Nationwide operational mapping of grassland mowing events combining machine learning and Sentinel-2 time series

Henry Rivas, Mathieu Fauvel, Vincent Thiérion, Millet Jerôme, Laurence Curtet

Nationwide operational mapping of grassland mowing events combining machine learning and Sentinel-2 time series

Henry RIVAS...

September 2023

1 Abstract

2 Introduction

Grasslands cover approximately 40% of the Earth's land area, encompassing nearly 70% of the global agricultural land area, and are distributed on all continents and across all latitudes (Suttie et al., 2005; White et al., 2000). Grassland dynamics influence global ecosystem functioning, and their impact is widely modulated by management practices intensity on these landscapes (Zhao et al., 2020). Management practices are primarily driven by grassland landscape maintenance, as well as by ecosystem service of provisioning offered by the grasslands. Grasslands are subject to management practices such as mowing or grazing or a combination of both. Therefore, monitoring grassland management practices is essential for assessing management intensity level, which in turn plays a critical role in studies related to biodiversity (XXXX), water (XXXXX) and carbon (XXXXX) cycling and others topics (XXXX). In France, the National Observatory of Mowed Grassland Ecosystems conducts birdlife monitoring in mowed grasslands, with a particular focus on the rise in breeding failures attributed to increasingly early mowing. Early mowing intercepts birds' reproductive period and interrupts their breeding process (Broyer et al., 2012). Usually, responsible agencies conduct occasional observation campaigns to support ecosystem-related public policies, but ground observations are not spatially exhaustive and are time-consuming. As an alternative source, synoptic remote sensing data enables regular and global-scale monitoring, enabling tracking of vegetation dynamics. Currently, Sentinel-2 mission provides cost-free high resolution data at 10m spatial resolution with a 5-day temporal frequency (10 days before 2017), allowing intra-plot level observations. Grassland mowing events timing and intensity have already been mapped using remote sensing-based time series, mainly from features sensitive to vegetation status, such as Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI), Leaf Area Index (LAI) and more. There have been several methods used to detect mowing events from satellite time series. These methods were mainly based on temporal changes in time series using threshold-based methods and anomalies detection approach. More recently, deep learning-based architectures were also used to detect mowing events timing.

Estel et al. (2018) assessed grassland use intensity spatial patterns across Europe. To extract annual mowing frequency, a temporal change analysis based on spline-adjusted MODIS NDVI time series was used. Their approach involved identifying mowing events as instances where a local minima exhibited a change, relative to its preceding peak, exceeding 10% of growing season amplitude. The results showed an overall accuracy of 80%, which decreases as the frequency of events increases. In northern Switzerland, Kolecka et al. (2018) also estimated mowing frequency employing similar temporal change analysis, but based on raw Sentinel-2 NDVI time series. Here, a drop in NDVI greater than 0.2, between two consecutive cloud-free acquisition dates, was counted as a mowing event. Their method accurately identified 77% of observed events and highlighted that false detection can occur due to residual cloud presence, while sparse time series led to the omission of mowing events. Regarding Griffiths et al. (2020), mowing events frequency and timing were mapped in Germany using 10-day composite Harmonized Landsat-Sentinel NDVI time series. Discrepancies between a hypothetical bell-shaped curve and the current polynomial-fitted curve were evaluated. An event was counted when the difference exceeded 0.2 NDVI. Findings revealed consistent spatial patterns in mowing frequency (indicating extensive and intensive management). However, estimated dates exhibited significant discrepancies compared to observed dates (MAE $>$ 50 days), which could be due to lower temporal resolution of Sentinel-2 before 2017 and the absence of reliable ground data for calibration and validation. Stumpf et al. (2020) mapped grassland management (grazing or mowing) and its intensity based on biomass productivity and management frequency, respectively. The latter were extracted from n-day composite Landsat ETM + and Landsat OLI NDVI time series. As in previous cases, a management event was counted when NDVI loss is higher than a threshold, which was based on the probability density function of all NDVI changes across the time series and was specified for p = 0.01. Their approach yielded management patterns consistent with several management-related indicators (species richness, nutrient supply, slope, etc). Recently, Watzig et al. (2023) estimated mowing events in Austria, using Sentinel-2 NDVI time series and implementing discrepancy analysis between a idealized unmowed trajectory and actual NDVI values. An event was recorded if the difference exceeded -0.061. Commission errors due to residual clouds were reduced via a subsequent binary classification of each estimated event using a gradient boosting algorithm trained over cloudy plots. Findings indicated an overall accuracy of 80% in correct event detection, with estimated dates closely aligning with observed dates (MAE $<$ 5 days). Vroey et al. (2022) developed a algorithm for detecting mowing events across Europe. Here, raw Sentinel-2 NDVI and Sentinel-1 VH-coherence time series were used separately.

