applied sciences

Article

Harvesting Insights from the Sky: Satellite-Powered Automation for Detecting Mowing Based on Predicted Compressed Sward Heights

Killian Dichou [1,*,†], Charles Nickmilder [1,†], Anthony Tedde [1,2], Sébastien Franceschini [1], Yves Brostaux [1], Isabelle Dufrasne [3], Françoise Lessire [3], Noémie Glesner [4] and Hélène Soyeurt [1]

[1] TERRA Research and Teaching Centre, Gembloux Agro-Bio Tech, University of Liège, 5030 Gembloux, Belgium; charles.nickmilder@uliege.be (C.N.); antho.tedde@gmail.com (A.T.); y.brostaux@uliege.be (Y.B.)
[2] National Funds for Scientific Research, 1000 Brussels, Belgium
[3] Faculty of Veterinary Medicine, University of Liège, Quartier Vallée 2, 4000 Liège, Belgium; isabelle.dufrasne@uliege.be (I.D.); flessire@uliege.be (F.L.)
[4] Fourrages Mieux ASBL, Horritine 1, Michamps, 6600 Bastogne, Belgium; glesner@fourragesmieux.be
* Correspondence: killian.dichou@uliege.be
† These authors contributed equally to this work.

Featured Application: The prospective identification of mowing events holds promise for application across diverse domains, including the assessment of arthropod biodiversity as an explanatory factor and the evaluation of general agricultural practices. An examination of their occurrences over time has the potential to enhance the efficient allocation of inherent territorial resources for animal feeding.

Abstract: The extensive identification of mowing events on a territory holds significant potential to help monitor shifts in biodiversity and contribute to assessing the impacts of drought events. Additionally, it provides valuable insights into farming practices and their consequential economic and ecological effects. To overcome challenges in obtaining reference grazing information directly from the field, this study introduces a novel methodology leveraging the compressed sward height (CSH) derived from Sentinel-1, Sentinel-2, and meteorological data, boasting an accuracy of 20 mm. Our central hypothesis posits that the mowing status of a parcel can be automatically discerned by analyzing the distribution and variation of its CSH values. Employing a two-step strategy, we first applied unsupervised algorithms, specifically k-means and isolation forest, and subsequently amalgamated the outcomes with a partial least squares analysis on an extensive dataset encompassing 194,657 pastures spanning the years 2018 to 2021. The culmination of our modeling efforts yielded a validation accuracy of 0.66, as ascertained from a focused dataset of 68 pastures. Depending on the studied year and with a threshold fixed at 0.50, 21% to 57% of all the parcels in the Wallonia dataset were tagged as mown by our model. This study introduces an innovative approach for the automated detection of mown parcels, showcasing its potential to monitor agricultural activities at scale.

Keywords: machine learning; compressed sward height; mowing

check for updates

Citation: Dichou, K.; Nickmilder, C.; Tedde, A.; Franceschini, S.; Brostaux, Y.; Dufrasne, I.; Lessire, F.; Glesner, N.; Soyeurt, H. Harvesting Insights from the Sky: Satellite-Powered Automation for Detecting Mowing Based on Predicted Compressed Sward Heights. *Appl. Sci.* **2024**, *14*, 1923. https://doi.org/10.3390/app14051923

Academic Editors: Samuel Adelabu, Romano Lottering and Kabir Peerbhay

Received: 22 January 2024
Revised: 21 February 2024
Accepted: 25 February 2024
Published: 26 February 2024

1. Introduction

With an average net annual productivity of 161 gC year^{-1} m^{-2} [1], grasslands emerge as crucial focal points in climate change mitigation efforts [2]. On the European stage, this significance is recognized as a pivotal driver in the implementation of greening policies [3,4]. Beyond environmental considerations, the economic facet is underscored by the integration of grass into cow diets, leading to a reduction in feed ration costs [5]. The frequency of mowing events, the timing of their occurrence, and the number of times they are executed can significantly influence forage yield and quality [6,7]. Additionally, mowing practices impact grassland biodiversity, affecting fauna heterogeneity, with potential consequences for arthropods [8] and specific cases such as butterflies, where it hinders the destruction of larvae and eggs [9]. Regarding the flora of grasslands and pastures, managing well the mowing events can promote plant species diversity [10], though other factors like hydrological considerations, as mentioned in [11], must be considered. Whether pertaining to cattle diets or biodiversity implications, awareness of mowing events can inform better management practices. From a societal perspective, grass plays a crucial role in shaping public perceptions of farming practices [3,12]. Furthermore, the inclusion of grass in the diet of dairy cows triggers alterations in milk fatty acid composition, such as an increase in C18:3, contributing to enhanced organoleptic and nutritional qualities attributed to pasture grazing or fresh grass intake [13,14]. To align with contemporary consumer expectations, dairy companies have established protected designations of origin mandating specific grazing parameters, either in terms of time periods or available areas.

