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

Henry Rivas, Hélène Touchais, Vincent Thierion, Jerome Millet, Laurence Curtet, Mathieu Fauvel

Remote Sensing of Environment 315 (2024) 114476

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

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

Henry Rivas [a], Hélène Touchais [a], Vincent Thierion [a], Jerome Millet [b], Laurence Curtet [c], Mathieu Fauvel [a,*]

[a] Centre d'Etudes Spatiales de la Biosphère (CESBIO), Université de Toulouse, CNES/CNRS/INRAE/IRD/UT3-Paul Sabatier, 31401, Toulouse, France
[b] Office Français de la Biodiversité (OFB), Direction de la Recherche et de l'Appui Scientifique, 79360, Villiers-en-Bois, France
[c] Office Français de la Biodiversité (OFB), Direction de la Recherche et de l'Appui Scientifique, 01330, Montfort, Birieux, France

ARTICLE INFO

Edited by Marie Weiss

Keywords:
Regression
Deep-learning models
Mowing dates mapping
Grassland management intensification
Satellite image time series

ABSTRACT

Grassland dynamics are modulated by management intensity and impact overall ecosystem functioning. In mowed grasslands, the first mowing date is a key indicator of management intensification. The aim of this work was to assess several supervised regression models for mapping grassland first mowing date at national-level using Sentinel-2 time series. Three deep-learning architectures, two conventional machine learning models and two threshold-based methods (fixed and relative) were compared. Algorithms were trained/calibrated and tested from field observations, using a spatial cross-validation approach. Our findings showed that time aware deep-learning models – Lightweight Temporal Attention Encoder (LTAE) and 1D Convolutional Neural Network (1D-CNN) – yielded higher performances compared to Multilayer Perceptron, Random Forest and Ridge Regression models. Threshold-based methods under-performed compared to all other models. Best model (LTAE) mean absolute error was within six days with a coefficient of determination of 0.52. Additionally, errors were accentuated at extreme (late/early) mowing dates, which were underrepresented in the data set. Oversampling techniques did not improve predicting extreme mowing dates. Finally, the best prediction accuracy was obtained when the number of clear dates surrounding the mowing event was greater than 2. Our outputs evidenced time aware deep-learning models' potential for large-scale grassland first mowing event monitoring. A national-level map was produced to support bird-life monitoring or public policies for biodiversity and agro-ecological transition in France.

1. Introduction

Grasslands cover approximately 40% of the Earth's land area, encompassing nearly 70% of the global agricultural land area (Suttie et al., 2005; White et al., 2000). Grasslands are subject to management practices, such as mowing or grazing or a combination of both, influencing overall ecosystem functioning (Zhao et al., 2020; Bengtsson et al., 2019), biodiversity (Klein et al., 2020), carbon sequestration (Yang et al., 2019), water quality (Huang et al., 2013), and more (Zhao et al., 2020). The choice of a specific practice impacts biodiversity (Wang et al., 2019; Metera et al., 2010). For instance, in mowed grasslands, the first mowing event date is a key indicator of the intensification level of plot's management, i.e., increase of nitrogen intake (Broyer et al., 2012). Consequently, this timing information is critical for assessing above-mentioned topics (Buri et al., 2016; Klein et al., 2020; Broyer et al., 2012). Usually, this monitoring is supported by local field observation campaigns which are time-consuming and difficult to repeat regularly. As a consequence, they are not spatially/temporally exhaustive and they need to be complemented with other data acquisition processes.

Remote sensing data enable recurrent and global-scale monitoring, facilitating tracking of vegetation dynamics at high spatial resolution and frequent revisit. For instance, Sentinel-2 mission provides cost-free high resolution data (10 m as spatial resolution and 5 days revisit), allowing intra-plot level observations. Such satellite data have proven their interest to monitor vegetation at large scale (Kooistra et al., 2024).

Grassland mowing events and intensity estimation have been investigated using satellite image time series, mainly through features sensitive to vegetation status, such as Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI), Leaf Area Index (LAI) and more (Reinermann et al., 2020). These methods usually exploit the temporal information contained in satellite image time series to detect mowing events: a significant variation is usually associated to an

* Corresponding author.
E-mail address: mathieu.fauvel@inrae.fr (M. Fauvel).

https://doi.org/10.1016/j.rse.2024.114476
Received 22 April 2024; Received in revised form 25 September 2024; Accepted 14 October 2024
Available online 30 October 2024

event. Methods differ in how such drop is computed (Estel et al., 2018; Kolecka et al., 2018; Griffiths et al., 2020; Stumpf et al., 2020; Watzig et al., 2023). Such methods exploit only optical modality and can be limited by clouds cover. A strategy to reduce cloud-related gap in optical time series is to combine optical data from different sensors. For instance, Schwieder et al. (2022) combined Sentinel-2 and Landsat −8 EVI time series for mowing events detection in Germany. To cope with optical sensors limitation, Vroey et al. (2022) developed an algorithm for detecting mowing events across Europe using jointly raw Sentinel-2 NDVI and Sentinel-1 VH-coherence time series. Reinermann et al. (2022) mapped mowing frequency across Germany, from Sentinel-2 EVI and Sentinel-1 PolSAR entropy time series separately. Alternatively, Garioud et al. (2019) jointly used Sentinel-1 and Sentinel-2 satellite image time series as well as climatic and topographic data to reconstruct continuous Sentinel-2 NDVI time series for mowing date estimation, based on NDVI drop analysis.

While threshold-based methods have been widely investigated, they are limited in their used because of manual tuning of the threshold parameter, which can be difficult when working on large areas. This issue is exacerbated when multi-modal time series are used. Supervised learning approaches have been investigated in the last decade to circumvent this issue. 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. Among the fourteen features used, NDVI and the harmonic mean of VV and VH coherence were found to be the most relevant. Lobert et al. (2021) also used a similar deep-learning model (1-D CNN) on Sentinel-2/Landsat −8 and Sentinel-1-based features time series for mowing event frequency and timing detection. Among all tested feature combinations, the highest overall accuracy in terms of F1-score was reached when combined NDVI, backscatter cross-ratio and coherence. Following a similar approach, Holtgrave et al. (2023) tested four machine learning algorithms for mowing event detection in Germany. Sentinel-2/Landsat −8, Sentinel-1 and weather-based features time series were analyzed. Mowing events were detected by a binary classification approach (mown or unmown) for each observation in the time series, using the adjacent observations as predictors. 1D-CNN and Long Short-Term Memory algorithms provided the best results in terms of F1-score classification accuracy.

Recent research have clearly showed the superiority of supervised machine learning-based approaches in mowing detection from remote sensing data. Such approaches were mainly formulated as a binary classification problem of temporal sequences[1]: *was there a mowing event during a given temporal sequence or not?* (Lobert et al., 2021; Holtgrave et al., 2023). Yet, the number of temporal sequences to be processed, their length and their possible overlap impact significantly the final accuracy. It thus poses challenges in optimization at large-scale because they cannot be optimized end-to-end and trials and errors must be conducted. Furthermore, there is no consensus in the literature on the optimal satellite data for mowing event detection in grasslands with deep learning: (Lobert et al., 2021) showed benefit using jointly optical and radar data with deep learning algorithms while authors of Holtgrave et al. (2023) recently observed that best mowing detection accuracy were obtained using optical data alone. They concluded that optical data alone can effectively detect mowing events, provided that enough cloud-free observations are available.

This paper targets the estimation of the first mowing event date using optical time series and machine learning algorithm formulated as a regression problem. The choice to use only optical satellite image time series was motivated by the significant additional computing cost to perform jointly radar and optical data at large-scale in regards to the limited reported improvement in the literature. Furthermore, regression algorithms were selected in contrast to classification ones in order to reduce the number of non-learnable parameters (such as those related to the temporal sequences) and to ease the learning step at large-scale. A large field survey (more than 2 000 plots) was conducted on agro-ecological and climatic different and distant landscapes, to construct one of the largest data set of mowing events for mainland France. We investigated several algorithms for predicting first mowing date: two threshold-based, two conventional machine learning-based and three deep-learning-based from which two process the temporal information specifically. We also compared the accuracy of the prediction when the problem is cast to a classification problem: each class being a non overlapping temporal interval. The accuracy of each method was assessed using spatial cross-validation for several metrics. We provided an analysis of the influence of the clouds cover on the prediction accuracy for the best regression model. A national-level map was produced using the learned model and qualitative and quantitative accuracy were discussed.

2. Materials

2.1. Study area

Our study area covers permanent grasslands across the mainland France (except Corsica), which represent 68.5% of the total grassland area – including permanent, temporary and other grasslands –, declared in the Land Parcel Identification System — LPIS (Cantelaube and Carles, 2014) in 2022 (Fig. 1). LPIS provides spatialized information on agricultural plot boundaries and crop types, but does not provide information about management practices. According to climate normals, annual rainfall is around 800–1000 (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 °C in summer and 5–10 °C in winter (https://meteofrance.com/climat/normales/france).

In LPIS, permanent grasslands are defined as surfaces with uninterrupted herbaceous cover for at least 6 years and are identified at the plot level with class code 18. These permanent grasslands alone account for approximately 27.5% (76 835 km^2) of the entire agricultural area reported in 2022. Permanent 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 mainland France, permanent 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 (Fig. 1). According to the LPIS, at least 75% of permanent grassland plots cover 2.80 hectares or less, and the largest plots – exceeding 20.0 hectares – are concentrated mainly in the center and eastern regions. In mainland France, grassland growing season spans from spring to autumn (March to October) and mowed grasslands are mainly managed extensively, with one or two mowing events per year (up to six mowing events in intensive management).

From bird-life diversity view point, the first mowing event date could be more important than mowing events frequency along growing season (Broyer et al., 2012). In our study area, intensification of grassland management is characterized by the first mowing event occurring around beginning of June (around 20 of May on lowland and around 20 of June on highland), whereas extensive management is characterized by the first mowing event occurring 20 to 30 days later (later to 15 of June on lowland and later to 15 of July on highland).

2.2. Satellite data

All available Sentinel-2 (L2 A) surface reflectance images, captured throughout the growing season (from the beginning of January to the end of September 2022) and intercepting mainland France, were used. Seven tiles intercepting field observation sites were used for training

[1] A temporal sequence is defined a set of continuous temporal acquisitions, e.g., all Sentinel-2 acquisitions between t_{start} and t_{end}

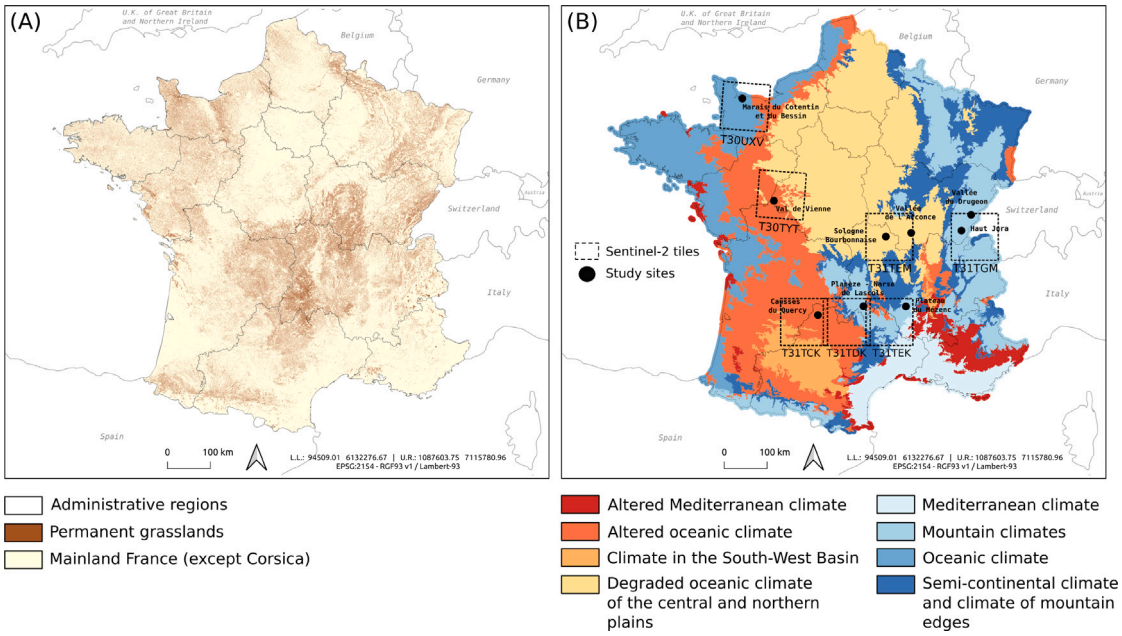

Fig. 1. (A) Study area location. The gray color represents the delimitation of mainland France (except Corsica), while the green color represents the permanent grassland plots declared in the LPIS 2022. (B) Study sites location. The black dots represent the observation sites in 2022. The boxes in dashed gray lines represent the Sentinel-2 tiles that intercept each observation site. The color palette represents the eco-climatic regions in mainland France, as defined in Joly et al. (2010). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Satellite data used for each model: input is a single pixel time series. $b_b(t)$ indicates the Sentinel-2 spectral band number b at time t, and $b'_b(t)$ is its derivative w.r.t. time. *Raw* for temporal acquisition means the Sentinel data are the original ones and *interpolated* means the Sentinel-2 acquisitions are interpolated on a 10-day regular grid. t_1 and t_T are the first and last dates respectively. For both *raw* and *interpolated* t_1 is equal to 2022-01-01 while t_T is equal to 2022-09-30 and 2022-09-28 for *raw* and *interpolated*, respectively.