A mowing event was deemed when temporal change exceeded -0.15 NDVI and 1.0 VH-coherence standard deviation, multiplied by a factor (3.0×10^{-7}), respectively. VH-coherence standard deviation was calculated from residuals of the six preceding observations. These residuals capture disparities between linear-fitted values and actual values. In the final product, Sentinel-1 outputs were considered when Sentinel-2 omitted events due to cloud cover. Results demonstrated synergy between optical and radar data in detecting mowing events (F1-score of 79%). Using only Sentinel-2 data achieved maximum precision, but combining both sensors boosted recall significantly. In the same focus to evaluate optical and radar data synergy, Reinermann et al. (2022) mapped mowing frequency across Germany, from Sentinel-2 EVI and Sentinel-1 PolSAR entropy time series separately. A mowing event was counted when temporal change exceeded -0.07 EVI, which was calculated between two consecutive critical points (local minima and its preceding local maxima). S1-based detection was used to find potentially missed mowing events in cloudy gaps ($>$ 25 days) in optical observations. Here, the change needed to exceed 0.05 entropy between a peak and a preceding trough. Findings showed that S2-based method correctly detected 60.3% of mowing events with an F1-Score of 0.64. However, combining S1 and S2 increased recall but also caused more false positives, lowering precision. To reduce cloudy gap in optical time series, Schwieder et al. (2022) combined Sentinel-2 and Landsat-8 EVI time series for mowing events detection in Germany. They analyzed the discrepancies between actual observations and an idealized temporal profile (unmowed regime). An event was recorded when difference exceeded the mean value of all absolute residuals. Also, the detected point needed a loss greater than 1.0 standard deviation (of actual time series) compared to the previous point. Overall, detected mowing dates exhibited an average absolute difference < 12 days compared to observed dates. Mowing events were detected with an average F-score of 0.60, while the estimation of their frequency showed a mean error $< 40\%$ of the actual number of mowing events. They highlighted that performance was lower in areas with less cloud-induced observations. While threshold-based methods have shown promise in detecting mowing events, some more complex approaches have also been explored. Komisarenko et al. (2022) estimated mowing events timing at plot level in Estonia, using a 1-D Convolutional Neural Networks (CNN) on Sentinel-2- and Sentinel-1-based features time series. Although fourteen features were used, NDVI and the harmonic mean of VV and VH coherence were considered the most relevant. Their approach yielded an accuracy of 73%, outperforming similar ones. Here, most of the incorrectly estimated events were observed when optical time series were sparse or the size of the plot was small. For three regions in Germany, Lobert et al. (2021) also used a similar deep learning approach (1-D CNN) on Sentinel-2/Landsat-8- (data cube) and Sentinel-1-based features time series for mowing event frequency and timing detection. Among all studied feature combinations, the highest overall accuracy was reached when combined NDVI, backscatter cross-ratio and coherence with an F1-Score of 0.84. Estimated mowing dates showed a MAE of 3.79 days compared with the observed dates. In terms of management intensity, low-intensity grasslands were overestimated, while high-intensity grasslands were underestimated.

********** Our approach:

In France, local remote sensing-based studies have already been conducted to discriminate grassland management practices in the northwest (Dusseux et al., 2014) and detect mowing events in the southeast (Courault et al., 2010).

Why first event only??? Considering environmental challenges in mowed grassland, as an indicator of management intensity, fulfillment of ecological policies for avifauna conservation, to support grassland management monitoring system at national level in France

Here, as a complement to previous efforts, we focused on mapping grassland first mowing event date using Sentinel-2-based features time series, primarily to support grassland management monitoring system at national level in France. Based on the state-of-the-art, we conducted a comprehensive evaluation of both machine learning- and threshold-based approaches, either already implemented in mowing event detection or chosen for their promising potential in this specific task. These methods were implemented via Iota^2 (https://docs.iota2.net/).

How is the document organized?

This paper is organized as follow : ...

3 Methods

3.1 Study area

Our study area covers grassland across France, which represent about 32% (84 225 km^2) of the agricultural area declared in the metropolitan Land Parcel Identification System (Cantelaube & Carles, 2014) in 2022 (Figure 1). Grasslands cover regions that are less suitable for agricultural activities due to unfavorable climatic or site conditions (high altitudes, steep slopes, poor or wet soils). In France, grasslands are found in mountain chains in the center (Massif Central), western (Massif Armoricain), eastern (Jura and Vosges), Alps and Pyrenees, as well as in plains and wet regions. These grasslands are often divided into small plots, with 75% of them covering less than 3.0 hectares. Larger plots, exceeding 5.0 hectares, are mainly concentrated in specific regions like the Massif Central and Jura. Here, grasslands predominantly fall into two categories: permanent grasslands, characterized by uninterrupted herbaceous cover for over 5 years (84% of declared grassland area in 2022), and temporary grasslands. Both permanent and temporary grasslands are subject to management practices such as mowing or grazing or a combination of both. The intensity of these practices varies widely between plots, influenced by factors such as climate, altitude, accessibility, and individual farmer preferences. Lower altitudes tend to offer more favorable conditions for mowing and more intensive management. In France, grassland growing season spans from spring to autumn (March to October) and mowed grasslands are generally managed extensively, with one or two mowings per year, but in some cases they can be managed more intensively, with up to six mowings per year. From an ecological point of view, timing of first mowing event is more important than frequency of mowing events along growing season. In intensive regime, first mowing event happens before June 15th, while in extensive one, it occurs after that date. Extensive management practices are beneficial for biodiversity (Petermann & Buzhdygan, 2021) and birdlife (Broyer et al., 2012) and are actively promoted by the Common Agricultural Policy through incentivized payment mechanisms.

In our study area, according to Köppen–Geiger classification (Peel et al., 2007), the climate is mainly oceanic, with warm summers throughout the country and a Mediterranean climate in the south. Annual rainfall is around 800-1 000 (mm), with a contrast between the western ($>$ 1 000 mm) and the southeastern (600-800 mm) regions. The average annual temperature is about 11-13 °C, with 20-25 degrees in summer and 5-10 degrees in winter (https://meteofrance.com).