Despite the existence of these specifications, there is a notable absence of factual verification regarding adherence to these conditions. As a response to this challenge, the Chronopâture tool was recently developed with the objective of automatically quantifying the number of days a herd spends grazing, utilizing GPS collars [15]. Unfortunately, the widespread deployment of such equipment encounters hindrances, as it entails a considerable financial investment. An alternative proposition by [16] suggests detecting a grass-based diet through milk mid-infrared spectral data, although its relevance is confined to dairy herds. Consequently, this has led to the exploration of mowing event detection through satellite data in various studies. Sentinel-1 (S1) [17] and Sentinel-2 (S2) data [18,19] are frequently deliberated upon, sometimes supplemented by Landsat 8 [20], and occasionally integrated with weather time series [21]. Various algorithms, including deep learning with neural networks [20,22], random forest, and support vector machine [21], have been tested. Key constraints of this approach highlighted in the literature include optical data availability, as seen in [18], where missing data were supplemented using S1 data. In other studies [20], availability and performance were only monitored over available data periods. A second constraint is the lack of comprehensive reference data [18], requiring grazing/management calendars. Both [17,23] underscore this challenge. However, various satellite-based proxies, such as the normalized difference vegetation index and the enhanced vegetation index values [21,23], and biomass assessments, like the leaf area index [23,24], have been employed for mowing event detection. Extending the range of proxies to grass height, as explored in [25,26], or even compressed sward height (CSH), as in [27], offers the opportunity to gather more reference data with reduced time and effort. CSH, measured under a rising plate meter, is more reproducible than classical grass height assessment and is often considered a reliable proxy for biomass [28]. Our team has recently developed models predicting CSH with around 20 mm accuracy from S1, S2, and meteorological data [27], facilitating on-site measurement circumvention with the rising plate mater. These models are integrated into a prediction platform, enabling longitudinal CSH prediction during the grazing period for all parcels in Wallonia [29]. Therefore, recognizing the documented impacts of mowing on forage availability, quality, and biodiversity [9,30], this study seeks to explore the feasibility of mowing event detection using the evolution of predicted CSH across the Walloon territory. The innovative approach promises improved accuracy by addressing the scarcity of reference data.

2. Materials and Methods

All computations in this study were conducted utilizing R v4.1.1 software [31] within the RStudio development environment [32]. Additionally, Python v3.9 [33] was employed in the Spyder development environment [34]. A comprehensive list of the packages utilized is provided in Table 1. For enhanced clarity in navigating the various stages of this study, the general workflow is outlined in Figure 1.

Table 1. List of packages used.

R			Python	
Software	**Version**	**Reference**	**Software**	**Reference**
R	4.1.1	[31]	os v3.9	[35]
raster	3.4-13	[36]	Re v3.9	[37]
sf	1.0-2	[38]	Pandas v1.4.1	[39]
data.table	1.14.0	[40]	Numpy v1.21.5	[41]
caret	6.0-93	[42]	Dask v2023.4.1	[43]
			scikit-learn v1.0.1	[44]

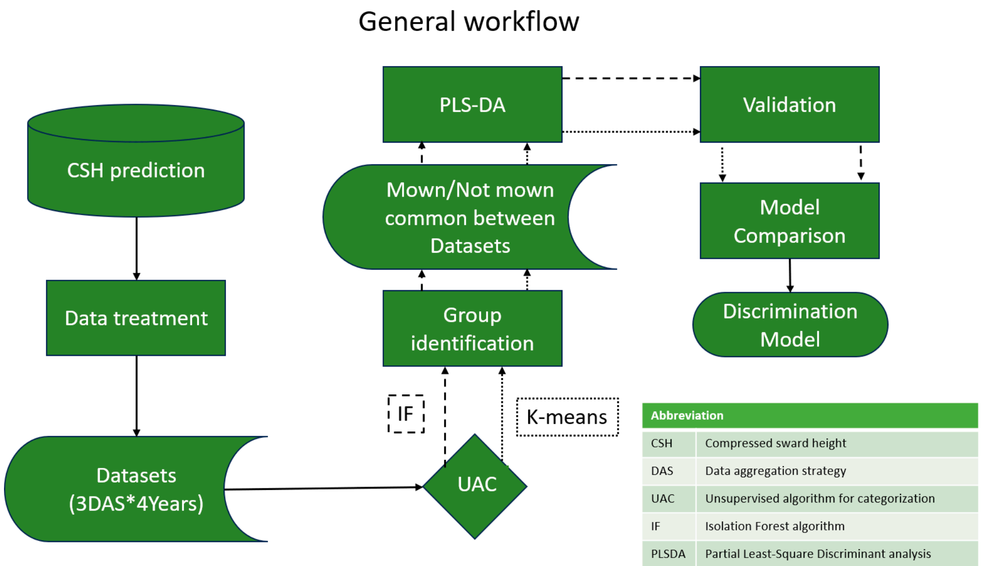

Figure 1. General workflow of the study.

2.1. Initial Data

The estimation process was anchored in the methodology outlined in [29], encompassing a substantial dataset of 10,883,214,548 pixel-based CSH predictions. These predictions, pertaining to 100 m^2 square pasture pixels on specific test dates, were documented between 2018 and 2021, covering 194,656 parcels situated in the Walloon region of Belgium. The overarching procedure involved retrieving S1 GRD products from the European Space Agency API. While the Scihub version [45] initially facilitated this operation, it was subsequently replaced by the Copernicus dataspace ecosystem [46]. The retro diffusivity of VV and VH polarization from S1 data was geocoded and computed to serve as inputs in the CSH prediction model. Simultaneously, S2 data in L2A format were obtained from the Theia [46] platform. Both the prepared S1 and S2 data underwent resampling to a 10 m resolution defined over the entire Walloon region of Belgium, with a specific mask applied to parcels designated as pastures in 2018. Subsequently, daily meteorological data encompassing air temperature (°C), wind speed (m/s), solar radiation (J/cm^2), precipitation (mm), relative air humidity (%), and potential evapotranspiration (mm/day) were retrieved from the Agromet platform [47]. Additionally, degree days, with a lower threshold of 0 °C and an upper threshold of 35 °C, were calculated. Cumulative degree days and precipitation over 3, 7, and 15 days were also computed. Employing a data augmentation strategy based on the prior use of data for up to 4 days in the past, with S2 acquisition serving as the reference for data aggregation initiation, both S1 and S2 data were aggregated. Following this, meteorological data were incorporated, associating parcels with the nearest meteorological station. Once the dataset was assembled for each acquisition date, predictions were executed at a pixel level using a Cubist prediction model, resulting in an estimated accuracy of approximately 20 mm [27].