Methods	Machine Learning	Threshold-Based
Spectral Bands	$b_2, b_3, b_4, b_5, b_6, b_7, b_8, b_{8A}, b_{11}, b_{12}$	None
Spectral features	None	NDVI
Temporal acquisition	Interpolated	Raw or interpolated.
Temporal features	First derivative	None
Features vectors $\mathbf{x}$	$b_2(t_1), \dots, b_2(t_T), b_3(t_1), \dots, b_{12}(t_T), b'_2(t_1), \dots, b'_2(t_T), b'_3(t_1), \dots, b'_{12}(t_T)$	$\text{ndvi}(t_1), \dots, \text{ndvi}(t_T)$
Size (nb of feature)	560	199

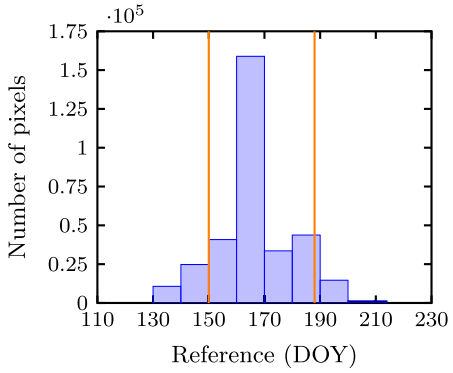

Fig. 2. Distribution of pixel-level first mowing event dates observed in *mowed* plots across all sites in 2022. The orange vertical lines indicate the 10th and 90th percentiles of the data, respectively.

and testing models (Fig. 1). An average of sixty images were available for each tile for the considered temporal period. All spectral bands (except B1, B9 and B10) were used, after resampling 20 m resolution bands to 10 m resolution[2] to uniform pixel sizes on a common geographical grid.

These images have 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). All images were provided with a mask layer for clouds and shadows. For each tile, cloud- and shadow-free time series with a regular 10-day time interval were generated using a linear interpolator, as done in Inglada et al. (2017) or Bellet et al. (2023a). In addition to spectral bands, we also computed their temporal derivative – using finite differences –, as well as the Normalized Difference Vegetation Index — NDVI (Rouse et al., 1974) (Table 1).

[2] using a bi-cubic interpolation, as implemented in the Orfeo ToolBox and its SuperImpose application (Team, 2023)

2.3. Reference data

In 2022, the French Biodiversity Agency (https://www.ofb.gouv.fr) coordinated an intensive campaign of field observations throughout the mainland French territory, involving local government agencies participating in the National Observatory of Mowed Grasslands Ecosystem network. A total of eight sites (from north to south: Marais du Cotentin et du Bessin, Val de Vienne, Sologne Bourbonnaise, Vallée de l'Arconce, Vallée du Drugeon, Haut Jura, Plateau du Mézenc and Planèze - Narse de Lascols) were monitored. They come from four different eco-climatic regions (Fig. 1) and they have a significant altitudinal gradient (Table 2). Observations were conducted once a week from May to August, for a total of 2 227 plots, with 1 605 mowed plots and a balanced distribution among sites (Table 2). For each specific site, observed plots were chosen based on accessibility and the local observer's prior knowledge of the area. Additionally, Causses du Quercy site was included (in the south, Fig. 1), where 38 plots were observed with a lower temporal resolution. Here, observations were provided by the local observatory of the *Parc Naturel Régional des*

Table 2
Statistical description of the observed sites. The values represent the number of mowed plots (# plots), average plot area (Av. area), and approximate altitude. Tile column represents Sentinel-2 tile intercepting an observed site.

Site	Tile	# mowed plots	Av. area (Ha)	Altitude (m)
Marais du Cotentin et du Bessin	T30UXV	136	1.39	2–50
Val de Vienne	T30TYT	239	1.06	30
Sologne Bourbonnaise	T31TEM	119	2.47	230–280
Vallée de l'Arconce	T31TEM	174	2.23	280–390
Vallée du Drugeon	T31TGM	219	3.87	800–850
Haut Jura	T31TGM	213	2.55	800–950
Plateau du Mézenc	T31TEK	255	1.66	1100–1300
Planèze - Narse de Lascols	T31TDK	217	1.40	1000–1050
Causses du Quercy	T31TCK	33	0.50	309–775
Total		1605		

Causses du Quercy, independently of the main observation campaign at the above-mentioned sites.

Plot boundaries were obtained from the 2020 LPIS. For permanent grasslands, a declared plot can be managed with two practices simultaneously (i.e., spatially separated mowing and grazing within the same plot). Therefore, prior to the field observation campaign, we visually assessed each chosen plot using a national database of aerial imagery (BD ORTHO, https://geoservices.ign.fr/bdortho) and Google Earth, to identify, as far as possible, sub-plots with the most homogeneous spatial structure. For each actual plot, a total of eleven observations were conducted throughout the growing season. At each weekly visit, current management practice (mowing or grazing) and the corresponding date were recorded.

Based on these records, each plot was labeled as *mowed*, *grazed* or *mixed* (mowing + grazing). Then, these labeled plots were grouped into two management practice categories: *mowed* – including *mowed* and *mixed* plots (70.5% of plots) – and *unmowed* – including *grazed* plots –. A management practice could be ongoing during the visit or have occurred between the current visit and the previous visit. Consequently, in these *mowed* plots, the observed date for a mowing event may have an uncertainty of few days. Here, 87% of *mowed* plots had one mowing event. The remaining plots had two mowing events and only the first event was used in the experiments.

A unique site identifier was assigned to plots located within the same Sentinel-2 tile in order to separate training and testing samples based on the tile membership. In the following, the term *site* is used to denote plots belonging to the same Sentinel-2 tile.

The prediction of first mowing event date is done at pixel-level. Thus, all pixels in observed *mowed* plots were selected, and their spectro-temporal profiles were extracted along with their corresponding first mowing event dates, serving as predictor and target values, respectively. Mowing event dates span from May 10th, day of the year (DOY) 130, to August 2nd, DOY 214, comprising a total of 328 451 pixels derived from the 1 605 observed *mowed* plots (Fig. 2). 80% of the mowing dates fell between DOY 150 and 188. The remaining occurrences (20%) fell at the extremes of the distribution. The average observed date was June 16th (DOY 167), while the median was June 15th (DOY 166). The total number of pixels is 328 451.

3. Methods

3.1. Mowing events prediction

Several pixel-wise supervised regression models were investigated, from conventional machine learning to recent deep-learning ones, and unsupervised threshold-based methods. Prediction performances against ground observations were then compared. Following Fauvel et al. (2020), we set up a spatial cross-validation to estimate the prediction accuracy (Fig. 3). All observations from a site were excluded from reference data before training/calibrating models. Then, models were tested and assessed using the excluded site-specific observations. In other words, all observations from excluded site were used as testing data, and all observations from non-excluded sites were used as training data. This split was repeated seven times, so that each site was excluded once and considered as testing data (i.e., 7-fold spatial cross-validation). We computed average prediction accuracy using all sites scores (spatial folds), with each site's score being estimated as the average of individual evaluations on fifty bootstrap test set.

A comprehensive overview of algorithms used to predict the first mowing event date is given in the following.

3.1.1. Machine learning approach

We implemented five supervised regression models from the literature: two conventional machine learning models – Random Forest (RF) and Ridge Regression – and three widely investigated deep-learning architectures — Multilayer Perceptron (MLP), 1D Convolutional Neural Network (1D-CNN, Pelletier et al., 2019) and Lightweight Temporal Attention Encoder (LTAE, Garnot and Landrieu, 2020).

Ridge Regression and MLP were used as a baseline, while RF was chosen due to its good behavior in large-scale prediction (Inglada et al., 2017; Fauvel et al., 2020). 1D-CNN and LTAE were selected for their capacity to model temporal information, leveraging convolutional techniques and attention mechanisms, respectively. A brief review of these models is given in Appendix A.

The hyperparameters for each model are presented in Table 3. Deep-learning models were trained on 200 epochs with a batch size of 4 096, using the Adam optimization algorithm (Kingma and Ba, 2015). In learning process, 10% of training data were used to form a validation set, which was used to perform early stopping and to reduce the learning rate by a factor of 10 when learning stagnated.[3] For each model, features and target variables were standardized (i.e. zero mean and unit variance).

For all supervised models, oversampling techniques for minority ranges of mowing dates were also investigated. More specifically, two oversampling techniques for classification problems were tested: SMOTE (Chawla et al., 2002) and ADASYN (He et al., 2008) algorithms, relying on a convex combination of existing samples. Since these oversampling techniques were defined for classification problem, we created 10 *fake* classes by dividing the interval of mowing date values into 10 sub-intervals of equal width and assigned each pixel to a class corresponding to the number of the interval in which its label fell.[4] For each oversampling technique, we used the implementation provided by the `imbalanced-learn` library (Lemaître et al., 2017). All classes were oversampled, except for the majority class, to obtain an equal number of samples in each class.

Lastly, in order to quantify the learning capacity of the above machine learning algorithms, we also reported results for naive regressor,[5] that makes prediction using a simple rule: predict the mean value of the training data. It can be seen as a linear model with only an intercept parameter. Theoretically, such model should have a R^2 equal to zero. It is called SimpleMean in the following.

[3] *Reduce on plateau* strategy implemented here https://pytorch.org/docs/stable/generated/torch.optim.lr_scheduler.ReduceLROnPlateau.html

[4] Other number of bins were investigated, as well as different binning strategy, providing similar or worst results. For clarity we only reported results for 10 equal bins.

[5] https://scikit-learn.org/stable/modules/generated/sklearn.dummy.DummyRegressor.html

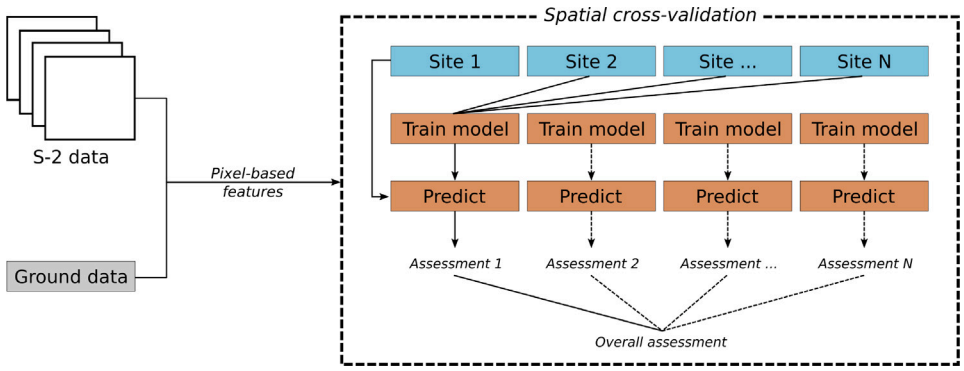

Fig. 3. Workflow for predicting first mowing event date. The box in dashed lines represents the spatial cross-validation approach implemented in this study. Here, the different steps (model training, prediction and assessment) implemented for a given site (spatial fold) are illustrated. The solid lines represent implementation for site 1, while dashed lines represent implementation for all remaining sites. In this example, all observations from site 1 were used as testing data, while all observations from the remaining sites were used as training data.

Table 3
Algorithm hyperparameters. The **Value** column reports the selected value or the search range for the algorithm, with the following notation `start:step:end`. For Ridge Regression and threshold methods, cross-validation was used to select the best value.

