., Flynn, K.C., García Meza, D., Gargano, D., Gherardi, L., Girkin, N.T., González, L., Graff, P., Hagenberg, L.W.C., Halbritter, A.H., Havrilchak, N.A., Herdoiza, N., Hersch-Green, E., Hopping, K., Jentsch, A., Jimoh, S.O., Kerby, J., Kirkman, K., Knops, J.M.H., Koerner, S.E., Koltz, A., Komatsu, K.J., Koura, B.I., Kruse, S., Laanisto, L., Lannes, L.S., Li, W., Liang, M., Lkhagva, A., López-Olmedo, L., Lorenzo, P., Lortie, C.J., Loydi, A., Luo, W., Macek, P., Malfasi, F., Mariotte, P., Martina, J.P., Martínez-Blancas, A., Martinson, H., Martorell, C., Meave, J.A., Medina-Villar, S., Mganga, K.Z., Monsimet, J., Nerlekar, A.N., Niu, S., Ohlert, T., Oliveras Menor, I., Oñatibia, G.R., Ortega, Y.K., Osborne, B., Palpurina, S., Pascual, J., Pennings, S.C., Pérez-García, E., Peri, P.L., Petit Bon, M., Petraglia, A., Pijcke, F., Prober, S.M., Quiroga, R.E., Ramirez, J.I., Reed, S., Rosado, B.H.P., Roscher, C., Rowley, D.W., Sereda, I., Small, D.M., Smith, N.G., Song, Y., Stevens, C., Suarez Jimenez, L.E., te Beest, M., Tedder, M., Terry, R.S., Thornton, K.S., Tian, D., Titcomb, G., Valkó, O., Ciska Veen, G.F., Virtanen, R., Welti, E.A.R., Wheeler, G.R., Wolf, A.A., Wolff, P., Young, A.L., Young, H.S., Zeglin, L.H., Zhu, K., Zong, S., Siewert, M.B., 2026. The global extent of the grassland biome and implications for the terrestrial carbon sink. Nat. Ecol. Evol. 10, 246–257. https://doi.org/10.1038/s41559-025-02955-6.
Masek, J., Ju, J., Roger, J.-C., Skakun, S., Vermote, E., Claverie, M., Dungan, J., Yin, Z., Freitag, B., Justice, C., 2021. HLS operational land imager surface reflectance and TOA brightness daily global 30m v2. 0 [data set]. In: NASA EOSDIS Land Processes Distributed Active Archive Center. HLSL30. 002.
Neyret, M., Fischer, M., Allan, E., Hölzel, N., Klaus, V.H., Kleinebecker, T., Krauss, J., Le Provost, G., Peter, S., Schenk, N., Simons, N.K., van der Plas, F., Binkenstein, J., Börschig, C., Jung, K., Prati, D., Schäfer, D., Schäfer, M., Schöning, I., Schrumpf, M., Tschapka, M., Westphal, C., Manning, P., 2021. Assessing the impact of grassland management on landscape multifunctionality. Ecosyst. Serv. 52, 101366. https://doi.org/10.1016/j.ecoser.2021.101366.
Petermann, J.S., Buzhdygan, O.Y., 2021. Grassland biodiversity. Curr. Biol. 31, R1195–R1201. https://doi.org/10.1016/j.cub.2021.06.060.
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.
Reinermann, S., Gessner, U., Asam, S., Ullmann, T., Schucknecht, A., Kuenzer, C., 2022. Detection of grassland mowing events for Germany by combining Sentinel-1 and Sentinel-2 time series. Remote Sens 14, 1647. https://doi.org/10.3390/rs14071647.
Reinermann, S., Asam, S., Gessner, U., Ullmann, T., Kuenzer, C., 2023. Multi-annual grassland mowing dynamics in Germany: spatio-temporal patterns and the influence of climate, topographic and socio-political conditions. Front. Environ. Sci. 11. https://doi.org/10.3389/fenvs.2023.1040551.
Reinermann, S., Boos, C., Kaim, A., Schucknecht, A., Asam, S., Gessner, U., Annuth, S.H., Schmitt, T.M., Koellner, T., Kiese, R., 2025. Grassland yield estimations - potentials and limitations of remote sensing, process-based modelling and field measurements. EGUsphere 2025, 1–34. https://doi.org/10.5194/egusphere-2024-4087.
Rösemann, C.V.C., Haenel, H.-D., Dämmgen, U., Döring, U., Wulf, S., Eurich-Menden, B., Döhler, H., Steuer, B., Osterburg, B., Fuß, R., 2025. Calculations of Gaseous and Particulate Emissions from German Agriculture 1990–2023 : Report on Methods and Data (RMD) Submission 2025. https://git-dmz.thuenen.de/vos/emissionsagriculture2025.
Schils, R.L.M., Bufe, C., Rhymer, C.M., Francksen, R.M., Klaus, V.H., Abdalla, M., Milazzo, F., Lellei-Kovács, E., Berge, H.t., Bertora, C., Chodkiewicz, A., Dămătîrcă, C., Feigenwinter, I., Fernández-Rebollo, P., Ghiasi, S., Hejduk, S., Hiron, M., Janicka, M., Pellaton, R., Smith, K.E., Thorman, R., Vanwalleghem, T., Williams, J., Zavattaro, L., Kampen, J., Derkx, R., Smith, P., Whittingham, M.J., Buchmann, N., Price, J.P.N., 2022. Permanent grasslands in Europe: land use change and intensification decrease their multifunctionality. Agric. Ecosyst. Environ. 330. https://doi.org/10.1016/j.agee.2022.107891.
Schwieder, M., Wesemeyer, M., Frantz, D., Pfoch, K., Erasmi, S., Pickert, J., Nendel, C., Hostert, P., 2022. Mapping grassland mowing events across Germany based on combined Sentinel-2 and. Landsat 8 time series. Remote Sens. Environ. 269, 112795. https://doi.org/10.1016/j.rse.2021.112795.
Singh, G., Moncrieff, G., Venter, Z., Cawse-Nicholson, K., Slingsby, J., Robinson, T.B., 2024. Uncertainty quantification for probabilistic machine learning in earth observation using conformal prediction. Sci. Rep. 14, 16166. https://doi.org/10.1038/s41598-024-65954-w.
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.
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, 542. https://doi.org/10.3390/rs11050542.
Stumpf, F., Schneider, M.K., Keller, A., Mayr, A., Rentschler, T., Meuli, R.G., Schaepman, M., Liebisch, F., 2020. Spatial monitoring of grassland management using multi-temporal satellite imagery. Ecol. Indic. 113, 106201. https://doi.org/10.1016/j.ecolind.2020.106201.
Szwarcman, D., Roy, S., Fraccaro, P., Gíslason, Þ.E., Blumenstiel, B., Ghosal, R., Oliveira, P.H.d., Almeida, J.L.d.S., Sedona, R., Kang, Y., Chakraborty, S., Wang, S., Gomes, C., Kumar, A., Gaur, V., Truong, M., Godwin, D., Khallaghi, S., Lee, H., Hsu, C.Y., Asanjan, A.A., Mujeci, B., Shidham, D., Balogun, R.O., Kolluru, V., Keenan, T., Arevalo, P., Li, W., Alemohammad, H., Olofsson, P., Mayer, T., Hain, C., Kennedy, R., Zadrozny, B., Bell, D., Cavallaro, G., Watson, C., Maskey, M., Ramachandran, R., Moreno, J.B., 2026. Prithvi-EO-2.0: A Versatile Multitemporal Foundation Model for Earth Observation Applications. IEEE Transactions on Geoscience and Remote Sensing 64, 1–20. https://doi.org/10.1109/TGRS.2025.3642610.
Taravat, A., Wagner, M.P., Oppelt, N., 2019. Automatic grassland cutting status detection in the context of spatiotemporal Sentinel-1 imagery analysis and artificial neural networks. Remote Sens 11, 711. https://doi.org/10.3390/rs11060711.
Tetteh, G.O., Gocht, A., Erasmi, S., Schwieder, M., Conrad, C., 2021. Evaluation of Sentinel-1 and Sentinel-2 feature sets for delineating agricultural fields in heterogeneous landscapes. IEEE Access 9, 116702–116719. https://doi.org/10.1109/ACCESS.2021.3105903.
Tetteh, G.O., Schwieder, M., Erasmi, S., Conrad, C., Gocht, A., 2023. Comparison of an Optimised Multiresolution Segmentation Approach with Deep Neural Networks for Delineating Agricultural Fields from Sentinel-2 Images. PFG – Journal of Photogrammetry. Remote Sensing and Geoinformation Science 91, 295–312. https://doi.org/10.1007/s41064-023-00247-x.
Tsardanidis, I., Koukos, A., Sitokonstantinou, V., Drivas, T., Kontoes, C., 2025. Cloud gap-filling with deep learning for improved grassland monitoring. Comput. Electron. Agric. 230, 109732. https://doi.org/10.1016/j.compag.2024.109732.
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., Schumacher, U., Schreiber, C., Lauterbach, R., Franke, A., Ostrowski, A., Miriam, T., 2025. Land use in grasslands: raw data of yearly owner interviews. Version 69. Biodiversity Exploratories Information Syst. https://www.bexis.uni-jena.de (Dataset ID= 26487).