Records with CSH predictions falling outside the defined calibration range, specifically [0;250] mm, were systematically removed to confine extrapolation (resulting in the deletion of 0.02% of records). Subsequently, records associated with parcels possessing an area smaller than 4900 m^2 were excluded to enhance the probability that most pixels constituting the parcel fully align with its boundaries (leading to the removal of 24.89% of parcels). The final filtering step involved retaining only records with more than three CSH predictions per parcel for the computation of descriptive statistics (resulting in the deletion of 0.004%). The cleaned dataset encompassed 39,578,872 pixels and 145,724 parcels. To conclude, these pixel-based data were aggregated per parcel and date to calculate the mean (MEAN_DAY) and standard deviation (SD_DAY). The descriptive statistics for this refined dataset are detailed in Table 2. Both MEAN_DAY and SD_DAY offer a straightforward yet effective means to differentiate potential groups based on predicted CSH.

Table 2. Summary statistics (mean ± SD) for predicted compressed sward height (CSH) at the pixel level, and aggregated averages per parcel and date (MEAN_DAY) with corresponding standard deviations (SD_DAY) across the 4-year period.

Year	Nparcels [1]	Pixel CSH (mm)	MEAN_DAY (mm)	SD_DAY (mm)
2018	86,218	56.0 ± 19.9	56.2 ± 17.5	7.0 ± 5.3
2019	86,617	63.8 ± 21.6	62.8 ± 17.6	8.1 ± 6.2
2020	91,224	60.1 ± 20.1	59.6 ± 16.6	7.5 ± 6.0
2021	136,892	59.2 ± 18.5	58.8 ± 15.7	7.3 ± 5.0

[1] Nparcels = number of parcels.

2.2. Data Preparation

Given the time-intensive nature of collecting grazing calendars and the resulting acquisition of a limited reference dataset, this study leveraged unsupervised learning to explore the potential for identifying the mowing status of a parcel based on the temporal variation of predicted CSH. The central working hypothesis posits that mowing events can be discerned from abrupt differences in CSH or related descriptive statistical parameters within the grazing period. To facilitate the identification of these potential patterns, the dataset underwent initial processing. Therefore, we investigated three data aggregation strategies (DAS) grounded in the distribution and variation of MEAN_DAY and SD_DAY over the grazing period. These strategies were tested both in combination and separately, with a reduction method applied to these features. The primary objective was to optimize computational resources and address concerns related to model overfitting. Specifically, the three DAS applied per parcel and year were the following: In Strategy 1, we aggregated MEAN_DAY and SD_DAY to compute the mean, standard deviation (sd), skewness (sk), and kurtosis (ku) for the considered grazing period (yearstats dataset: 8 features (4×2 features)). In Strategy 2, as the temporal density of data was not the same per parcel and year, we standardized MEAN_DAY and SD_DAY weekly (31 features as there are 31 weeks during the grazing season) for each studied trait (week_std dataset: 62 features (31×2 features)) by subtracting the mean and dividing by the standard deviation. Finally, in Strategy 3, fast Fourier transform (FFT), a dimensional reduction method, was applied to the week_std dataset to reduce the number of components to 10, which is a good compromise to limit the mirror effect of this kind of modeling (week_fft dataset: 20 features (10×2 features)). To optimize computational resources, all values in the datasets underwent rescaling before dimension reduction. This was achieved by dividing each value by 250, the maximum value within the calibration set used to construct the Cubist model for predicting CSH.

For the daily standardization outlined in Strategy 2, the recording days were organized in a daily-spaced dataset covering the period from 1st April to 3rd November (e.g., 217 days). The choice of 3rd November was made to evenly divide the dataset into seven segments, facilitating comparison with other time-standardized datasets. Subsequently, date gaps (highlighted in red in Figure 2) where MEAN_DAY and SD_DAY were unavailable due to meteorological reasons were filled using linear interpolation, as we made the hypothesis that grass growth is linear within a short period, summarized as follows:

$$CSH_i = CSH_{prev} + (CSH_{next} - CSH_{prev}) \times \frac{dt}{ndays}$$

where CSH_i = the interpolated date, CSH_{prev} = the previous date of CSH prediction, CSH_{next} = the next available date of CSH prediction, dt = the number of days between the previous prediction and the interpolated date, and $ndays$ = the length of the time gap between the previous and next predictions. For values between the first record and the first date and between the last record and the last date, linear interpolation was not possible, so the closest MEAN_DAY or SD_DAY value was attributed, as illustrated in blue in Figure 2. For the week standardization, the MEAN_DAY and SD_DAY values covered in the 217-day standardized dataset were averaged for each of the 31 adjacent periods of 7 days.

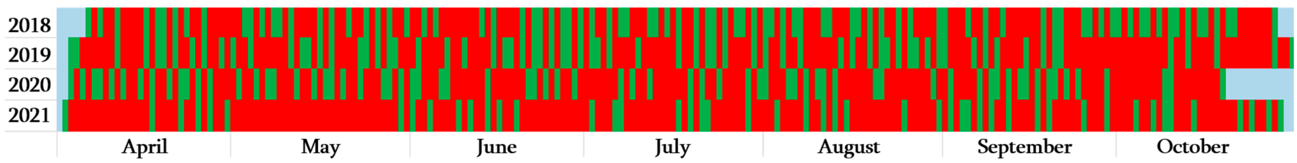

Figure 2. Temporal density of the CSH predictions across the four studied grazing seasons. Green marks actual prediction dates, red marks temporal areas with a linear interpolation to fill the gaps, and light blue marks the periods with a closest value replacement to fill the gaps.