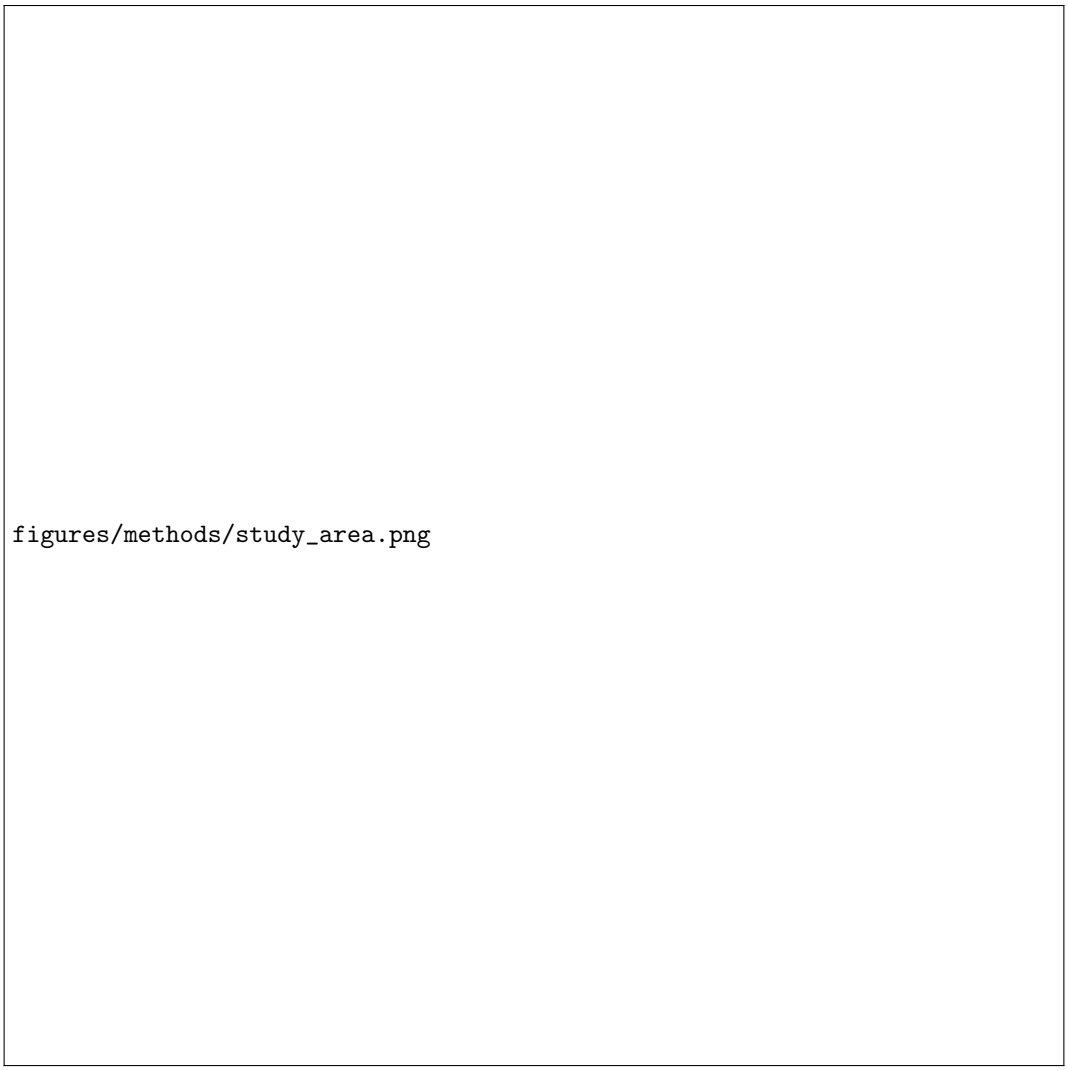

Figure 1: Study area.

3.2 Reference data

In 2022, the French Agency of Biodiversity https://www.ofb.gouv.fr) coordinated an intensive campaign of ground observations throughout the country, involving local government agencies participating in the National Observatory of Mowed Grasslands Ecosystem network. Observations were conducted once a week from May to August, covering a total of 2 227 plots evenly distributed across eight specific sites (Figure 1 and Table 1). These sites were selected to ensure diversity in ecology, topography, and management intensity (Figure 1 and Table 1). For each specific site, permanent grassland plots were obtained from the 2020 Land Parcel Identification System, a database derived from farmers' official declarations (Cantelaube & Carles, 2014). This database provides information about plot polygons and crop types. However, each plot may have more than one management at a given time (mowing and grazing). Therefore, we visually assessed each plot using a national database of aerial imagery (BD ORTHO, https://geoservices.ign.fr/bdortho) and Google Earth to identify and separate sub-plots with homogeneous spatial structure. For each actual plot, a total of eleven observations were conducted throughout growing season. At each weekly visit, current management practice (mowing or grazing) was recorded. An event could be ongoing during the visit or have occurred between the current and previous visits. Consequently, observed date for a mowing event may have an uncertainty of up to 7 days (Δt). At the end of observation campaign, each plot was labeled as mowed, grazed or mixed (mowing + grazing) and the date of occurrence of each mowing event was recorded. Here,

grasslands were mowed up to twice per year, but most of them were mowed only once (75% of plots). Finally, an additional site (site 9) was included, where 38 plots were observed with a lower temporal resolution. These grasslands are located in a national park (Parc Naturel Régional des Causses du Quercy) in the south of the country and were monitored by local contributors (Table 1). Therefore, we established a database where every plot was labeled in a management class. In the case of mowed plots, we additionally recorded the dates of each observed mowing event.

	Number of plots	Average area (Ha)	Mowed plots (%)	Altitude (m)	Key features
site 1	212	1.39	64	2-50	low-altitude flood grasslands
site 2	325	1.06	73	30	low-altitude flood grasslands
site 3	267	2.47	44	230-280	intensive low-plateaux grasslands
site 4	288	2.23	60	280-390	low-altitude unflood grasslands
site 5	272	2.55	78	800-950	mid-plateaux grasslands
site 6	267	3.87	82	800-850	mid-altitude flood grasslands
site 7	300	1.66	85	1100-1300	mid-plateaux grasslands
site 8	296	1.40	73	1000-1050	mid-plateaux grasslands
site 9	38	0.50	86		

Table 1: Reference data.

3.3 Satellite data

All available Sentinel-2 surface reflectance L2A images captured throughout the growing season (from January to September 2022), intersecting our study area were used. All spectral bands (except B1, B9 and B10) were used after resampling at 10 m to standardize pixel size between remaining bands. These images had been preprocessed using MAJA algorithm (Lonjou et al., 2016) for atmospheric correction and cloud detection, and were downloaded from THEIA platform (https://www.theia-land.fr). Here, images were masked from clouds and shadows resulting in a cloud-free time series. We processed a total of ninety tiles, each having sixty images on average. Additional preprocessing were performed separately according to the task (Table 2). We determined task-specific optimal preprocessing and features based on top-performing results from several initial tests. For classification task, we generated a linearly interpolated time series, with a regular 10-day time interval. In addition to original spectral bands, we also computed three spectral indices: Normalized Difference Vegetation Index - NDVI (Rouse et al., 1974), Normalized Difference Water Index - NDWI (McFeeters, 1996) and Brightness Index - BI (Escadafal, 1989). For mowing detection task, we generated an independent data-cube according to applied approach. For machine learning methods, we produced an interpolated time series as in classification task above. Here, however, we only used original spectral bands and their first derivative. For threshold-based methods, we computed NDVI time series using non-interpolated raw data.