Algorithm	Hyperparameters	Value	Package
Random Forest	Number of trees	100	Scikit-Learn
Ridge Regression	Regularization	1000:500:15500	Scikit-Learn
1D-CNN	Batch size Optimizer Learning rate	4096 Adam 1e−3	Pytorch
LTAE	Batch size Optimizer Learning rate	4096 Adam 1e−3	Pytorch
MLP	Batch size Optimizer Learning rate	4096 Adam 1e−4	Pytorch
Fixed threshold	Minimum loss of NDVI	0.10:0.01:0.40	
Relative threshold	Minimum loss of NDVI	10:5:50%	

3.1.2. Threshold-based approach

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). The main idea developed in Vroey et al. (2022) was to quantify temporal loss of NDVI, and to consider a mowing event when this loss is higher than a threshold value. In our study, the threshold value was set automatically using grid-search on training data (Table 3). Two types of thresholds methods were used: 1. A fixed threshold corresponding to absolute loss of NDVI. Threshold value is expressed in NDVI. 2. A relative threshold corresponding to relative loss of NDVI, taking into account the pixel amplitude (minimum and maximum value). Threshold value is expressed as a percentage of the amplitude.

It was adapted to detect only the first mowing event date, since it was primarily designed to detect mowing event time interval. The main differences compared to original algorithm are detailed in Appendix B.

Threshold-based algorithms were calibrated and tested using the same training and testing data used for machine learning-based algorithms, respectively.

3.2. Assessment of mowing events

The deviation between predicted and observed first mowing dates was assessed using nine metrics (Pontius, Jr., 2022): Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Max error, the coefficient of determination (R^2), the Mean Deviance (MD), the coefficient of Correlation (Corr), the Slope(S) and the 50-percentile and the 95-percentile of the Absolute Error (Q_{50} and Q_{95}).[6] The last two metrics are robust (to outliers in the reference) version of MAE and Max Error. These metrics are estimated as:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}(|y_i - \hat{y}_i|),$$

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

$$\text{Max error} = \max_{i\in\{1,\dots,n\}}(|y_i - \hat{y}_i|),$$

$$\mathrm{R}^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2},$$

$$\mathrm{MD} = \bar{y} - \bar{\hat{y}},$$

$$\mathrm{Corr} = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\left[\sum_{i=1}^{n}(y_i - \bar{y})^2\right]\left[\sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})^2\right]}},$$

$$\mathrm{S} = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sum_{i=1}^{n}(y_i - \bar{y})^2},$$

$$Q_{50} = \text{median}\{|y_1 - \hat{y}_1|, \dots, |y_n - \hat{y}_n|\},$$

$$Q_{95} = \text{percentile}_{95}\{|y_1 - \hat{y}_1|, \dots, |y_n - \hat{y}_n|\},$$

where $\hat{y}_i$ and y_i are predicted and observed first mowing dates at pixel i, respectively, n is the number of pixels in testing data. $\bar{y}$ and $\bar{\hat{y}}$ are respectively the average of observed dates and the average of predicted dates, in testing data. MAE, RMSE and Max_error. The metrics were computed on the original mowing dates scale, i.e. after un-standardization of the model predictions.

The MAE, RMSE, Max error, MD, Q_{50} and Q_{95} are represented in days and should be minimized, except MD which should be canceled. R^2, Corr and S should be maximized, with R^2 and Corr ranging between −1 and 1, and S$\in \mathbb{R}$.

[6] The Q_{50} is also known as the Median Absolute Error

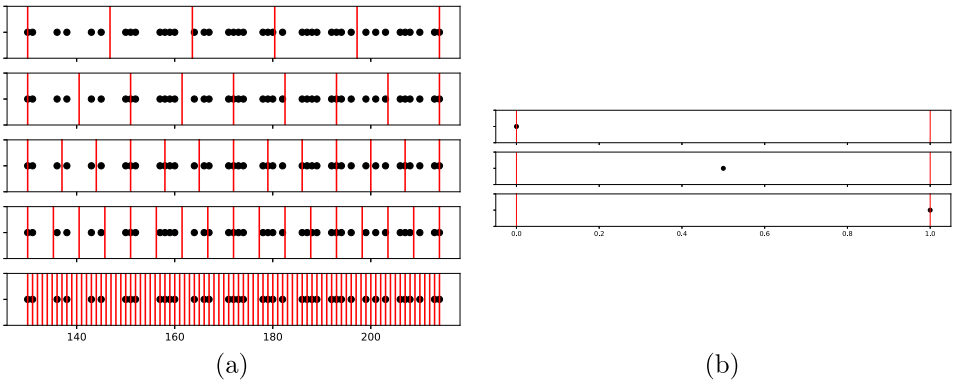

Fig. 4. (a) Illustration of the construction of classes from mowing event dates. From top to bottom, different temporal sequences of size 15, 10, 7, 5 and 1 days respectively. Black dots are the mowing dates and the vertical red line are the limit of the temporal sequences. Each samples belongs to one and only one temporal sequence/class. The smaller the size of the sequence, the higher the number of classes. For interval of size 1 day, many sequences are empty, i.e., no mowing date fall into such sequences. (b) Illustration of the position in the interval of the earliest first mowing event. From top to bottom: beginning, middle and end of the sequence. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.3. Classification set-up

The problem of estimating the date of first mowing event was cast into a classification problem. Inspiring by Lobert et al. (2021), Holtgrave et al. (2023), the mowing data have been sliced into non-overlapping temporal sequences of constant length. The Fig. 4 illustrates this procedures for different interval sizes. Different size for classes were investigated: 15, 10, 7, 5 and 1 days. The proposed set-up is somewhat different from Lobert et al. (2021), Holtgrave et al. (2023) where the pixel time series was sliced into overlapped temporal sequences which was then classified using a binary classifier discriminate between mown/unmown sequence. In our classification set-up, the full time series was used.

The effect on the accuracy of the starting date for the first class was also considered: the earliest first mowing date of our data set was either positioning at the beginning, middle or end of the first temporal sequence/class. To each sequence is assign one class and the classification problem is to classify the pixel reflectance time series to one class/sequence. The same spatial cross-validation strategy, using exactly the same pixels for training and testing, was used than for regression.

4. Results

In this section, a comprehensive overview of the quantitative results derived from all implemented algorithms is presented. We further provide a more detailed analysis of the results obtained for the best model. Finally, a qualitative assessment of the prediction map is conducted.

4.1. Evaluation of algorithms for mowing events prediction

Fig. 5 shows best results obtained for each model with and without oversampling. Results for all models and configurations can be found in Appendix D, Table D.6. Results were averaged over the 7-fold spatial cross-validation runs, as detailed in 3.1.

From Fig. 5, non-linear machine learning models obtained the best performances, by a clear margin, with a R^2 above 0.4, a correlation above 0.65 and a Q_{50} less than 5 days. Linear models reached lower accuracy, Ridge Regression model being slightly better than SimpleMean, as expected. Worst results, for any quality index, were obtained for threshold-based methods. Among non-linear models, deep-learning ones (LTAE, 1D-CNN and MLP) yielded higher performances compared to conventional RF. Furthermore architectures taking into account the temporal dimension (LTAE and 1D-CNN) were the most accurate models, with a slight advantage for LTAE. Results in terms of the others metrics followed the same trend than for R^2.

Oversampling techniques improved accuracy only for MLP; other models did not show notable accuracy gains. Additionally, there was no significant difference in accuracy improvement between SMOTE and ADASYN. Oversampling did not improve predictions for extreme mowing dates (early/late), as shown in the next section.

In terms of maximum error, among the best algorithms, 1D-CNN demonstrated the lowest values, followed by MLP and LTAE. Using robust estimator Q_{95}, the differences are smaller. If LTAE offered the best predictions on average, it occasionally resulted in higher errors for certain metrics.

Fig. D.13 in Appendix D presents the prediction accuracy at the site level (spatial fold). Within-site variability can be observed, and LTAE did not offer the highest accuracy for each site. Yet, on average, LTAE provided the best results for MAE, RMSE or R^2. In the next section, we provide a more detailed analysis of LTAE predictions.

4.2. Mowing events prediction across sites for LTAE

Fig. 6 shows the joint density between predicted and observed dates for each site, along with the marginal density for both predicted and observed dates, as well as for the corresponding training data. On average, predictions were accurate since the modes of the joint density were on the identity line. Yet, for smaller R^2 (T30UXV, T30TYT and T31TEM), we observed a clear overestimation of early dates (predicted dates were later than observed ones), as well as an underestimation of late dates (predicted dates were earlier than observed ones).

Fig. 7 shows MAE and first mowing event distribution as a function of the number of cloud-free dates in a temporal interval surrounding the event.[7] The worst average MAE was obtained for very low number of cloud-free dates (0 to 2) and early mowing event (before DOY 150). Once the number of cloud-free dates reached 3, MAE remained low, with a median value below 7 days. A slight degradation of MAE was observed when the number of cloud-free dates is greater than 8, corresponding to late mowing events.

Fig. 8 shows pixel-wise average of LTAE attention score. The attention score reflects which part of the temporal signal was used for the prediction, and it is computed at pixel-level. Implemented LTAE used 4 heads, i.e., 4 ranges of the temporal signal can be selected. For the first attention head, selected observations – from interpolated satellite image time series – were located in the month of July, whatever the predicted date. We observed more variation for the three

[7] 0 cloud-free dates means that the 10-day linearly interpolated satellite image time series used for prediction was constructed using Sentinel-2 acquisitions outside the temporal interval.

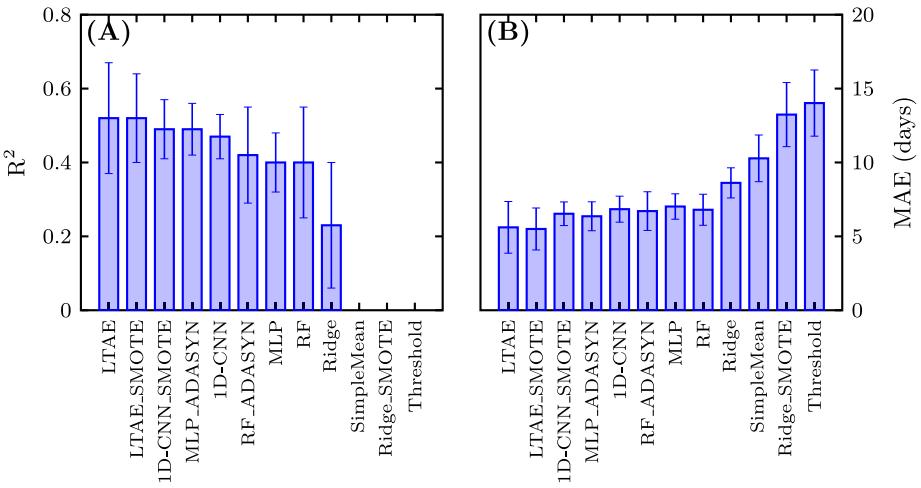

Fig. 5. Algorithm-specific statistical summary in terms of (A) R^2 and (B) MAE. The values represent weighted means of all sites. A site-specific score was weighted using the number of pixels used for the evaluation. A site-specific score represents the mean of fifty individual evaluations (from 50 folds by bootstrapping 70% of observations). Here, values less than zero are not shown.