2.3. Unsupervised Analysis to Detect Mowing Events

The 12 datasets (resulting from 4 years × 3 DAS) served as input for two distinct unsupervised algorithms for categorization (UACs): isolation forest (IF) and k-means. The IF algorithm, proposed by [48], employs random partitioning of the dataset to identify anomalies. Given the expectation that mown parcels would exhibit different behavior than grazed parcels, it can be assumed a priori that mown parcels would be detected as an abnormal population. To this end, the contamination threshold was set to 0.2, a value aligned with the prevailing management practices in the Walloon region of Belgium [49]. Indeed, the predominant management modality for permanent pasture, comprising 42.7% of the usable agricultural area, involves either grazing or a combination of grazing and mowing, which is ten times more extensive than temporary pastures in terms of surface area [50]. The second unsupervised algorithm tested was based on a k-means algorithm [51]. The k-means algorithm is a clustering method that segregates a dataset into k distinct, non-overlapping subsets called clusters by iteratively assigning data points to the cluster with the nearest mean and updating the cluster centroids. As this method necessitates specifying the number of clusters beforehand, 2 to 5 clusters were tested to evaluate the partitioning. From the requirement for interpretability, a 5-cluster limitation was chosen to parameterize the k-means.

The interpretation of the clusters generated by both UACs relied on the following hypothesis: groups exhibiting higher average values for MEAN_YEAR and SD_YEAR should predominantly consist of mown parcels. This was inferred from the expectation that grass height in mown parcels would be significantly higher than in grazed parcels until the mowing event, and the variability in this parameter would be more pronounced. Following the identification and labeling of mown and pasture clusters, all parcels categorized within these groups over the 4-year period were incorporated into three distinct datasets corresponding to week_std, week_fft, and year-stats datasets.

Subsequently, supervised learning was conducted using partial least squares discriminant analysis (PLS-DA) on the three datasets, where the mown character was designated as "1" and the pasture character as "0". PLS-DA aims to identify patterns indicative of mowing events by maximizing the correlation between these variables and the occurrence of mowing events. To address the imbalanced proportion between mown and pasture modalities, the up-sampling technique from the caret package was applied before the PLS-DA. Outputs from the stratified ten-fold cross-validation informed the determination of the number of components in the PLS-DA model, mitigating overfitting and providing an assessment of the overall accuracy of the obtained classification. In addition to cluster predictions, PLS-DA enabled the prediction of the probability of being in the mown cluster. To evaluate the similarity in signals produced by these different models, correlations between the probabilities obtained from the six constructed PLS-DA models were calculated. Furthermore, for the three types of datasets, we examined the number of occurrences that would be labeled as mown with a probability exceeding 50% for k-means and IF.

2.4. Description of the Validation Set

Segmenting and interpreting clusters provide a method to distinguish between mown and unmown parcels. However, it is essential to recognize that the machine's perception might not always align with actual field reality. To externally validate the findings of our study, we utilized grazing calendars collected by our partners at the "Centre des Technologies Agronomiques" (CTA, Strée, Belgium) and the "Fourrages Mieux" ASBL (Bastogne, Belgium), covering the same timeframe as the one used for CSH predictions. The validation set comprised 55 distinct parcels with confirmed mowing events, situated in the Walloon region of Belgium. Given that certain parcels were affiliated with the same farms, Figure 3 provides an overview of their spatial distribution. For most parcels, we achieved the precision of predicting the mowing date within one or two days. In the case of three parcels, our knowledge was limited to the occurrence of mowing events during the summer period. In total, this dataset encompasses 55 distinct parcels, and Table 3 provides a detailed description of this validation dataset. From this validation set, the rate of correctly classified mown parcels was estimated for each UAC and type of dataset.

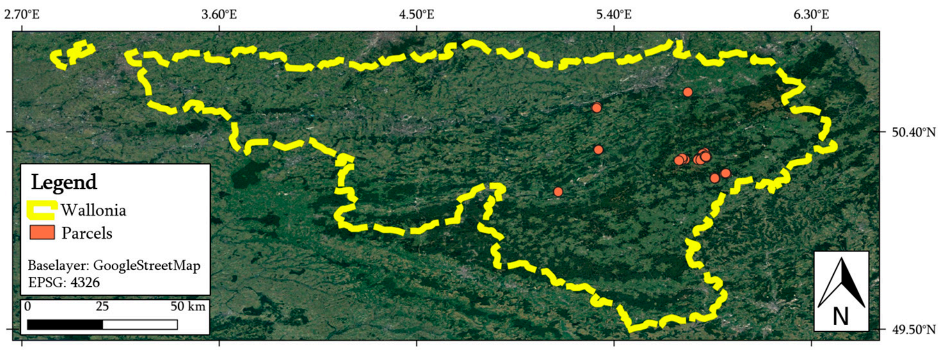

Figure 3. Walloon region of Belgium and parcels position.

Table 3. Description of the validation set.

Temporality	Category	2018	2019	2020	2021
With a precise date	N events	12	11	44	0
	N farms	2	4	5	0
	N parcels	12	11	11	0
	N refusal mowing [1]	0	0	8	0
Without the precise date	N events = N parcels	3	3	3	3

[1] refusal mowing = mowing events occurring to cut the grass patches that were not grazed on a pasture.

3. Results

As was explained in the introduction, training models to precisely detect mowing events over the year requires extensive knowledge of the mowing dates. This proves to be a challenge to acquire on a large scale. Therefore, this study introduced an alternative approach using CSH predicted from S1, S2, and meteorological data. Employing a multiyear perspective spanning from 2018 to 2021, the study focused on entire grazing seasons, detecting distinctive patterns through features derived from the predicted CSH to categorize parcels as mown or not mown. Three feature datasets—week_std, week_fft, and yearstats—were tested, along with two UACs, namely, k-means and IF.