Task	Approach	Spectral bands	Derived features	Data
Classification		B2, B3, B4, B5, B6, B7, B8, B8A, B11, B12	NDVI, NDWI, BI	linear
Mowing detection	Machine learning	B2, B3, B4, B5, B6, B7, B8, B8A, B11, B12	band-specific 1st derivative	linear
	Threshold-based		NDVI	

Table 2: Satellite data.

3.4 Grassland management map

A map of grassland management practices was generated to constrain mowing date estimation to areas of mowed grassland. We performed a pixel-based classification task within a nationwide grassland mask (Figure 1), derived from permanent grassland plots declared in the 2022 Land Parcel Identification System (Cantelaube & Carles, 2014). This database only provides information about plot polygons and crop types, but not about management practices. We trained a Shark Random forests classifier (Breiman, 2001) using

a grassland management practices dataset (see section 3.2). From this dataset, we redefined two distinct classes : "mowed" including mowed and mixed grasslands (1 605 plots), and "unmowed" including grazed grasslands (660 plots). Reference data were split into a 70% training dataset and a 30% test dataset, ensuring classes and sites representation through stratified sampling. Task-specific Sentinel-2-based time series were used as predictor (section 3.3 and Table 2). Hence, every pixel served as a sample, its temporal profile served as the features, and corresponding plot label served as target value. The classifier algorithm parameters were set to their default values as specified in the Orfeo Toolbox - TrainImagesClassifier module documentation (https://www.orfeo-toolbox.org), and were implemented via IOTA^2 software (https://docs.iota2.net/master/). Grassland management map achieved an overall precision of 90%, with "mowed" class showing an F-score of 0.93 and "unmowed" class exhibiting an F-score of 0.81. Findings showed that "mowed" class was slightly overestimated. In addition, in each plot of the initial grassland mask, resulting classes exhibited a coherent spatial structure, showing unimodal or bimodal intra-plot management patterns.

3.5 Mowing events detection

We tested several generic regression methods, as well as specific mowing event detection methods found in recent literature, including both machine learning- and threshold-based approaches (Figure 2). Here, we estimated first mowing event date for each pixel classified as "mowed" grassland. All methods share one underlying hypothesis: mowing event induces a sudden shift in the temporal patterns of features, leading to a decrease in values for features related to vegetative activity (e.g. NIR, vegetation indices...). Machine learning- and threshold-based methods were trained/calibrated and tested using the same corresponding dataset. Reference data (1 605 plots) included first mowing event date of plots labeled as "mowed" in initial dataset (see section 3.2), and was split into a 70% training dataset and a 30% test dataset, ensuring sites representation through stratified sampling. Task- and approach-specific Sentinel-2-based time series were used as predictor (Table 2). Overall performances of tested methods were compared on the common test dataset. We also assessed models' ability to generalize using a spatial cross-validation approach known as "leave-one-site-out". Here, a site-specific observations were excluded from reference data before training and testing models. Therefore, models were evaluated separately on actual test dataset (obtained from n-1 sites) and on external test dataset, which was obtained from a specific and independent site not involved in learning process. This exercise was repeated nine times, so that each site was excluded once and considered as external test dataset. In next sections, we will provide a comprehensive overview of the steps involved in implementing and evaluating each method.