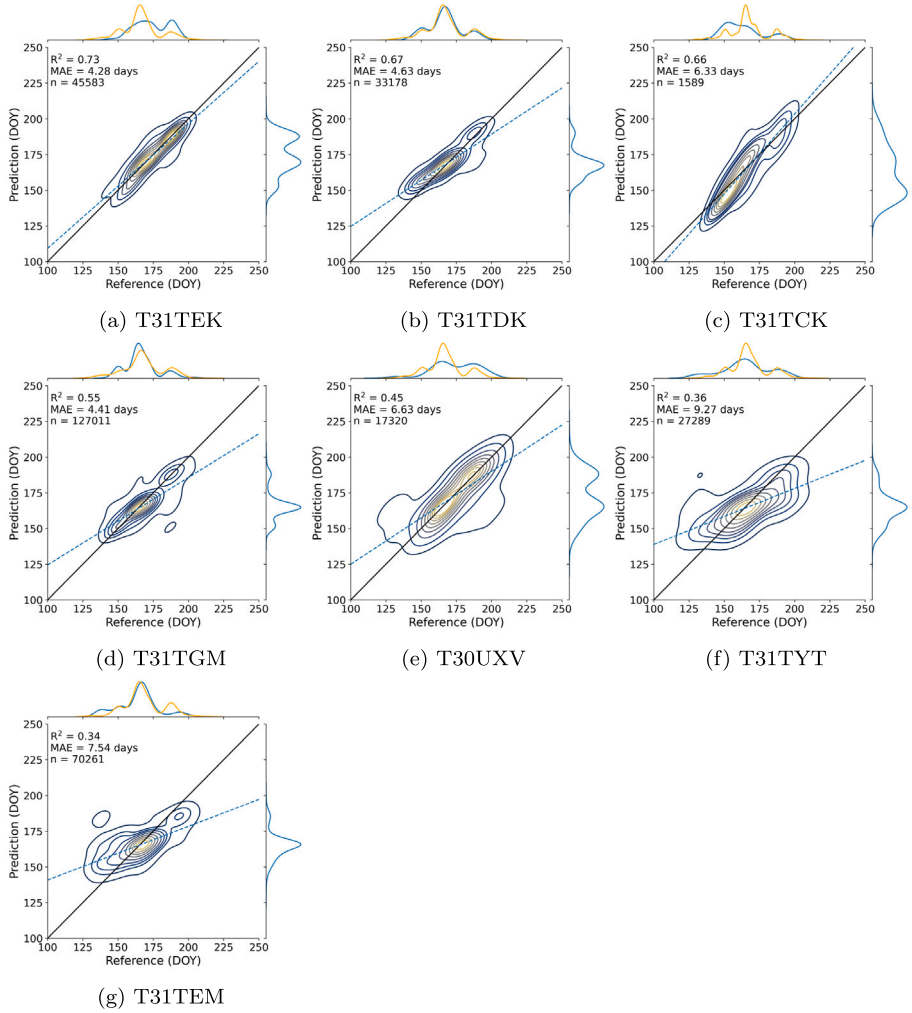

Fig. 6. Site-specific LTAE pixel-level predictions. For each graph, horizontal axis represents the reference datasets (testing in blue and training in orange), while the vertical axis represents the prediction. The black continuous line is the identity line and the dashed blue line is the regression line between the reference and prediction. Each graph corresponds to a site (spatial fold). For instance, in (A), T31TEK testing data included all samples from that specific site, while training data included all samples from all remaining sites, as discussed in Section 3.1. Predicted and observed dates are expressed in days of year (DOY). Joint and marginal densities were computed using Scipy `gaussian_kde` function (Virtanen et al., 2020).

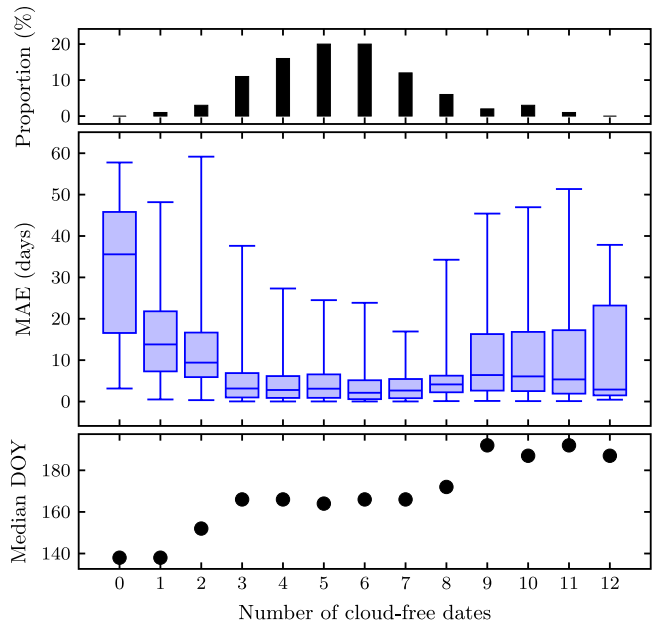

Fig. 7. MAE distribution as a function of the number of cloud-free dates in a temporal interval surrounding the event (in blue). The temporal interval was defined as $[y-20$ days, $y+20$ days$]$ with y the observed mowing date, and MAE is defined in Section 3.2. The number of cloud-free dates can reach up to 12 within a 40-day period due to Sentinel-2 orbit overlap. Proportion of each category (number of cloud-free dates), expressed as a percentage of the total number of samples, is shown in black in the top figure. The bottom figure shows the median DOY for the first mowing event for each category.

Table 4
Classification results. ACC and F1 are respectively the accuracy score and F1 score computed on the test set. The "w" version are the same metrics computed using the class proportion. Finally, the MAE was computed by considering the middle of the interval/class as the predicted date (e.g., for a class corresponding to dates between DOY 150 and 159, the predicted date is 155).

	Origin	# Class.	ACC	ACCw	F1	F1w	MAE
RF	L	5	**0.77**	**0.65**	**0.77**	**0.45**	20.63
	R		0.76	0.58	0.76	0.43	19.77
	C		0.65	0.61	0.65	0.41	21.29
LTAE	L		0.75	0.45	0.75	0.42	20.73
	R		0.75	0.46	0.75	0.41	19.82
	C		0.62	0.43	0.62	0.35	21.31
RF	L	8	0.60	0.44	0.60	0.31	12.42
	R		0.59	0.48	0.59	0.33	12.15
	C		0.64	0.42	0.64	0.29	11.67
LTAE	L		0.56	0.33	0.56	0.27	12.84
	R		0.57	0.35	0.57	0.29	12.56
	C		0.62	0.34	0.62	0.26	12.12
RF	L	12	0.53	0.40	0.53	0.22	6.28
	R		0.56	0.29	0.56	0.20	6.32
	C		0.52	0.34	0.52	0.19	**5.60**
LTAE	L		0.51	0.23	0.51	0.18	7.26
	R		0.53	0.24	0.53	0.17	7.31
	C		0.51	0.27	0.51	0.20	6.37
RF	L	16	0.46	0.28	0.46	0.16	5.96
	R		0.45	0.28	0.45	0.16	6.73
	C		0.44	0.26	0.44	0.13	5.87
LTAE	L		0.44	0.19	0.44	0.13	7.17
	R		0.44	0.19	0.44	0.13	7.74
	C		0.45	0.21	0.45	0.12	6.82
RF	C	84	0.26	0.16	0.26	0.07	5.70
LTAE	C		0.30	0.09	0.30	0.04	6.69

other heads. For the second attention head, selected observations were distributed between April and June (included). For the third attention head, selected observations were located between February and June, for predicted early dates; while between June and September for predicted late dates. For the last attention head, selected observations were distributed between April and June. Overall, no clear pattern was observed linking attention scores and predicted dates.

4.3. Performance comparison with classification algorithm

For this comparison, Random Forest and LTAE algorithms were used only: LTAE provides the best results for the regression case and RF is known to perform well in large scale classification. Results are reported in Table 4. It can be seen that the size of the interval, and at a lesser degree, the origin of the intervals have an influence on the final accuracy. The higher the number of intervals/classes, the higher accuracy but it also comes with higher incertitude w.r.t. the dates. For 8 classes, the intervals have a size of 10 days. For extreme case, when the size of the intervals is one day, the classification accuracy is low, but the estimated MAE is in line with those obtained with regression method.

4.4. Spatialized prediction analysis for LTAE

In this section, the prediction map was analyzed qualitatively with a particular focus on the spatial distribution of intra-plot predictions. For each spatial cross-validation fold (testing data), an individual prediction map was generated for each observed plot. All plot-level prediction maps are fully available for visualization in the supplementary material, which contains prediction for LTAE, 1D-CNN and RF. After a careful visual analysis of LTAE predictions, four main cases were identified and illustrated in Fig. 9.

1. In most of the plots, intra-plot predictions were homogeneous showing a coherence in terms of spatial configuration and were in agreement with the reference, as illustrated in the first row (A.1 to A.3) of Fig. 9.
2. For some plots, predictions exhibited two distinct spatial patterns within a given plot. Usually, one spatial pattern matched with the reference and the other one did not. This case is illustrated in the second row (B.1 to B.3) of Fig. 9.
3. Occasionally, heterogeneous intra-plot predictions were obtained with a global disagreement with the reference. This situation is shown in the third row (C.1 to C.3) of Fig. 9.
4. Finally, homogeneous intra-plot predictions with large deviation from the reference can be found. For most of such plots, it corresponded to early/late mowing dates, as it is displayed in the last row (D.1 to D.3) of Fig. 9.

The two last cases occurred most often at site T31TEM, where prediction accuracy was the lowest (R^2=0.34 and MAE=7.54 days).

4.5. Mowing events prediction across mainland France

The LTAE model was selected to generate a prediction map across mainland France, because of its performances compared to others models previously evaluated in this study. All referenced samples (328 451 pixels across seven sites) were used to train model. Then the learned model was used to predict the first grassland mowing date for all *mowed grassland* pixels. Such pixels corresponds to pixels classified as *mowed* in a grassland management map, as described in Appendix C. The prediction map is freely accessible at https://zenodo.org/records/11034387. An overview of the map is given in Fig. 10.

It is difficult to assess the quality of the map at large-scale. Accuracy metrics and spatial homogeneity have been reported in the previous section for the 7 tiles. As a sanity check, mowing dates for 45 plots

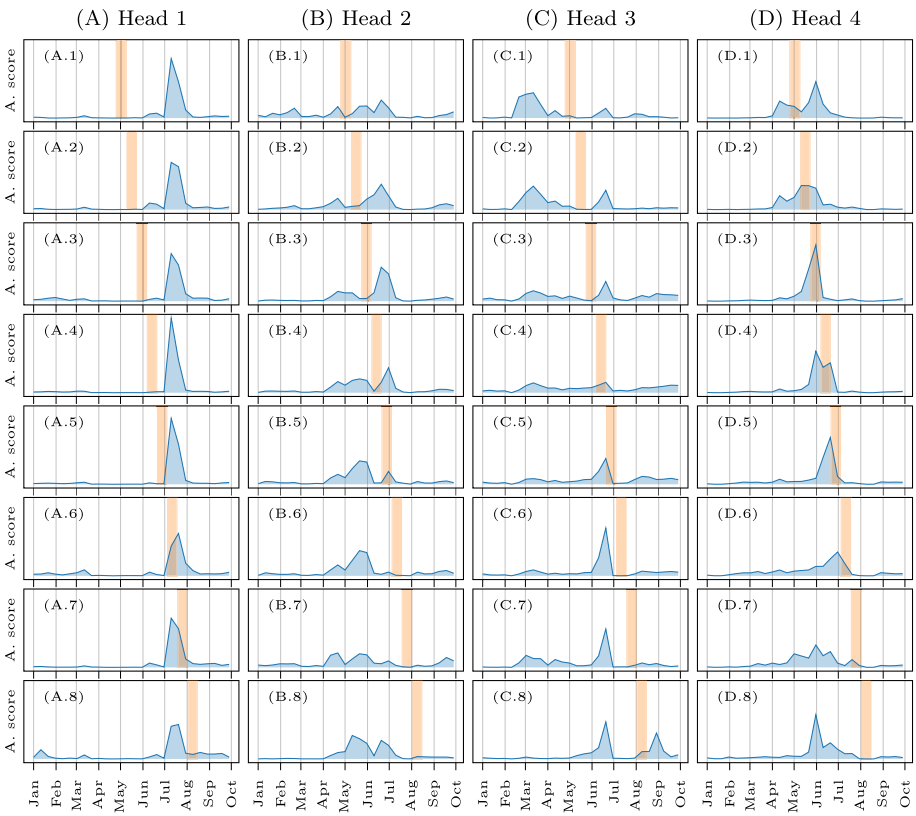

Fig. 8. LTAE attention mechanism across 4 heads. For each observation in the interpolated satellite image time series (x-axis), attention score (y-axis) was computed at pixel-level; and the average attention score from all pixels – in testing data – is depicted in blue. Satellite image time series observations with high attention score were the most significant for the prediction process. Each column (A–D) represents a specific attention head, while each row (1–8) represents a specific range of predicted dates (in orange). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

(corresponding to 27 478 pixels) have been extracted from an experimental research platform[8] located in the tile T31TDL not seen during the training process, for the same year. These are two experimental sites (Laqueuille and Marcenat) and are actively monitored by local research units. Observed dates were the true mowing date (without uncertainty).

Prediction metrics for these unseen parcels are in agreement with those of the previous section: $R^2 = 0.67$, MAE=5.17, RMSE=7.72, MD=1.29, $r = 0.83$ and slope=0.63. The scatter plot of the prediction is given in Fig. 11. In particular, the tile T31TDL is spatially closed to the tile T31TDK and the accuracy metrics were similar.