Applying the threshold based on standard practices in grazing management in southern Belgium, IF divided 80% of the samples into a normal population (referred to as Group 0) and 20% into an abnormal population (referred to as Group 1). Mown parcels are anticipated to exhibit concurrently higher values for MEAN_YEAR and SD_YEAR. Some descriptive statistics of these traits for each of the two groups identified by IF are presented in Table 4. Across all tested datasets, MEAN_YEAR and SD_YEAR were consistently higher for Group 1, indicating that this group likely comprises the mown parcels.

Table 4. Descriptive statistics of MEAN_YEAR and SD_YEAR of predicted compressed sward height (in mm) for each isolation forest (IF). Groups with potential mown parcels are highlighted in bold. Groups with the opposite characteristics are underlined.

		week_fft		week_std		yearstats	
Year	**IF**	**MEAN_YEAR**	**SD_YEAR**	**MEAN_YEAR**	**SD_YEAR**	**MEAN_YEAR**	**SD_YEAR**
2018	Group 0	<u>55.21</u>	<u>6.8</u>	<u>55.04</u>	<u>6.71</u>	<u>55.58</u>	<u>6.99</u>
	Group 1	**59.05**	**8.64**	**59.78**	**9.01**	**57.85**	**8**
2019	Group 0	<u>61.41</u>	<u>7.71</u>	<u>61.17</u>	<u>7.5</u>	<u>61.86</u>	<u>7.83</u>
	Group 1	**68.19**	**10.35**	**69.2**	**11.21**	**66.7**	**10.02**
2020	Group 0	<u>58.78</u>	<u>7.33</u>	<u>58.22</u>	<u>6.93</u>	<u>58.69</u>	<u>7.26</u>
	Group 1	**62.77**	**9.01**	**64.99**	**10.58**	**63.12**	**9.29**
2021	Group 0	<u>57.21</u>	<u>6.76</u>	<u>56.77</u>	<u>6.53</u>	<u>57.54</u>	<u>6.85</u>
	Group 1	**63.18**	**9.26**	**64.51**	**9.95**	**62.2**	**9.03**

The k-means algorithm identified five clusters, designated as clusters 0 to 4 in Table 5. However, the assignment of cluster labels may vary across datasets, as it is influenced by the distinct starting points and differences in observations and features inherent to k-means. Consequently, the first step is to highlight the clusters associated with mowing events for each dataset. Following a methodology akin to the one utilized for IF, data labeling was performed by interpreting the descriptive statistics of MEAN_YEAR and SD_YEAR. This analytical process allowed the identification of 1 to 3 clusters associated with mown events for each dataset, as delineated in Table 5. By positing the working hypothesis that a parcel tagged as mown in each DAS for the same year is more likely to be a mown parcel, we could establish an indirect reference dataset. Conversely, a similar rationale was applied for the detection of grazed parcels. Table 6 provides the count of common parcels identified as mown or grazed based on the studied year and the UAC used. We were thus able to consider the number of parcels flagged as mown by the three datasets. For each of them, across the 4 years, the mean variation of the percentage of tagged mown parcels ranged from 0.39 to 0.40 for k-means, and from 0.19 to 0.23 for IF, as presented in Table 7.

Assuming the correctness of the parcel labels, we constructed a comprehensive reference dataset that encompassed all parcels labeled for each studied year. Leveraging this new dataset enabled the application of supervised learning, which was executed using PLS-DA. The prediction performance, obtained through a 10-fold cross-validation, is presented in Table 8 for each feature matrix. The discrimination between mowing and grazing events proved to be excellent, with observed accuracies surpassing 93% for all models. Models utilizing datasets generated from the k-means segregation technique exhibited higher accuracy compared to those using the IF method.

Table 5. Descriptive statistics of MEAN_YEAR and SD_YEAR of predicted compressed sward height (in mm) for each k-means clustering. Clusters with potentially mown parcels are highlighted in bold. Clusters with the opposite characteristics are underlined.

Year	K-Means	week_fft		week_std		yearstats	
		MEAN_YEAR	SD_YEAR	MEAN_YEAR	SD_YEAR	MEAN_YEAR	SD_YEAR
2018	Cluster 0	53.3	5.92	56.67	10.58	57.65	8.12
	Cluster 1	57.03	7.14	57.92	7.69	57.26	6.8
	Cluster 2	55.43	7.18	53.35	6.86	**60.68**	**9.38**
	Cluster 3	54.73	5.96	54.73	5.96	54.73	5.96
	Cluster 4	**59.09**	**8.86**	**59.09**	**8.86**	59.09	8.86
2019	Cluster 0	61.47	7.35	65.18	9.08	57.93	6.3
	Cluster 1	58.18	6.39	62.82	7.57	58.61	6.37
	Cluster 2	**62.84**	**8.29**	**69.54**	**11.38**	**64.74**	**9.17**
	Cluster 3	**62.75**	**8.25**	62.75	8.25	62.75	8.25
	Cluster 4	**62.87**	**8.27**	62.87	8.27	62.87	8.27
2020	Cluster 0	53.28	5.68	**63.38**	**9.56**	55.48	5.67
	Cluster 1	**63.61**	**9.47**	55.6	5.93	**62.73**	**8.12**
	Cluster 2	57.54	6.95	61.24	7.37	**61.33**	**8.75**
	Cluster 3	59.76	7.73	59.76	7.73	59.76	7.73
	Cluster 4	59.74	7.71	59.74	7.71	59.74	7.71
2021	Cluster 0	56.3	6.28	54	6.03	59.02	6.61
	Cluster 1	**62.71**	**8.72**	**58.76**	**10.8**	**64.81**	**10.32**
	Cluster 2	59.87	7.7	58.93	6.75	62.82	8.18
	Cluster 3	58.7	7.34	58.7	7.34	58.7	7.34
	Cluster 4	58.66	7.36	58.66	7.36	58.66	7.36

Table 6. Number of parcels labeled as mown or grazed simultaneously by models using the 3 different feature matrices.