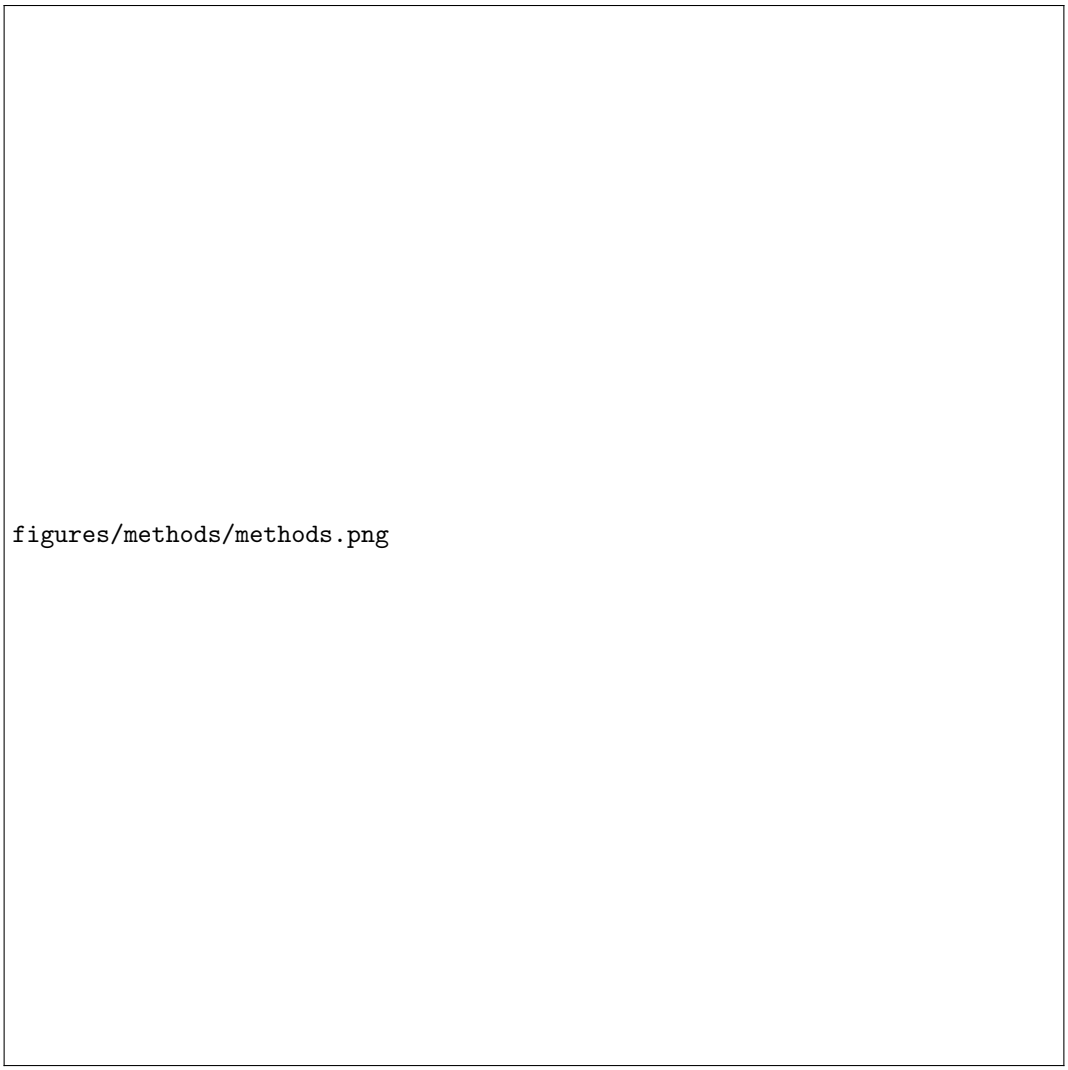

Figure 2: Workflow.

3.5.1 Machine learning approach

Machine learning algorithms are versatile for both classification and regression task. In recent years, despite their growing implementation in similar topics, few studies have applied machine learning-based methods to detect mowing event date, and traditional methods continue to dominate the field (Wang et al., 2022). We implemented seven generic models from the literature and adapted them to detect mowing event date, since most were primarily designed for classification and non-remote sensing-based applications (see section 7.1). These models were carefully selected based on their established capabilities in similar tasks or their adaptability to detect mowing event date. A comprehensive set of machine learning approaches, including traditional algorithms such as Random Forest, Ridge, and Least Absolute Shrinkage and Selection Operator (LASSO), alongside cutting-edge deep learning architectures like Fully Convolutional Network (FCN 1-D CNN), Lightweight Temporal Attention Encoder (LTAE), and Multilayer Perceptron (MLP), were rigorously evaluated in this study. A brief review of these models and their principal applications is given in section 7.1. As mentioned in preceding section, all models were trained and tested using the same corresponding dataset, as well as the same input dataset. Hyperparameters tuning was performed via grid search strategy, which provides optimal values from the space of possible values (Table 3). To detect mowing event date, Sentinel-2 spectral bands time series and their first derivative were used as input dataset (Table 2). First derivative quantifies temporal change in reflectance values. Therefore, this band-specific feature is relevant

for mowing events detection, since biomass extraction induces an abrupt shift in spectro-temporal profile of grasslands (Dusseux et al., 2014). Specifically, bands sensitive to vegetation dynamics exhibit a decrease in reflectance values.

3.5.2 Threshold-based approach

Threshold-based methods are well-known in vegetation dynamics studies, and were widely used in mowing event frequency and timing detection (Wang et al., 2022). We implemented a recent specific mowing event detection algorithm introduced by Vroey et al. (2022) as an integral monitoring tool within Sen4CAP program (http://esa-sen4cap.org). Here, it was adapted to detect mowing event date, since it was primarily designed to detect mowing event time interval, which provides current and preceding satellite acquisition dates. The main differences compared to original method are detailed in section 7.2. In this study, we evaluated Sentinel-2 NDVI time series to detect mowing event date (Table 2). NDVI is correlated to vegetation biomass dynamics throughout growing season (Dusseux et al., 2014). Therefore, the main idea is to quantify temporal loss of NDVI, and to consider a mowing event when this loss is higher than a threshold, which was specifically determined for our study area (Table 3). We employed two types of thresholds : the fixed threshold that arbitrarily determines a fixed value expressed as NDVI, and the relative threshold that determines a percentage of the NDVI seasonal amplitude. The method was calibrated and tested using the same dataset employed for training and testing machine learning models. To determine the optimal threshold, we tested a range of possible values and selected the value yielding highest performance (Table 3).

Approach	Model	Hyperparameters	Value range	Selected value
Machine learning	Random Forest	N trees, deep...	0-100	100
	Lasso	degree of regularization (λ)	0-100	100
	Rigde	degree of regularization (λ)	0-100	100
	MLP L1 Loss	epoch, ...	0-100	100
	MLP L2 Loss	epoch, ...	0-100	100
	FCN (1-D CNN)	epoch, ...	0-100	100
	LTAE L2 Loss	epoch, ...	0-100	100
Threshold-based	Fixed threshold	Minimum loss of NDVI	0.10, 0.11, ... 0.40	0.30
	Relative threshold	Minimum loss of NDVI	10%, 15%, ... 50%	15%

Table 3: Implemented methods and corresponding hyperparameter values.

3.6 Assessment of mowing events

We evaluated the performance of each method in detecting first mowing event date using the same test dataset. Assessment compared detected dates against observed dates at pixel level. For this exercise, we rasterized test dataset to match Sentinel-2 geometric grid, assigning the date observed in each plot to all intersecting pixels. These dates were expressed in Day Of Year (DOY). To measure deviation between detected dates and observed dates, we used statistical metrics commonly applied in regression task, such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and Bias, where discrepancies are measured in days. To assess spatial sensitivity of methods in detecting mowing event date, we calculated the percentage of observed dates explained by the detected dates using the coefficient of determination (r^2). These statistical metrics were calculated as follow :

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{1}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{2}$$

$$Bias = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i) \tag{3}$$

$$r^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{4}$$

Where $\hat{y}_i$ and y_i are detected date and observed date at pixel i, n is the number of pixels. In r^2 formula, $\bar{y}$ is the average of observed dates obtained from all pixels.