5. Discussion

5.1. Machine learning algorithms

Deep-learning models shown superior results in terms of prediction accuracy. Contrary to RF, they have the ability to automatically learn data representations at multiple levels of abstraction through the use of multiple processing layers (LeCun et al., 2015), which could enable them to capture patterns within the data more effectively compared to conventional machine learning models. In mowing event detection, Holtgrave et al. (2023) have also reported this superior capability of deep-learning models. In our study, architectures that exploited temporal dimension (LTAE and 1D-CNN) demonstrated higher performance compared to a baseline deep-learning model (MLP). This temporal information's contribution to improving mowing events prediction was also observed in Komisarenko et al. (2022), Lobert et al. (2021). This could be due to their ability to capture temporal patterns and dependencies (e.g., biomass variation) within the time series, as already noted in other applications (Lin et al., 2020; Pelletier et al., 2019; Liao et al., 2020; Zhong et al., 2019). However in our study, temporal derivatives of each spectral band were provided as additional feature to each algorithm, thus LTAE and 1D-CNN may capture other temporal dependency than first temporal derivative.

Attention scores of LTAE, show in Fig. 8, tend to indicate that acquisition between April and September (included) are the most important for the prediction, whatever the mowing date to be predicted. From the first head, it seems that the month of July contains very predictive information, since it is used for every range of prediction. This result indicates that, at the national scale, the most useful spectral information to predict mowing event is located during July: the state of the vegetation during that period is enough (for the LTAE algorithm) to predict the mowing date, being post or past this period. To a lesser extent, May and June data are also very important for the prediction. If early dates are considered, one attention head use acquisition from mid-February to end of March. However, this observation should be mitigated with the limited accuracy associated to this range of dates.

Threshold-based method exhibited the worst performance in detecting first mowing event date compared to all supervised models. The lack of cloud-free observations during the event is an inherent limitation of optical remote sensing data (it can be clearly visualized in the supplementary material). Although this affects both thresholding and machine learning methods, the latter have the ability to learn from the available data (with or without cloud-free observations during the event) which could overcome this limitation to some extent. More critical is that threshold-based method assumption is not always satisfied, in particular in natural grasslands and late event. In that situation, the senescence is already effective and dry grass is visible, resulting in no or

[8] INRAE-Herbipole experimental farm https://doi.org/10.15454/1.5572318050509348E12

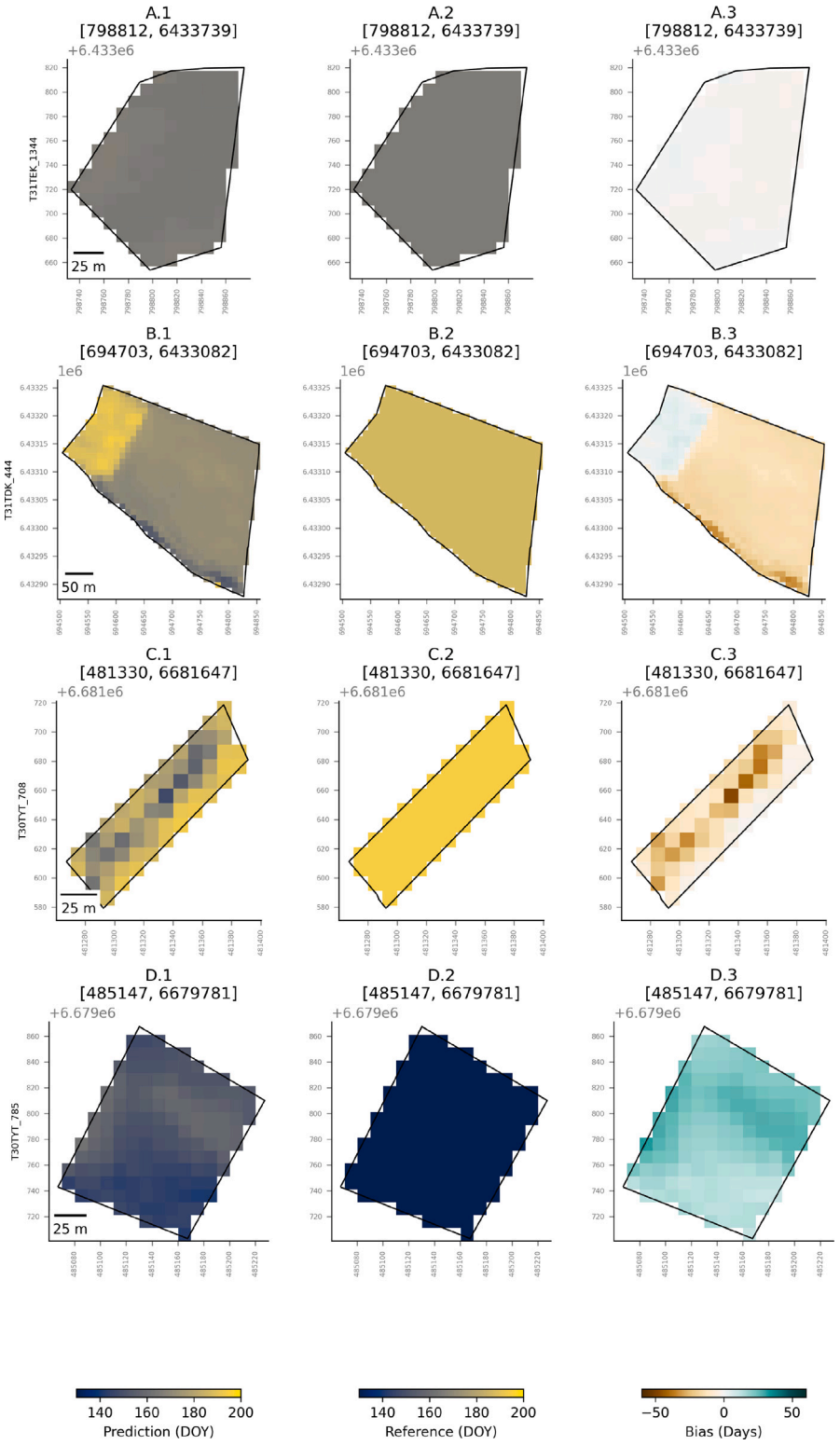

Fig. 9. Intra-plot LTAE outputs. The columns (1–3) represent prediction, reference and bias (prediction – reference) values, respectively. The predicted and observed dates are expressed in Days Of Year (DOY), and bias is expressed in days. A negative bias indicates the predicted date was earlier than the observed date, while a positive bias means the predicted date was later than the observed date. The rows (A–D) represent four selected plots (T31TEK_1344, T31TDK_444, T30TYT_708 and T30TYT_785).

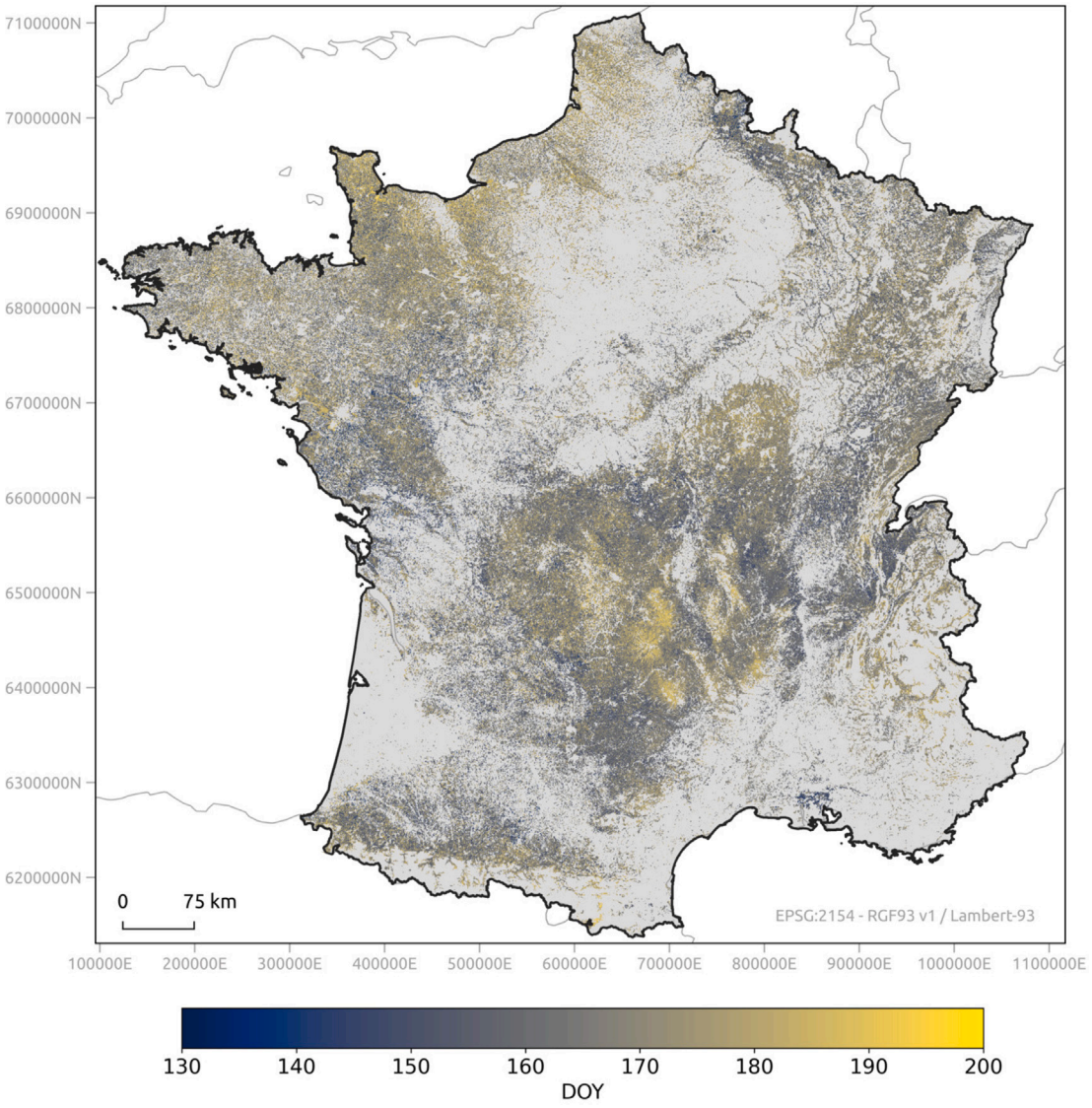

Fig. 10. Map of grassland first mowing date in mainland France in 2022. This map was generated by combining LTAE model and Sentinel-2 time series.

reduce loss of NDVI value after the mowing. The Fig. 12 below shows three plots corresponding to such situation.

5.2. Classification or regression?

For some configurations, the classification accuracy matches with those obtained with regression algorithms. One objective of this work was to demonstrate that regression set-up is easier to implement than classification to predict the date of the first mowing event at large scale. In our opinion, training classification models for predicting the date of the first mowing event has two major limitations:

1. Some parameters related to the classification problem cannot be learned from the data: number, size and origin of the intervals. To choose them, a full exhaustive search involving a training and testing steps must be run. In comparison, in regression, internal parameters of RF, LTAE or 1D-CNN can be learned from the data. One can remark that these algorithms have also pre-fixed internal parameters, such as the number of trees for RF or the number of head for LTAE. However, in practice it exists for them a large range of values for which the results are similar and therefore they are easier to tune (For instance changing the number of trees from 100 to 400 does not change the final regression accuracy).
2. During the learning step, classification error does not take into account the "*size*" of the error: roughly speaking, the prediction is true if the model assigns the pixel to the right class, and it is wrong if it is not the right class. It is does not take into account if the predicted class is adjacent or not to the right class. When using a regression loss, for instance the mean square error, it gives more importance in the learning process to large errors, i.e., predicted dates far from the true ones.

5.3. Accuracy of the mowing events prediction

Errors were more important when predicting early and late mowing dates. A clear trend in overestimating early mowing dates as well as underestimating late mowing dates was observed for each model. Owing to the reference data distribution (see Fig. 2) early/late mowing dates were underrepresented and therefore more difficult to learn for the supervised models. Yet rare, maximum errors were obtained for such extreme dates.