Year	K-Means		Isolation Forest	
	Seemingly Mown Parcels	Seemingly Grazed Parcels	Seemingly Mown Parcels	Seemingly Grazed Parcels
2018	2901	3086	5343	54,267
2019	4335	19,880	5923	54,904
2020	8082	1296	5610	56,698
2021	750	2454	10,601	88,201
TOTAL	16,068	26,716	27,477	250,070

Table 7. Percentage of parcels tagged as mown in each dataset with both UACs.

Year	K-Means			Isolation Forest		
	Weekstd	Weekfft	Yearstats	Weekstd	Weekfft	Yearstats
2018	0.38	0.42	0.44	0.20	0.21	0.28
2019	0.38	0.41	0.39	0.20	0.21	0.22
2020	0.56	0.57	0.50	0.18	0.19	0.27
2021	0.23	0.21	0.29	0.17	0.17	0.14
Mean	0.39	0.40	0.40	0.19	0.19	0.23

Table 8. Prediction performance of the partial least squares discriminant analysis to predict seemingly mowing events.

Algorithm	Dataset	N Components	Calibration Accuracy	Validation Accuracy (N = 68)
K-means	week_std	4	0.99 ± 0.0004 (N = 42,784)	0.48 (N = 33)
	week_fft	7	0.99 ± 0.0004 (N = 42,784)	0.66 (N = 45)
	yearstats	6	0.98 ± 0.0011 (N = 42,784)	0.54 (N = 37)
Isolation forest	week_std	7	0.95 ± 0.0006 (N = 277,547)	0.18 (N = 12)
	week_fft	9	0.95 ± 0.0005 (N = 277,547)	0.31 (N = 21)
	yearstats	6	0.94 ± 0.0006 (N = 277,547)	0.40 (N = 27)

To assess the coherence among the different models, we computed the correlations of their probabilities of being mown parcels, depicted in Figure 4. The observed positive correlations suggested a similar signal across the various models, even though they were all executed using a reference dataset created based on the UAC outputs. However, these correlations did not reach 1, indicating some discrepancies between the labels assigned to parcels by the different algorithms. The highest similarity was noted between the PLS-DA models developed from the week_std k-means and week_fft k-means datasets. As shown in Figure 4, PLS-DA models created from datasets generated through the IF and k-means techniques did not exhibit strong convergence; their correlation was generally below 0.80 in most cases. This contrast is further highlighted when comparing the reference datasets used to execute the PLS-DA. Only 1858 parcels/year were consistently labeled as mown by these two methods, representing 11.56% in the k-means dataset and 6.76% in the IF dataset.

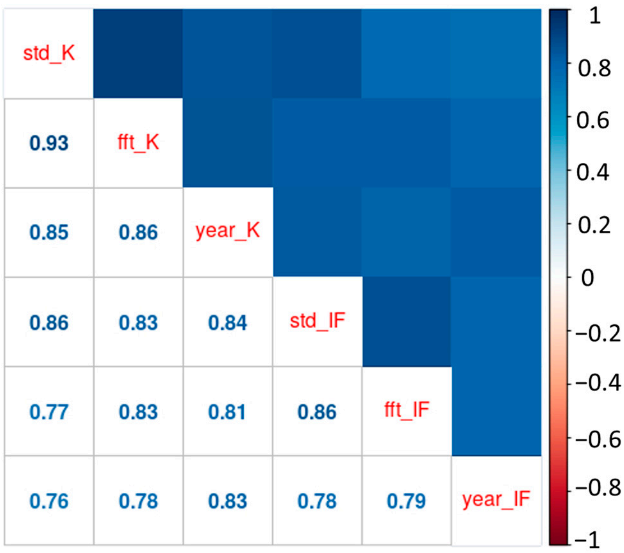

Figure 4. Correlogram of the probability of being a mown parcel following the 6 partial least squares discriminant analyses performed.

To evaluate the performance of the models, external validation was required as the reference values were established through unsupervised learning. According to the validation performance detailed in Table 8, PLS-DA demonstrated superior performance, particularly when utilizing datasets derived from k-means, with a validation accuracy reaching up to 0.66, a criterion for discriminating the best model. An indirect validation approach involves applying the developed PLS-DA models to the entire dataset to observe the number of detected mown parcels. Table 9 presents the count of parcels labeled as mown by the best PLS-DA algorithm for each studied year. The probability threshold set to label a parcel as mown was defaulted to 0.5. However, testing alternative threshold values resulted in noticeable variations in the number of parcels tagged as mown, as shown in Table 9. Consequently, fixing this threshold would be necessary. Unfortunately, the available validation set did not provide the number of mowing events associated with each parcel.

Table 9. Percentage of parcels labeled as mown in the entire dataset using the PLS-DA based on the week_fft dataset with the k-means approach for several thresholds of mown probability (in parentheses).