4 Results

5 Discussion

5.1 Mowing event date

Time interval graph wiht delta t observed Expliquer que l'on estime la fin de l'intervale observé Faire une figure explicative Interpreter les dates estimées avec le biais. Si negative, delta t est diminué et donc moins d'incertitude. Si positive, delta t est augmenté et donc plus d'incertitude.

Laisser claire que les dates estimées sont en effet la fin de l'intervale de temps observée.

5.2 False positive mowing events

Remaining clouds and snow :

Kolecka et al. (2018) found that the highest accuracy for detection of mowing events was achieved using additional clouds masking and size reduction of parcels, which allowed correct detection of 77% of mowing events. Additionally, we found that using only standard cloud masking leads to significant overestimation of mowing events (false positive).

5.3 Period of mowing event detection

Kolecka et al. (2018) : We found that more than 40% of the study area was mown before 15 June, while the remaining part was either mown later, or was not mown at all.

5.4 Number of valid observations

Kolecka et al. (2018) the detection based on sparse time series does not fully correspond to key events in the grass growth season.

5.5 Traditional and ML approach

Kolecka et al. (2018) Our approach : First, it is not conditional upon the availability of reference data, which is often missing, contrary to widely used machine learning strategies, which require training data. Understanding seasonal and phenological aspects of management practices in the study area was sufficient to allow discrimination between grass and non-grass clusters and, together with careful investigation of satellite imagery, allowed for development of a rule set for mowing event detection.

Komisarenko et al. (2022) MLP showed lower performances than their CNN

Lobert et al. (2021) NDVI time series alone mostly under performed in comparison to optical/SAR combinations but clearly outperformed input-sets that were solely based on SAR features.

Vroey et al. (2022) : In this study, the Planet image interpretation approach allowed to rapidly gather a large reference dataset (n = 803) to validate the mowing detections in six countries along the whole season (April to October 2019). ??

here, our approach had more reference dataset (n = 1600)

Nationwide mapping

Even though remote sensing-based approaches have been shown valuable to gather such information, large- scale mapping approaches are still scarce (Reinermann et al., 2020).

When lower cloud-free obs is available, mowing event frequency led to a systematic underestimation of mowing events, when the general data availability was not as high as in the following years (compare Fig. 2). This becomes increasingly prob lematic towards the South and may thus hamper comparable analyses in the context of CAP in other European countries. The launch of Landsat 9 and the planned launches of Sentinel-2C, and –2D in 2021/24/25 would enable to increase the density of optical time series — but only if all sensors remain active.

While the usefulness of SAR data for detecting grassland management has already been tested in several studies with a regional focus (De Vroey et al., 2021; Tamm et al., 2016, Lobert et al. accepted), mowing detection algorithms that make use of SAR and op tical data together are still scarce. Even though mowing events can be identified in SAR time series, additional factors such as topography, parcel size and shape

influence the results and there are still signal in teractions that need to be further explored (De Vroey et al., 2021).

In most common grasslands, exploitation activities (grazing or mowing) start from mid-April. In grasslands of high biological interest, supported by the EU CAP, mowing is only allowed after the 16th of June, for flowering purposes, and before the 31th of October.

Limitations : reference data with 7 days error We compared estimated date and observed date. However, observed date correspond to the end of observation time interval $\Delta t = 7days$.

6 Conclusion

References

Ahmad, I., Singh, A., Fahad, M., & Waqas, M. M. (2020). Remote sensing-based framework to predict and assess the interannual variability of maize yields in pakistan using landsat imagery. *Computers and Electronics in Agriculture*, *178*, 105732.

Barriere, V., & Claverie, M. (2022). Multimodal crop type classification fusing multi-spectral satellite time series with farmers crop rotations and local crop distribution. *arXiv preprint arXiv:2208.10838*.

Belgiu, M., & Drăguţ, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS journal of photogrammetry and remote sensing*, *114*, 24–31.

Breiman, L. (2001). Random forests. *Machine Learning*. https://doi.org/10.1023/a:1010933404324

Broyer, J., Curtet, L., & Boissenin, M. (2012). Does breeding success lead meadow passerines to select late mown fields? *Journal of Ornithology*, *153*, 817–823.

Cantelaube, P., & Carles, M. (2014). Le registre parcellaire graphique: Des données géographiques pour décrire la couverture du sol agricole. *Le Cahier des Techniques de l'INRA*, 58–64.

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. *Hydrology and Earth System Sciences*, *14*(9), 1731–1744.

Dusseux, P., Vertès, F., Corpetti, T., Corgne, S., & Hubert-Moy, L. (2014). Agricultural practices in grasslands detected by spatial remote sensing. *Environmental Monitoring and Assessment*. https://doi.org/10.1007/s10661-014-4001-5

Escadafal, R. (1989). Remote sensing of arid soil surface color with landsat thematic mapper. *Advances in space research*, *9*(1), 159–163.

Estel, S., Mader, S., Levers, C., Verburg, P. H., Baumann, M., & Kuemmerle, T. (2018). Combining satellite data and agricultural statistics to map grassland management intensity in europe. *Environmental Research Letters*. https://doi.org/10.1088/1748-9326/aacc7a

Garnot, V. S. F., & Landrieu, L. (2020). Lightweight temporal self-attention for classifying satellite images time series. In V. Lemaire, S. Malinowski, A. Bagnall, T. Guyet, R. Tavenard, & G. Ifrim (Eds.), *Advanced analytics and learning on temporal data* (pp. 171–181). Springer International Publishing.

Griffiths, P., Nendel, C., Pickert, J., & Hostert, P. (2020). Towards national-scale characterization of grassland use intensity from integrated sentinel-2 and landsat time series. *Remote Sensing of Environment*. https://doi.org/10.1016/j.rse.2019.03.017

Guidici, D., & Clark, M. L. (2017). One-dimensional convolutional neural network land-cover classification of multi-seasonal hyperspectral imagery in the san francisco bay area, california. *Remote Sensing*, *9*(6), 629.