Oversampling techniques were investigated in this study to alleviate such imbalanced issues in the training data (Jafarigol and Trafalis,

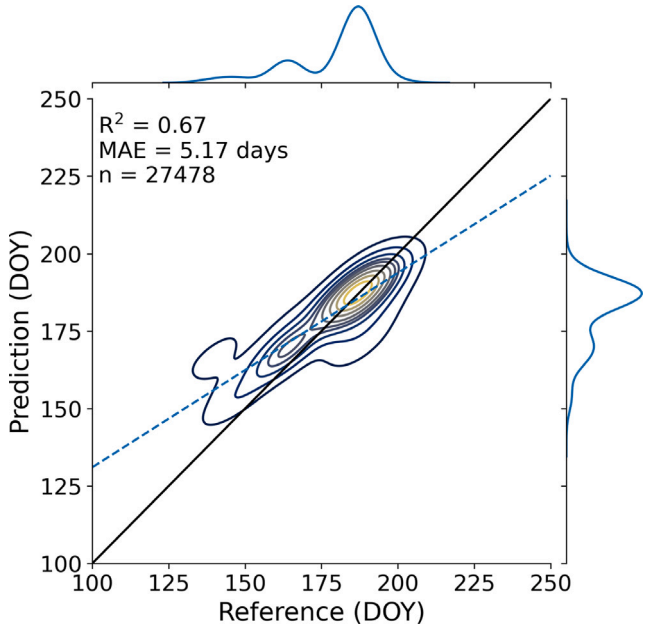

Fig. 11. Comparison between predicted (*y*-axis) and observed (*x*-axis) dates from the two independent sites. Observations from both sites were concatenated and are not shown separately in this graph. Predicted and observed dates are expressed in Days Of Year (DOY). The number of pixels used for the evaluation is represented by *n*. The black continuous line is the identity line and the dashed blue line is the regression line between the reference and prediction. Joint and marginal densities were computed using Scipy `gaussian_kde` function (Virtanen et al., 2020). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

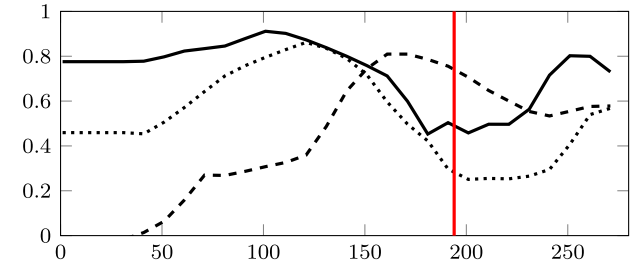

Fig. 12. The continuous, dotted and dashed black lines are NDVI as a function of time for three permanent grassland pixels with late first mowing event. The horizontal axis represents the day of the year and the vertical axis represents the NDVI. Black lines are the 10-day interval interpolated data. The vertical red line corresponds to the mowing event at DOY 194.

Table 5
Confusion matrix for intensive–medium–extensive post-classification. Intensive: prior to June 05; medium: between June 05 and 30; extensive: post to June 30. Rows and columns correspond to the assignment obtained with the field survey and the model prediction respectively. The matrix is row-normalized.

		Predicted		
		Intensive	Medium	Extensive
True	Intensive	0.79	0.20	0.01
	Medium	0.07	0.90	0.03
	Extensive	0.04	0.23	0.73

2023). Such approaches have found to be useful in classification problems (Mohammed et al., 2020; Rahman et al., 2023) and also for remote sensing data (Hao et al., 2023) to construct well balanced training data. However, these techniques did not help to reduce this issue in our large-scale regression scenario. Only MLP model exhibited an improvement when oversampling data, but its accuracy remained lower than those obtained with time aware deep-learning models.

For the bird-life diversity management viewpoint, the correct identification of early or late is determined by event occurring before beginning or after end of June, respectively (see 2.1). In order to assess the usability of the predicted mowing event date, we computed the confusion matrix of assignation to intensive (prior to June 05), medium or extensive management (post to June 30) w.r.t. the field survey using the predicted first mowing event.[9] The confusion matrix is given in Table 5. Even if early mowing event are less accurately recovered, the results are quite consistent and can be used to monitor intensification of practices.

5.4. Influence of the number of cloud-free dates on prediction accuracy

The number and distribution of cloud-free satellite observations during the growing season are critical for reliably detecting patterns of change, mainly during the event of interest. Cloud cover modulates optical satellite data availability, resulting in dense or sparse time series according to location. Temporal gap in optical time series affects both the threshold method and the supervised models performances.

Overall, the highest prediction errors for LTAE were observed when the number of cloud-free satellite observations was limited around the true mowing date (i.e., 0-2 cloud-free dates), which occurs mainly for early mowing event. Then they decreased as the number of cloud-free dates increased (i.e., 3-8 cloud-free dates), as illustrated in Fig. 7 and in agreement with Schwieder et al. (2022), Kolecka et al. (2018), Komisarenko et al. (2022). Surprisingly, prediction errors increased slightly on average when the number of cloud-free dates increased (i.e., 9-12 cloud-free dates). However, it mainly happens for late mowing event, which are less accurately estimated (see bottom Fig. 7). This pattern can be explained by the seasonality over the French metropolitan territory: there are more clouds in April/May than in July/August: hence early and late mowing dates will have respectively less and more cloud-free acquisitions. We therefore consider that the increased error for reduce (less than 3) or higher (more than 8) number of dates are mainly due to model limitation for early/late mowing dates. Yet, the higher errors for early event indicates that when no acquisition is available in the given period, the model is significantly negatively impacted.

Usually, main strategies implemented to cope with sparse time series due to clouds are (i) combining optical data from different sensors and (ii) combining optical and radar data. From an operational view point, including Sentinel-1 satellite image time series in the processing comes with a high computational burden, since Sentinel-1 and Sentinel-2 images need to be projected on the same pixels grid and Sentinel-1 satellite image time series required additional noise filtering. In mainland France, very cloudy pixels are rare (less than 5% of the pixels) and it is questionable that the effort is worth it. For a bird-life monitoring perspective, we believe that works should be done to reduce errors for early/late event either in improving the learning algorithm or in collecting specific data set.

5.5. Importance of the field survey

In our study, field observation campaign conducted on 2 265 plots (Section 2.3) aimed to represent the greatest possible diversity of mowing dates as well as site conditions (i.e., altitudinal gradient, flooded and dry grassland, early and late mowing dates, etc.), which required a great human and time effort. However, despite this effort, the representativeness of the reference data needs to be improved, mainly for minority ranges of mowing dates. In this context, efforts to

[9] The dates of June 05 and 30 are questionable at the scale of the French territory, but how to define intensive/extensive management is beyond the scope of the paper. However, by predicting the DOY it is possible to adapt these values depending on the landscape.

build a more diversified reference dataset could be oriented towards a citizen/collaborative science, involving citizens and established observation networks, such as the emerging initiatives of the French Biodiversity Agency,[10] or those already consolidated in similar topics for phenological observations (e.g., French Phenology Network-TEMPO,[11] National Phenology Network-USA/NPN,[12] UK Phenology Network,[13] Pan-European Phenological Network-PEP725,[14] etc.).

As shown in Fig. 9-B.(1-3), some grasslands have bimodal mowing management. Such situation was not considered neither in the learning nor in the validation processes. This inaccurate/incomplete reference map is mainly due to inaccessibility to the plot or non-visibility of the intra-plot bimodal mowing management during the observer's visit. The LTAE model was able to reliably capture the intra-plot mowing management, despite it was learned with few label noise (i.e. wrong mowing date). Yet rare, this situation was not taken into account in the validation process.

Regarding accuracy of observed dates, revisit frequency depends on the observation campaign protocol and the observer's availability. Here, observed dates included an average uncertainty of seven days, due to weekly revisits. Consequently, this uncertainty in the reference data is involved in the learning process, which could partially degrades the performance of the model. A more reliable approach would be to use the true mowing date rather than observed mowing date. Such field data is inaccessible by field survey unless farmers voluntarily declare and share it, which could be raised under public policies such as the CAP (https://agriculture.ec.europa.eu) in Europe for example. Hence, in our study, supervised models were trained to predict the observed date which is not necessarily the actual date of the event, but a close approximation of it. We believe this a cost to pay for covering significant heterogeneous grassland practices. It should be note that the quality of the prediction was good, and of the same magnitude, for the data set with the true mowing date, see Section 4.5. It indicates that the field survey uncertainty is not critical.

6. Conclusion

This paper focused on nationwide mapping of grassland first mowing event date, combining machine learning and Sentinel-2 time series. Among implemented algorithms, time aware architectures (LTAE and 1D-CNN) were the most accurate models. Overall, uncertainties were accentuated at extreme mowing dates (early/late), which were underrepresented in the reference data. Oversampling techniques demonstrated no significant improvement in predicting these extreme mowing dates, except for MLP. Regarding transferability, LTAE model exhibited reliable performance across all spatial folds, but accuracy may degrade for area underrepresented in training process. In our study, LTAE model's best prediction occurred when the number of cloud-free dates was greater than 2 within a 40-day temporal interval surrounding the mowing event (which occurred in more than 95% of tested samples). In comparison to classification based approaches, regression based set-up show slightly superior results and were easier to implement, with less hand-tuned parameters.

End-to-end learning such as in Bellet et al. (2023b) has shown significant improvement for classification purpose. Rather than interpolating the data before and independently of the learning process, the reconstruction is learned jointly with the classification task. Such approach should be considered in the future, either for a mono-sensor or multi-sensor, as it has outperformed common "reconstruct" then "learn" strategy discussed in previous studies. We expect it will leverage the slight performance loss observed when the number of cloud-free dates was high. Another strategy that could be considered would be based on Transformer, such as in Zhang et al. (2024) were irregular and unaligned SITS were properly handled by adapting Transformer architecture to automatically aligned temporal acquisition.

Our findings evidenced time aware deep-learning models' potential to nationwide grassland mowing monitoring. Although our approach should be adapted for predicting all mowing events during growing season, predicted first mowing event date is a key indicator of plot management intensification; and could support bird-life monitoring or public policies for biodiversity and agro-ecological transition in France. For long-term monitoring, LTAE transferability into an unknown year needs to be investigated. Self-supervised learning is a promising research direction for such application (Dumeur et al., 2024).

CRediT authorship contribution statement

Henry Rivas: Writing – original draft, Visualization, Validation, Investigation, Formal analysis. **Hélène Touchais:** Writing – review & editing, Visualization, Software, Data curation. **Vincent Thierion:** Writing – review & editing, Supervision, Investigation, Conceptualization. **Jerome Millet:** Writing – review & editing, Project administration, Funding acquisition. **Laurence Curtet:** Writing – review & editing, Project administration, Funding acquisition, Conceptualization. **Mathieu Fauvel:** Writing – review & editing, Project administration, Methodology, Funding acquisition, Conceptualization.

Acknowledgments

The authors would like to thank OFB's fieldworker (departmental services 03, 14, 15, 25, 37, 39, 43, 50, 71) and Permanent Center for Environmental Initiatives of Val de Vienne for the field campaign, the INRAe-Herbipôle Team for sharing their data set, the Theia Center for making available the Sentinel-2 satellite image time series, the CNES for funding H. Touchais and H. Rivas work and for providing the HPC facility. M. Fauvel wants to thank the CESBIO AI Team for fruitful discussions and the IOTA^2 core team for their helps, and the French Biodiversity Agency (OFB) for co-funding this project.

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.

Appendix A. Review of models

A.1. Conventional machine learning models

Ridge Regression is a regularized linear model that seeks a linear relationship between the predictors (here the Sentinel-2 spectro-temporal features) and the output (here the observed mowing date) (Hastie et al., 2001). A regularized version was used to cope with the high number of spectro-temporal features (Hastie et al., 2001, Chapter 3). This method serves as a baseline for supervised model: its learning capacity is limited w.r.t. other non-parametric regression methods but has provided accurate results for some case, such as *chlorophyll-a* concentration mapping (Ivanda et al., 2021). The regularization parameter value was selected using 10-folds cross-validation on the training data, as implemented in Scikit-learn (Pedregosa et al., 2011).

Random Forest is a non-parametric and non-linear regression model introduced by Breiman (2001). It is an ensemble-based model learning multiple independent decision trees, using bootstraps of training samples and features. It has been widely used in remote sensing time series applications, mainly for land cover/use mapping (Inglada et al., 2017) and estimation of continuous variables (Belgiu and Drăguţ,

[10] https://www.ofb.gouv.fr
[11] https://tempo.pheno.fr
[12] https://www.usanpn.org
[13] https://naturescalendar.woodlandtrust.org.uk
[14] http://www.pep725.eu

2016). Several hyperparameters can be selected for training. The most important one is the number of decision trees in the forest. As shown in Inglada et al. (2017), Fauvel et al. (2020), setting it to a large value is enough to provide accurate results. In this experiment we found that 100 trees was a good compromise: increasing the values did not lead to an improvement of the precision while the processing complexity (time and memory footprint) was much higher. Random Forest was implemented in Pedregosa et al. (2011).