Year	Total Parcels	Percentage of Parcels Flagged as Mown				
		week_fft (0.50)	week_fft (0.45)	week_fft (0.40)	week_fft (0.35)	week_fft (0.30)
2018	86,218	42	53	65	77	88
2019	86,617	41	48	57	66	78
2020	91,224	57	66	75	84	91
2021	136,892	21	27	36	47	65
Total	400,915	40	49	58	68	80

4. Discussion

As highlighted in the introduction, there is a growing interest in understanding grazing and mowing management practices, particularly to assess the potential environmental impacts [6,8,11]. This could be addressed with the creation of a grazing calendar, but its updating proves to be time-consuming, limiting its adoption among farmers [20]. Another possible solution is the utilization of GPS collars [15], but despite their effectiveness, they face restrictions due to the significant financial investment involved. An alternative that holds appeal for farmers is associated with the use of readily available spatial data, especially through projects like Copernicus [46]. However, the varying availability of optical data and the lack of comprehensive reference data pose constraints on the development of models [18,20] for detecting mowing events. The innovative aspect of this study lies in the application of unsupervised learning based on CSH predicted from S1, S2, and meteorological data. This approach aims to address the challenge of the scarcity of reference data, providing a novel solution for mowing event detection.

The detection of mowing clusters was possible with both UACs, but only 1858 parcels/year were consistently labeled as mown by these two techniques, representing 11.56% in the k-means dataset and 6.76% in the IF dataset. This difference could be explained by the fact that for the IF algorithm, the isolation of parcels considered mown was based on a fixed threshold (i.e., 20%). To solve this issue, we considered examining the percentage of parcels flagged as mown by the three datasets, as shown in Table 7. The results showed mean variation from 0.39 to 0.40 for k-means and from 0.19 to 0.23 for IF across the 4 years. The percentage observed for k-means is therefore more in line with the Walloon statistics [50].

This previous step, which allowed labeling of parcels that seem to be affected by mowing events, can be used to train a PLS-DA with the aim of automatically detecting mowing events based on a large dataset. Indeed, the dataset size was 42,784 for k-means and 277,547 for IF. This has never been used in past studies related to the detection of the mowing effect. Indeed, they used between 236 and 1200 reference records [17,18,20,21]. The size of our training set, reachable with the unsupervised learning method we presented, is interesting to improve the robustness of a model [52]. The PLS-DA models showed a good ability to detect seemingly mowing events, with a calibration accuracy ranging from 0.94 to 0.99. The ability of the k-means output to build the PLS-DA models was also better.

To check the ability to detect confirmed mowing parcels, a validation was conducted using a limited dataset (N = 68). The IF method employed in the study exhibited lower accuracy, reaching a maximum detection rate of 40%, as shown in Table 8. Two main hypotheses can be proposed to explain this result: either the model was trained with a fixed threshold that does not match enough with the actual number of mowing events, or the validation data had patterns that did not align well with the eventual patterns found by the IF. It raises a question about the effectiveness of the method, especially given the temporal nature of the data involved in this study. Consequently, our choice is to push the use of k-means during unsupervised learning because the validation accuracy reached 0.66. Even if this value was lower than the one observed for calibration, a 66% accuracy on a large scale is deemed satisfactory when examining the literature. For instance, in [18], continuous validation information resulted in the correct detection of 63% of mowing events. Higher detection rates, up to 86%, have been reported [17] using change detection in the time series of S1 and S2 data. However, it is important to note in this case that the precision of this model dropped to 57% when considering all grassland types; our study also considered all grassland types but considered them as pasture. Challenges in identifying mowing events in this study may stem from the mixed practices of farmers, as evidenced in [18], where half of the false positives were detected in pastures that were also grazed. Another potential factor affecting model performance is intense grazing over a short period, identified as a significant confounding factor in [17].

Now, we will examine the potential significance of the developed predictive models. The aim of this study was to create a model for identifying mown or grazed parcels using routinely recorded data (S1, S2, and meteorological data). This model serves the purpose of verifying compliance with label specifications such as "Happy Marguerite cow" in Belgium or "Lait de pâturage" in France, ensuring that cows are genuinely on pastures for a specific duration rather than being fed with mown grass [15]. While the 66% validation performance achieved might seem a bit low, it is important to note that the validation set consisted of only 68 records. Enhancing the performance of this potential labeling check could involve supplementing the obtained probability with information about milk composition. It is well established that milk composition, particularly the fatty acid profile, is influenced by diet.

This method opens up another potential application: estimating damages, such as those resulting from droughts. Utilizing a meticulously assessed small reference dataset, the study suggests that mowing events may diminish grassland resistance to drought [53]. For instance, mowing can limit the water uptake capacity of plants by directly reducing their evapotranspiration [54]. This application is of particular significance in Wallonia, where the extensive time required for drought damage assessments often results in a standardized rate assigned to the area without a precise evaluation of farmers' losses. Implementing this approach could offer advantages for both insurers and farmers, ensuring more accurate financial compensation that aligns with the actual losses incurred.

5. Conclusions

This study aimed to develop models for detecting mown or grazed parcels using easily recorded routine data, showcasing variable detection performance based on the dataset and the unsupervised method employed. Predictions from the k-means technique outperformed those derived from the IF outputs. The highest validation performance, reaching 66%, was achieved by employing a PLS-DA with a dataset utilizing k-means and incorporating week_fft features. The study's approach aims to address the challenge of acquiring reference data by leveraging all data acquired in the validation process and employing unsupervised methods for model calibration. Utilizing a larger training dataset leads to increased variability, thus enhancing the model's robustness. Moreover, incorporating a proxy like the CSH, which considers multiple data sources from S1, S2, and meteorological data, reduces the models' dependency on external factors, a limitation often encountered when exclusively using remote sensing data. We suggest further refinement of the initial PLS-DA model by implementing an assessed gap-filling method and employing a larger validation set. These aspects, challenging to explore in depth in this study, are crucial steps before considering the widespread application of this algorithm in areas such as label marking, land-use changes, or drought assessment analyses.