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Applications to nonorthogonal problems. *Technometrics*, *12*(1), 69–82. Retrieved September 21, 2023, from http://www.jstor.org/stable/1267352

Hornik, K., Stinchcombe, M. B., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*. https://doi.org/10.1016/0893-6080(89)90020-8

Imani, M., & Ghassemian, H. (2015). Ridge regression-based feature extraction for hyperspectral data. *International Journal of Remote Sensing*, *36*(6), 1728–1742.

Ivanda, A., Šerić, L., Bugarić, M., & Braović, M. (2021). Mapping chlorophyll-a concentrations in the kaštela bay and brač channel using ridge regression and sentinel-2 satellite images. *Electronics*, *10*(23), 3004.

Kattenborn, T., Leitloff, J., Schiefer, F., & Hinz, S. (2021). Review on convolutional neural networks (cnn) in vegetation remote sensing. *ISPRS journal of photogrammetry and remote sensing*, *173*, 24–49.

Kiranyaz, S., Ince, T., Hamila, R., & Gabbouj, M. (2015). Convolutional neural networks for patient-specific ecg classification. *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2608–2611.

Kolecka, N., Ginzler, C., Pazúr, R., Price, B., & Verburg, P. H. (2018). Regional scale mapping of grassland mowing frequency with sentinel-2 time series. *Remote Sensing*. 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. *Scientific Reports*. https://doi.org/10.1038/s41598-022-04932-6

Kussul, N., Lavreniuk, M., Skakun, S., & Shelestov, A. (2017). Deep learning classification of land cover and crop types using remote sensing data. *IEEE Geoscience and Remote Sensing Letters*, *14*(5), 778–782.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278–2324.

Li, Z., Chen, G., & Zhang, T. (2019). Temporal attention networks for multitemporal multisensor crop classification. *Ieee Access*, *7*, 134677–134690.

Liao, C., Wang, J., Xie, Q., Baz, A. A., Huang, X., Shang, J., & He, Y. (2020). Synergistic use of multi-temporal radarsat-2 and venμs data for crop classification based on 1d convolutional neural network. *Remote Sensing*, *12*(5), 832.

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 Sensing of Environment*. https://doi.org/10.1016/j.rse.2021.112751

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Lonjou, V., Desjardins, C., Hagolle, O., Petrucci, B., Tremas, T., Dejus, M., Makarau, A., & Auer, S. (2016). Maccs-atcor joint algorithm (maja). https://doi.org/10.1117/12.2240935

McFeeters, S. K. (1996). The use of the normalized difference water index (ndwi) in the delineation of open water features. *International Journal of Remote Sensing*. https://doi.org/10.1080/01431169608948714

Ofori-Ampofo, S., Pelletier, C., & Lang, S. (2021). Crop type mapping from optical and radar time series using attention-based deep learning. *Remote Sensing*, *13*(22), 4668.

Peel, M. C., Finlayson, B. L., & McMahon, T. A. (2007). Updated world map of the köppen-geiger climate classification [Publisher: Copernicus GmbH]. *Hydrology and Earth System Sciences*, *11*(5), 1633–1644. https://doi.org/10.5194/hess-11-1633-2007

Pelletier, C., Webb, G. I., & Petitjean, F. (2019). Temporal convolutional neural network for the classification of satellite image time series. *Remote Sensing*, *11*(5), 523.

Petermann, J. S., & Buzhdygan, O. Y. (2021). Grassland biodiversity. *Current Biology*, *31*(19), R1195–R1201.

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 Sensing*, *14*(7). https://doi.org/10.3390/rs14071647

Rouse, J. W., Haas, R. H., Schell, J. A., Deering, D. W., et al. (1974). Monitoring vegetation systems in the great plains with erts. *NASA Spec. Publ*, *351*(1), 309.

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 Sensing of Environment*. https://doi.org/10.1016/j.rse.2021.112795

Stumpf, F., Schneider, M. K., Keller, A., Mayr, A., Rentschler, T., Meuli, R., Schaepman, M. E., & Liebisch, F. (2020). Spatial monitoring of grassland management using multi-temporal satellite imagery. *Ecological Indicators*. https://doi.org/10.1016/j.ecolind.2020.106201

Suttie, J. M., Reynolds, S. G., & Batello, C. (2005). *Grasslands of the world* (Vol. 34). Food & Agriculture Org.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *58*(1), 267–288.

Verrelst, J., Camps-Valls, G., Muñoz-Marı, J., Rivera, J. P., Veroustraete, F., Clevers, J. G., & Moreno, J. (2015). Optical remote sensing and the retrieval of terrestrial vegetation bio-geophysical properties–a review. *ISPRS Journal of Photogrammetry and Remote Sensing*, *108*, 273–290.

Vinayak, B., Lee, H. S., & Gedem, S. (2021). Prediction of land use and land cover changes in mumbai city, india, using remote sensing data and a multilayer perceptron neural network-based markov chain model. *Sustainability*, *13*(2), 471.

Vroey, M. D., Vendictis, L. D., 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 Sensing of Environment*. https://doi.org/10.1016/j.rse.2022.113145

Wang, Z., Ma, Y., Zhang, Y., & Shang, J. (2022). Review of remote sensing applications in grassland monitoring. *Remote Sensing*. https://doi.org/10.3390/rs14122903

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 Sensing of Environment*. https://doi.org/10.1016/j.rse.2023.113577

White, R. P., Murray, S., Rohweder, M., Prince, S., Thompson, K., et al. (2000). *Grassland ecosystems*. World Resources Institute Washington, DC, USA.

Zandler, H., Brenning, A., & Samimi, C. (2015). Quantifying dwarf shrub biomass in an arid environment: Comparing empirical methods in a high dimensional setting. *Remote Sensing of Environment*, *158*, 140–155.

Zhang, C., Pan, X., Li, H., Gardiner, A., Sargent, I., Hare, J., & Atkinson, P. M. (2018). A hybrid mlp-cnn classifier for very fine resolution remotely sensed image classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, *140*, 133–144.