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

Watzig, C., Schaumberger, A., Klingler, A., Dujakovic, A., Atzberger, C., Vuolo, F., 2023. Grassland cut detection based on Sentinel-2 time series to respond to the environmental and technical challenges of the Austrian fodder production for livestock feeding. Remote Sens. Environ. 292, 113577. https://doi.org/10.1016/j.rse.2023.113577.

Weber, D., Schwieder, M., Ritter, L., Koch, T., Psomas, A., Huber, N., Ginzler, C., Boch, S., 2023. Grassland-Use Intensity Maps for Switzerland. EnviDat. https://doi.org/10.16904/envidat.428.

Weber, D., Schwieder, M., Ritter, L., Koch, T., Psomas, A., Huber, N., Ginzler, C., Boch, S., 2024. Grassland-use intensity maps for Switzerland based on satellite time series: challenges and opportunities for ecological applications. Rem. Sens. Ecol. Conserv. 10, 312–327. https://doi.org/10.1002/rse2.372.

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., 't Hoen, P.A.C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., Mons, B., 2016. The FAIR guiding principles for scientific data management and stewardship. Sci. Data 3, 160018. https://doi.org/10.1038/sdata.2016.18.

Xiao, A., Xuan, W., Wang, J., Huang, J., Tao, D., Lu, S., Yokoya, N., 2025. Foundation models for remote sensing and earth observation: A survey. IEEE Geosci. Remote Sens. Magazine 13, 297–324. https://doi.org/10.1109/MGRS.2025.3576766.