A.2. Deep-learning models

One conventional and two advanced DL models were implemented: a Multilayer Perceptron (MLP), a 1D Convolutional Neural Network (1D-CNN) and the Lightweight Temporal Attention Encoder (LTAE), respectively. MLP was composed of three "linear layer + batchnormalization layer + rectified linear activation layer" modules and last linear output layer (Zhang et al., 2023). Such architecture has been widely used in remote sensing for land cover/use mapping (Kussul et al., 2017; Zhang et al., 2018, 2019) or land cover/use changes analysis (Vinayak et al., 2021).

1D-CNN was defined to perform along the temporal dimension, as in Kattenborn et al. (2021), Kussul et al. (2017), Guidici and Clark (2017), Zhong et al. (2019), Liao et al. (2020), Pelletier et al. (2019), to take into account the temporal dependence between the acquisition dates. From the MLP configuration, we replace the linear layer by a 1D convolutional layer and add max-pooling operation, as usually done with CNN models (Zhang et al., 2023).

LTAE used temporal attention mechanism to make use of the acquisition dates (Garnot and Landrieu, 2020). A temporal *attention head* computes a level of importance for each part/date of the satellite image time series w.r.t. the final task (here a regression task). It corresponds to the most important part of the signal which should be used for inference. In practice, it is computed through several linear and one softmax layers. Attention mechanism has showed to perform really well for land-cover mapping (Ofori-Ampofo et al., 2021; Garnot and Landrieu, 2020; Li et al., 2019; Bellet et al., 2023a). The same architecture proposed by Garnot and Landrieu (2020) was used in this work, the last layer and loss function was modified to perform regression rather than classification.

Appendix B. Threshold-based method

We implemented a specific mowing event detection algorithm introduced by Vroey et al. (2022) and integrated into the Sen4CAP toolbox (http://esa-sen4cap.org) to facilitate the monitoring of grassland management activities across Europe, aligning with the European Common Agricultural Policy. In our study, this method was adapted to detect first mowing event date, since it was primarily designed to detect mowing event time interval.

Vroey et al. (2022) proposed 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:

1. Each observation NDVI(t) is compared to the last available cloud-free observation NDVI($t-1$).
2. If the loss of NDVI, between NDVI(t) and NDVI($t-1$), is greater than 0.15 NDVI (NDVI(t) < NDVI($t-1$)-0.15), a mowing event is considered. As an additional condition, two consecutive mowing events must be separated by a minimum temporal distance of 28 days, and 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:

$$f(x; \min, \max) = \max - (\max - \min) \times \exp(-x), \tag{B.1}$$

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.

The first mowing event among the four most confident detection was retained, 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, additional checks in step 2 were ignored.

Appendix C. Grassland management map

A map of grassland management practices – *mowed* or *unmowed* – was generated to constrain mowing date prediction to areas of mowed grassland. We performed a pixel-based classification task within a nationwide grassland mask (Fig. 1), derived from permanent grassland plots declared in the 2022 LPIS (Section 2.1) from which landscape elements have been removed as much as possible by geometric operations (buffering, etc . . .). This database provides spatialized information on agricultural plot boundaries and crop types, but does not provide information about management practices.

Here, we trained a Random Forest classifier using a grassland management practices dataset, derived from our ground observations in 2022 described in Section 2.3. In this reference dataset, *mowed* class included 1 605 plots and *unmowed* class 660 plots (Table 2). Reference data were split at the plot level into a 70% training dataset and a 30% test dataset, ensuring classes and sites representation through stratified sampling. The split was done at the plot level, ensuring one plot to be either in the training or the test set preventing bias. Sentinel-2-based time series were used as predictor. In this dataset, in addition to 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).

The classification was done using IOTA^2 software (Inglada et al., 2016). 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.

It is important to note that all quantitative evaluation results presented in this paper were not based on the grasslands management map, as they were computed on the reference data (observed *mowed* plots). Hence, possible wrong declarations as well as non-homogeneous plots existing in the LPIS do not influence the quantitative accuracy assessment. This map was used for visual evaluation only, with all the limitations of such evaluation.

Appendix D. Tables and figures

See Tables D.6 and D.7 and Fig. D.13.

Appendix E. Supplementary materials

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.rse.2024.114476.The supplementary materials are available at the following link https://zenodo.org/records/11034843. The python scripts implementing the different algorithms are available at https://src.koda.cnrs.fr/pzzkfwbr/mowing-event-detection. Finally, the prediction raster map, as well as per French county vector files (plots summary) are provided here: https://zenodo.org/records/11034387.

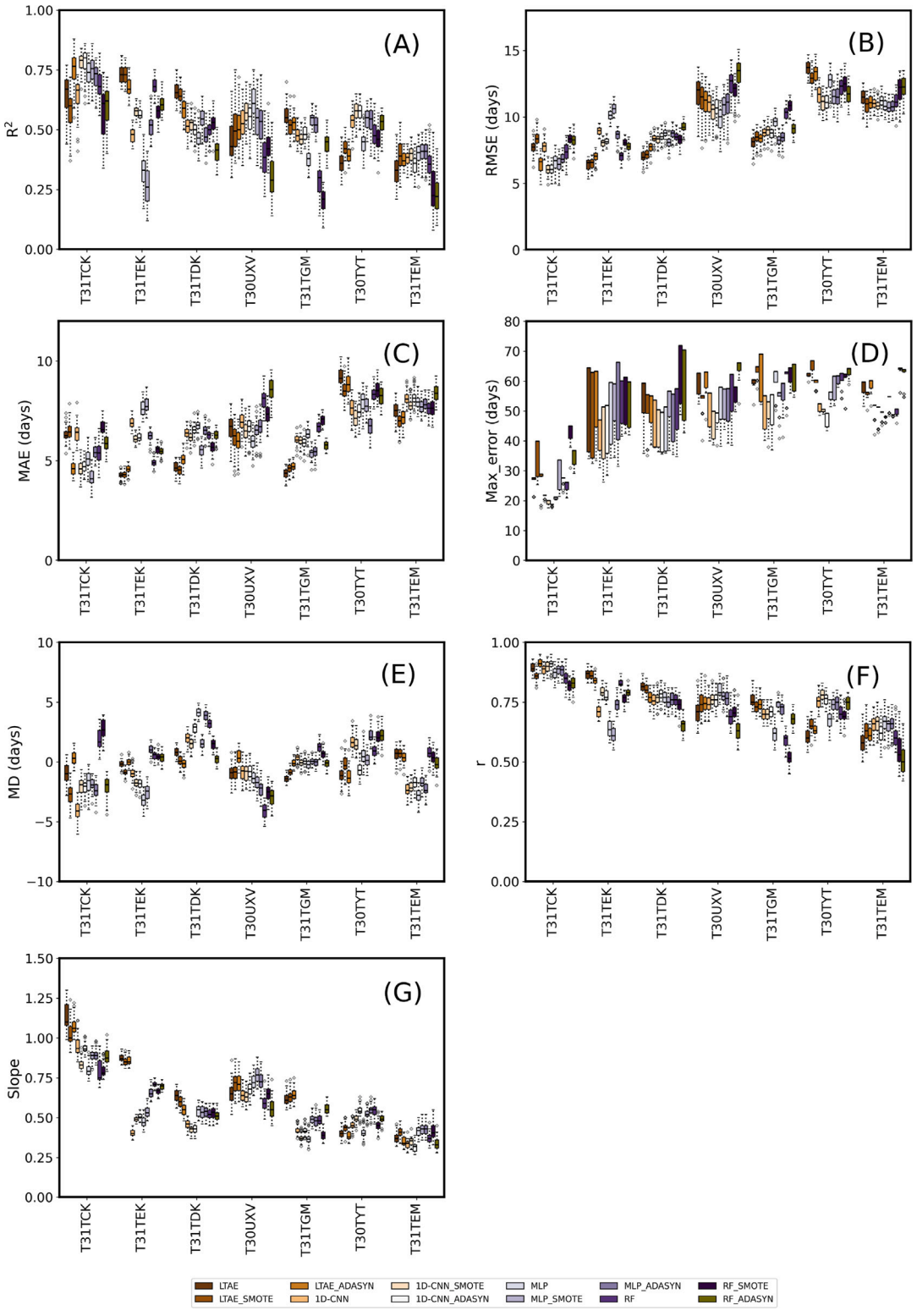

Fig. D.13. Site-specific outputs in terms of (A) R^2, (B) RMSE, (C) MAE, (D) Max_error, (E) Mean Deviation, (F) Correlation and (G) Slope. The sites are represented on the *x*-axis. The color palette represents the algorithms. For each site, fifty individual evaluations were conducted (from 50 folds by bootstrapping 70% of observations). Ridge Regression, SimpleMean and Threshold outputs are not shown in this figure.

Table D.6
Algorithm-specific statistical summary assessed at pixel-level. Each score value represents weighted mean of all sites. A site-specific score was weighted using the number of pixels used for the evaluation. A site-specific score represents mean of fifty individual evaluations (from 50 folds by bootstrapping 70% of observations). The lines are sorted based on R^2 values (descending order).

Algorithm	MAE	RMSE	R^2	MaxError	MD	Corr	Slope	Q_{50}	Q_{95}
LTAE_SMOTE	**5.50**	**9.11**	**0.52**	58.81	−0.37	**0.73**	**0.60**	**3.24**	20.21
LTAE	5.61	9.13	0.52	58.28	−0.50	**0.73**	0.59	3.35	**19.47**
LTAE_ADASYN	5.68	9.24	0.51	59.66	−0.07	0.72	0.58	3.28	20.77
MLP_ADASYN	6.36	9.39	0.49	54.00	0.10	0.72	0.52	4.22	20.38
1D-CNN_SMOTE	6.53	9.40	0.49	48.59	−0.50	0.72	0.41	4.43	20.28
1D-CNN_ADASYN	6.67	9.46	0.48	48.57	−0.42	0.71	0.44	4.62	20.28
MLP_SMOTE	6.50	9.50	0.47	54.23	−0.63	0.71	0.51	4.07	20.78
1D-CNN	6.84	9.60	0.47	49.30	−0.20	0.71	0.42	4.85	20.53
RF_ADASYN	6.71	10.04	0.42	61.30	**0.02**	0.66	0.52	4.25	21.60
RF	6.80	10.09	0.40	56.69	0.92	0.66	0.51	4.35	22.68
MLP	7.02	10.11	0.40	54.28	−0.77	0.66	0.44	4.83	21.71
RF_SMOTE	6.97	10.64	0.34	61.53	0.61	0.61	0.47	4.31	24.23
Ridge	8.62	11.45	0.23	72.57	1.18	0.58	0.43	6.58	23.36
SimpleMean	10.28	13.81	−0.10	**42.06**	0.19	0.00	0.00	5.52	26.78
Ridge_SMOTE	13.24	16.71	−0.76	87.89	1.74	0.43	0.51	11.02	32.80
Threshold	14.02	19.66	−1.36	73.24	−1.36	0.17	0.13	7.21	41.85
Ridge_ADASYN	17.11	21.17	−2.03	110.89	1.47	0.33	0.43	14.93	40.66

Table D.7
Algorithm-specific statistical summary assessed at plot-level. Each score value represents weighted mean of all sites. A site-specific score was weighted using the number of plots used for the evaluation. A site-specific score represents mean of fifty individual evaluations (from 50 folds by bootstrapping 70% of observations). The lines are sorted based on R^2 values (descending order).

Algorithm	MAE	RMSE	R^2	MaxError	MD	Corr	Slope	Q_{50}	Q_{95}
LTAE_ADASYN	**5.42**	**8.56**	**0.61**	41.32	**0.00**	**0.80**	0.59	3.49	18.38
LTAE_SMOTE	5.28	8.66	**0.61**	47.93	−0.37	0.79	**0.60**	3.38	17.99
LTAE	5.52	8.95	0.58	48.48	−0.48	0.76	0.59	**3.36**	**17.76**
RF_ADASYN	6.19	9.19	0.56	46.06	0.05	0.76	0.53	4.13	18.85
RF_SMOTE	6.51	9.59	0.51	46.73	0.48	0.72	0.49	4.27	19.75
1D-CNN_ADASYN	6.71	9.62	0.51	42.70	−0.37	0.74	0.44	4.52	20.56
MLP_SMOTE	6.46	9.66	0.51	45.73	−0.54	0.73	0.51	3.90	20.98
1D-CNN_SMOTE	6.60	9.68	0.51	43.09	−0.36	0.75	0.42	4.28	21.01
RF	6.61	9.73	0.50	45.76	0.75	0.72	0.51	4.47	20.24
MLP_ADASYN	6.45	9.75	0.50	48.05	0.25	0.74	0.52	4.17	21.05
1D-CNN	6.95	9.94	0.48	44.05	0.01	0.73	0.42	4.74	21.21
MLP	6.86	10.04	0.47	44.05	−0.45	0.71	0.45	4.57	21.15
Ridge	8.04	11.01	0.36	43.25	0.73	0.65	0.47	5.92	22.54
SimpleMean	10.78	14.47	−0.10	**41.14**	0.06	0.00	0.00	6.78	28.29
Ridge_SMOTE	11.57	14.63	−0.21	45.53	0.33	0.54	0.58	9.58	28.12
Threshold	12.98	17.24	−0.64	56.38	−3.92	0.33	0.25	9.79	35.04
Ridge_ADASYN	14.42	17.80	−0.93	50.07	0.43	0.46	0.53	12.57	34.47

Data availability

Data will be made available on request.