Zhang, C., Sargent, I., Pan, X., Li, H., Gardiner, A., Hare, J., & Atkinson, P. M. (2019). Joint deep learning for land cover and land use classification. *Remote sensing of environment*, *221*, 173–187.

Zhao, Y., Liu, Z., & Wu, J. (2020). Grassland ecosystem services: A systematic review of research advances and future directions. *Landscape Ecology*, *35*, 793–814.

Zhong, L., Hu, L., & Zhou, H. (2019). Deep learning based multi-temporal crop classification. *Remote sensing of environment*, *221*, 430–443.

7 Supplementary information

7.1 Review of models

Breiman (2001) introduced Random Forest algorithm for classification and regression tasks. It employs ensemble learning by training multiple decision trees independently, using randomly selected subsets of training samples and features. It has been widely used in remote sensing time series applications, mainly for land cover/use mapping and estimation of continuous variables (Belgiu & Drăguţ, 2016). Ridge is a linear regression model that represents a special case of Least Squares, implementing a penalty criteria on predictor coefficients, which aims to improve model generalization and modulate overfitting (Hoerl & Kennard, 1970). This method is not common in remote sensing applications but some examples can be mentioned, such as chlorophyll-a concentration mapping (Ivanda et al., 2021) or reduction of hyperspectral data dimension (Imani & Ghassemian, 2015). Tibshirani (1996) developed Least Absolute Shrinkage and Selection Operator (LASSO), a regression model with regularization criteria that constrain the predictor coefficients toward zero. This restriction forces the sum of the absolute value of the regression coefficients to be less than λ. Predictors with a regression coefficient of zero after shrinkage are excluded from the model. There are limited examples in the literature of using this model with remote sensing data. However, a few instances include estimating vegetation parameters (Verrelst et al., 2015), biomass (Zandler et al., 2015) or crop yields (Ahmad et al., 2020). Based on Convolutional Neural Network (CNN) architecture introduced by LeCun et al. (1998) for 2-D image categorization task, Fully Convolutional Network (FCN) is a deep learning architecture, which has been developed for 2-D image semantic segmentation task (Long et al., 2015), performing pixel-based labeling. These 2-D architectures exploit the spatial domain of the data. In signal processing, Kiranyaz et al. (2015) introduced an 1-D CNN architecture for patient-specific electrocardiogram classification, allowing to exploit the sequential order of the data. More recently, 1-D CNN architecture was adapted and implemented to exploit spectro-temporal domain of satellite data (Kattenborn et al., 2021), mainly for land cover/use classification tasks (Guidici & Clark, 2017; Kussul et al., 2017; Liao et al., 2020; Pelletier et al., 2019; Zhong et al., 2019). FCN (1-D CNN) refers to an architecture using temporal domain at pixel level. Garnot and Landrieu (2020) introduced Lightweight Temporal Attention Encoder (LTAE), a deep learning-based architecture using attention mechanism for remote sensing time series classification task. Similar architectures were also developed and implemented mainly in crop type mapping (Barriere & Claverie, 2022; Li et al., 2019; Ofori-Ampofo et al., 2021). Finally, Multilayer Perceptron was formalized by Hornik et al. (1989) and is the seminal architecture in deep learning topic. This versatile algorithm performs well in both classification and regression tasks, using backpropagation for training the network. It was implemented in remote sensing-based data analysis, such as in land cover/use mapping (Kussul et al., 2017; Zhang et al., 2018; Zhang et al., 2019) or land cover/use changes analysis (Vinayak et al., 2021).

7.2 Adapted threshold-based method

We implemented a recent specific mowing event detection algorithm introduced by Vroey et al. (2022) and integrated into the Sen4CAP toolbox (http://esa-sen4cap.org). In our study, this method was adapted to detect mowing event date, since it was primarily designed to detect mowing event time interval, which provides current and preceding satellite acquisition dates. In this section, we will provide a concise overview of the original method and emphasize the main distinctions between it and the adapted method.

Original method's primary purpose is to facilitate the monitoring of grassland management activities across Europe, aligning with the European Common Agricultural Policy. Vroey et al. (2022) employed two independent change detection algorithms, whereby raw Sentinel-2 NDVI and Sentinel-1 VH-coherence time series were evaluated separately. In the final product, Sentinel-1 outputs were considered only when Sentinel-2 omitted events due to cloud cover. Here, we reproduced and adapted their Sentinel-2-based algorithm for evaluating pixel-based time series, as opposed to the original method that used object-based approaches. To account for a mowing event, the original algorithm performed the following steps : (i) each observation NDVI(t) is compared to the last available cloud-free observation NDVI(t-1), (ii) an mowing event is considered when the loss of NDVI, between NDVI(t) and NDVI(t-1), is greater than 0.15 NDVI ($NDVI(t) < NDVI(t-1) - 0.15$), (iii) as an additional condition, two consecutive mowing events must be separated by a minimum temporal distance of 28 days, and (iv) if a mowing event is detected within the time

interval [t-1, t], it is assumed that the actual event took place within 60 days before t. If [t-1, t] spans more than 60 days, the detection interval is adjusted to [t-60, t]. For each detected mowing event, the confidence level was estimated through a normalization function as follows:

$$f(x; min, max) = max - (max - min) \times e^{-x} \tag{5}$$

Where x is the difference NDVI(t-1) - 0.15 - NDVI(t), [min, max] were set to fit the confidence limits from 0.5 to 1. We retained the first mowing event among the four most confident detections, as opposed to the original method that retained all four most confident detections. In contrast to the original method, where the time interval [t-1, t] was kept for each detected mowing event, we retained the specific date t. Therefore, in our study, step (iv) was ignored. Finally, original method was evaluated using performance metrics related to classification task (precision, recall...), where an interception of detected and observed time interval was considered as a correct detection. Here, however, we used performance metrics related to regression task (Mean Absolute Error, Root Mean Square Error...).