References

Belgiu, M., Drăguţ, L., 2016. Random forest in remote sensing: A review of applications and future directions. ISPRS J. Photogramm. Remote Sens. 114, 24–31.

Bellet, V., Fauvel, M., Inglada, J., 2023a. Land Cover Classification with Gaussian Processes using spatio-spectro-temporal features. IEEE Trans. Geosci. Remote Sens. http://dx.doi.org/10.1109/TGRS.2023.3234527, URL https://hal.science/hal-03781332.

Bellet, V., Fauvel, M., Inglada, J., Michel, J., 2023b. End-to-end Learning for Land Cover Classification using Irregular and Unaligned SITS by Combining Attention-Based Interpolation with Sparse Variational Gaussian Processes. IEEE J. Sel Top. Appl. Earth Observ. Remote Sens. URL https://hal.science/hal-04112115.

Bengtsson, J., Bullock, J.M., Egoh, B.N., Everson, C.S., 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 http://dx.doi.org/10.1002/ecs2.2582.

Breiman, L., 2001. Random forests. Mach. Learn. http://dx.doi.org/10.1023/a:1010933404324.

Broyer, J., Curtet, L., Boissenin, M., 2012. Does breeding success lead meadow passerines to select late mown fields? J. Ornithol. 153, 817–823.

Buri, P., Humbert, J.-Y., Stańska, M., Hajdamowicz, I., Tran, E., Entling, M.H., Arlettaz, R., 2016. Delayed mowing promotes planthoppers, leafhoppers and spiders in extensively managed meadows. Insect Conserv. Divers. http://dx.doi.org/10.1111/icad.12186.

Cantelaube, P., Carles, M., 2014. Le registre parcellaire graphique: des données géographiques pour décrire la couverture du sol agricole. Cahier Tech. l'INRA 58–64.

Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P., 2002. SMOTE: Synthetic minority over-sampling technique. J. Artificial Intelligence Res. http://dx.doi.org/10.1613/jair.953.

Dumeur, I., Valero, S., Inglada, J., 2024. Self-supervised spatio-temporal representation learning of satellite image time series. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 17, 4350–4367. http://dx.doi.org/10.1109/JSTARS.2024.3358066.

Escadafal, R., 1989. Remote sensing of arid soil surface color with landsat thematic mapper. Adv. Space Res. 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. Environ. Res. Lett. http://dx.doi.org/10.1088/1748-9326/aacc7a.

Fauvel, M., Lopes, M., Dubo, T., Rivers-Moore, J., Frison, P.-L., Gross, N., Ouin, A., 2020. Prediction of plant diversity in grasslands using sentinel-1 and -2 satellite image time series. Remote Sens. Environ. 237, 111536. http://dx.doi.org/10.1016/j.rse.2019.111536, URL https://www.sciencedirect.com/science/article/pii/S0034425719305553.

Garioud, A., Giordano, S., Valero, S., Mallet, C., 2019. Challenges in Grassland Mowing Event Detection with Multimodal Sentinel Images. In: 2019 10th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp). IEEE, Shanghai, France, pp. 1–4. http://dx.doi.org/10.1109/Multi-Temp.2019.8866914, URL https://hal.science/hal-02387167.

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

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. http://dx.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 Sens. 9 (6), 629.

Hao, X., Liu, L., Yang, R., Yin, L., Zhang, L., Li, X., 2023. A review of data augmentation methods of remote sensing image target recognition. Remote Sens. 15 (3), http://dx.doi.org/10.3390/rs15030827, URL https://www.mdpi.com/2072-4292/15/3/827.

Hastie, T., Tibshirani, R., Friedman, J., 2001. The elements of statistical learning. In: Springer Series in Statistics, Springer New York Inc., New York, NY, USA.

He, H., Bai, Y., Garcia, E.A., Li, S., 2008. ADASYN: Adaptive synthetic sampling approach for imbalanced learning. In: IEEE International Joint Conference on Neural Networks. http://dx.doi.org/10.1109/IJCNN.2008.4633969.

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. http://dx.doi.org/10.1016/j.rse.2023.113680.

Huang, J., Zhan, J., Yan, H., Wu, F., Deng, X., 2013. Evaluation of the impacts of land use on water quality: A case study in the chaohu lake basin. Sci. World J. http://dx.doi.org/10.1155/2013/329187.

Inglada, J., Vincent, A., Arias, M., Tardy, B., 2016. iota2-a25386. Zenodo, http://dx.doi.org/10.5281/zenodo.58150.

Inglada, J., Vincent, A., Arias, M., Tardy, B., Morin, D., Rodes, I., 2017. Operational high resolution land cover map production at the country scale using satellite image time series. Remote Sens. 9 (1), http://dx.doi.org/10.3390/rs9010095, URL https://www.mdpi.com/2072-4292/9/1/95.

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.

Jafarigol, E., Trafalis, T., 2023. A review of machine learning techniques in imbalanced data and future trends. arXiv preprint arXiv:2310.07917, abs/2310.07917, URL https://api.semanticscholar.org/CorpusID:263909438.

Joly, D., Brossard, T., Cardot, H., Cavailhes, J., Hilal, M., Wavresky, P., 2010. Les types de climats en France, une construction spatiale. Cybergeo: European Journal of Geography http://dx.doi.org/10.4000/cybergeo.23155, URL https://journals.openedition.org/cybergeo/23155#annexe, Publisher: CNRS-UMR Géographie-cités 8504.

Kattenborn, T., Leitloff, J., Schiefer, F., Hinz, S., 2021. Review on convolutional neural networks (CNN) in vegetation remote sensing. ISPRS J. Photogramm. Remote Sens. 173, 24–49.

Kingma, D., Ba, J., 2015. Adam: A method for stochastic optimization. In: International Conference on Learning Representations. ICLR, San Diego, CA, USA.

Klein, N., Theux, C., Arlettaz, R., Jacot, A., Pradervand, J.-N., 2020. Modeling the effects of grassland management intensity on biodiversity. Ecol. Evol. http://dx.doi.org/10.1002/ece3.6957.

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 Sens. http://dx.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. http://dx.doi.org/10.1038/s41598-022-04932-6.

Kooistra, L., Berger, K., Brede, B., Graf, L.V., Aasen, H., Roujean, J.-L., Machwitz, M., Schlerf, M., Atzberger, C., Prikaziuk, E., Ganeva, D., Tomelleri, E., Croft, H., Reyes Muñoz, P., Garcia Millan, V., Darvishzadeh, R., Koren, G., Herrmann, I., Rozenstein, O., Belda, S., Rautiainen, M., Rune Karlsen, S., Figueira Silva, C., Cerasoli, S., Pierre, J., Tanır Kayıkçı, E., Halabuk, A., Tunc Gormus, E., Fluit, F., Cai, Z., Kycko, M., Udelhoven, T., Verrelst, J., 2024. Reviews and syntheses: Remotely sensed optical time series for monitoring vegetation productivity. Biogeosciences 21 (2), 473–511. http://dx.doi.org/10.5194/bg-21-473-2024, URL https://bg.copernicus.org/articles/21/473/2024/.

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

LeCun, Y., Bengio, Y., Hinton, G.E., 2015. Deep learning. Nature http://dx.doi.org/10.1038/nature14539.

Lemaître, G., Nogueira, F., Aridas, C.K., 2017. Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning. J. Mach. Learn. Res. 18 (17), 1–5, URL http://jmlr.org/papers/v18/16-365.html.

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 Sens. 12 (5), 832.

Lin, T., Zhong, R., Wang, Y., Xu, J., Xu, J., Ying, Y., Rodríguez, L.F., Ting, K.C., Li, H., 2020. DeepCropNet: A deep spatial-temporal learning framework for county-level corn yield estimation. Environ. Res. Lett. http://dx.doi.org/10.1088/1748-9326/ab66cb.

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. http://dx.doi.org/10.1016/j.rse.2021.112751.

Lonjou, V., Desjardins, C., Hagolle, O., Petrucci, B., Tremas, T., Dejus, M., Makarau, A., Auer, S., 2016. MACCS-ATCOR joint algorithm (MAJA). In: Comerón, A., Kassianov, E.I., Schäfer, K. (Eds.), In: Remote Sensing of Clouds and the Atmosphere XXI, vol. 10001, International Society for Optics and Photonics, SPIE, 1000107. http://dx.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. Int. J. Remote Sens. http://dx.doi.org/10.1080/01431169608948714.

Metera, E., Sakowski, T., Sloniewski, K., Romanowicz, B., 2010. Grazing as a tool to maintain biodiversity of grassland - a review. Animal Sci. Pap. Rep. 28, 315–334.

Mohammed, R., Rawashdeh, J., Abdullah, M., 2020. Machine learning with oversampling and undersampling techniques: Overview study and experimental results. In: 2020 11th International Conference on Information and Communication Systems. ICICS, pp. 243–248. http://dx.doi.org/10.1109/ICICS49469.2020.239556.

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

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

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

Pontius, Jr., R.G., 2022. Metrics That Make a Difference: How to Analyze Change and Error. Springer International Publishing, http://dx.doi.org/10.1007/978-3-030-70765-1.

Rahman, A.A., Prasetiyowati, S.S., Sibaroni, Y., 2023. Performance analysis of the imbalanced data method on increasing the classification accuracy of the machine learning hybrid method. Jipi (J. Ilmiah Penelitian Dan Pembelajaran Inform.) http://dx.doi.org/10.29100/jipi.v8i1.3286.

Reinermann, S., Asam, S., Kuenzer, C., 2020. Remote sensing of grassland production and management—A review. Remote Sens. http://dx.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 (7), http://dx.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 Sens. Environ. http://dx.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. Ecol. Indic. http://dx.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..

Team, O.D., 2023. Orfeo ToolBox 8.1.2. Zenodo, http://dx.doi.org/10.5281/zenodo.8178641.

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.

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S.J., Brett, M., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., Carey, C.J., Polat, İ., Feng, Y., Moore, E.W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A.H., Pedregosa, F., van Mulbregt, P., SciPy 1.0 Contributors, 2020. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. Nature Methods 17, 261–272. http://dx.doi.org/10.1038/s41592-019-0686-2.

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 Sens. Environ. http://dx.doi.org/10.1016/j.rse.2022.113145.

Wang, X., Li, F.Y., Tang, K., Wang, Y., Guga, S., Bai, Z., Baoyin, T., 2019. Land use alters relationships of grassland productivity with plant and arthropod diversity in inner mongolian grassland. Ecol. Appl. http://dx.doi.org/10.1002/eap.2052.

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. http://dx.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.

Yang, Y., Tilman, D., Furey, G.N., Lehman, C., 2019. Soil carbon sequestration accelerated by restoration of grassland biodiversity. Nature Commun. http://dx.doi.org/10.1038/s41467-019-08636-w.

Zhang, A., Lipton, Z.C., Li, M., Smola, A.J., 2023. Dive into Deep Learning. Cambridge University Press, https://D2L.ai.

Zhang, H.K., Luo, D., Li, Z., 2024. Classifying raw irregular time series (CRIT) for large area land cover mapping by adapting transformer model. Sci. Remote Sens. 9, 100123. http://dx.doi.org/10.1016/j.srs.2024.100123, URL https://www.sciencedirect.com/science/article/pii/S2666017224000075.

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 J. Photogramm. Remote Sens. 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 Sens. Environ. 221, 173–187.

Zhao, Y., Liu, Z., Wu, J., 2020. Grassland ecosystem services: a systematic review of research advances and future directions. Landsc. Ecol. 35, 793–814.

Zhong, L., Hu, L., Zhou, H., 2019. Deep learning based multi-temporal crop classification. Remote Sens. Environ. 221, 430–443.