Author Contributions: Conceptualization, K.D., C.N., A.T. and H.S.; methodology, K.D., C.N. and H.S.; software, K.D. and C.N.; validation, K.D. and C.N.; formal analysis, K.D., C.N., S.F. and Y.B.; investigation, I.D., F.L. and N.G.; resources, I.D., F.L. and N.G.; data curation, K.D. and C.N.; writing—original draft preparation, K.D. and C.N.; writing—review and editing, K.D., C.N. and H.S.; visualization, K.D. and C.N.; supervision, H.S.; project administration, H.S.; funding acquisition, H.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Service Public de Wallonie with the projects ROAD-STEP [grant number D31-1393/S1] and WALLeSmart [grant number D65-1435] and by the Luxembourg National Research Fund with the project SIMBA [grant number T.0221.19].

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Dataset available on request from the authors. The raw data supporting the conclusions of this article will be made available by the authors on request.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Gourlez de la Motte, L.; Jérôme, E.; Mamadou, O.; Beckers, Y.; Bodson, B.; Heinesch, B.; Aubinet, M. Carbon Balance of an Intensively Grazed Permanent Grassland in Southern Belgium. *Agric. For. Meteorol.* **2016**, *228–229*, 370–383. [CrossRef]
2. Erb, K.-H.; Kastner, T.; Plutzar, C.; Bais, A.L.S.; Carvalhais, N.; Fetzel, T.; Gingrich, S.; Haberl, H.; Lauk, C.; Niedertscheider, M.; et al. Unexpectedly Large Impact of Forest Management and Grazing on Global Vegetation Biomass. *Nature* **2018**, *553*, 73–76. [CrossRef] [PubMed]
3. Lessire, F.; Jacquet, S.; Veselko, D.; Piraux, E.; Dufrasne, I. Evolution of Grazing Practices in Belgian Dairy Farms: Results of Two Surveys. *Sustainability* **2019**, *11*, 3997. [CrossRef]
4. European Comission EU Agrees to Increase Carbon Removals. Available online: https://ec.europa.eu/commission/presscorner/detail/en/IP_22_6784 (accessed on 1 February 2023).
5. Dillon, P.; Hennessy, T.; Shalloo, L.; Thorne, F.; Horan, B. Future Outlook for the Irish Dairy Industry: A Study of International Competitiveness, Influence of International Trade Reform and Requirement for Change. *Int. J. Dairy Technol.* **2008**, *61*, 16–29. [CrossRef]
6. Fulkerson, W.J.; Michell, P.J. The Effect of Height and Frequency of Mowing on the Yield and Composition of Perennial Ryegrass—White Clover Swards in the Autumn to Spring Period. *Grass Forage Sci.* **1987**, *42*, 169–174. [CrossRef]
7. Rotz, C.A. How to Maintain Forage Quality during Harvest and Storage. *Adv. Dairy Tecnol.* **2003**, *15*, 227–239.
8. Impact of Mowing Frequency on Arthropod Abundance and Diversity in Urban Habitats: A Meta-Analysis—ScienceDirect. Available online: https://www.sciencedirect.com/science/article/pii/S1618866722002576 (accessed on 16 January 2024).
9. Influence of Mowing on the Persistence of Two Endangered Large Blue Butterfly Species—JOHST—2006—Journal of Applied Ecology—Wiley Online Library. Available online: https://besjournals.onlinelibrary.wiley.com/doi/full/10.1111/j.1365-2664.2006.01125.x (accessed on 16 January 2024).
10. Effectiveness of Mowing for the Flora Diversity Preservation: A Case Study of Steppe-like Grassland in Croatia (NATURA 2000 Site) | Biodiversity and Conservation. Available online: https://link.springer.com/article/10.1007/s10531-016-1252-1 (accessed on 16 January 2024).
11. Kołos, A.; Banaszuk, P. Mowing May Bring about Vegetation Change, but Its Effect Is Strongly Modified by Hydrological Factors. *Wetl. Ecol. Manag.* **2018**, *26*, 879–892. [CrossRef]
12. Michaud, A.; Plantureux, S.; Baumont, R.; Delaby, L. Les prairies, une richesse et un support d'innovation pour des élevages de ruminants plus durables et acceptables. *INRAE Prod. Anim.* **2020**, *33*, 153–172. [CrossRef]
13. Chilliard, Y.; Ferlay, A.; Mansbridge, R.M.; Doreau, M. Ruminant Milk Fat Plasticity: Nutritional Control of Saturated, Polyunsaturated, Trans and Conjugated Fatty Acids. *Ann. Zootech.* **2000**, *49*, 181–205. [CrossRef]
14. Elgersma, A.; Tamminga, S.; Ellen, G. Modifying Milk Composition through Forage. *Anim. Feed Sci. Technol.* **2006**, *131*, 207–225. [CrossRef]
15. Caraes, C.; Hebert, J. Agriculture et Élevage de Précision: Comptabiliser Le Temps de Pâturage Grâce à Un Collier Connecté: Chronopâture 2020. Available online: https://normandie.chambres-agriculture.fr/ (accessed on 20 July 2020).

16. Soyeurt, H.; Gerards, C.; Nickmilder, C.; Bindelle, J.; Franceschini, S.; Dehareng, F.; Veselko, D.; Bertozzi, C.; Gengler, N.; Marvuglia, A.; et al. Prediction of Indirect Indicators of a Grass-Based Diet by Milk Fourier Transform Mid-Infrared Spectroscopy to Assess the Feeding Typologies of Dairy Farms. *Animals* **2022**, *12*, 2663. [CrossRef]
17. De Vroey, M.; Radoux, J.; Defourny, P. Grassland Mowing Detection Using Sentinel-1 Time Series: Potential and Limitations. *Remote Sens.* **2021**, *13*, 348. [CrossRef]
18. Reinermann, S.; Gessner, U.; Asam, S.; Ullmann, T.; Schucknecht, A.; Kuenzer, C. Detection of Grassland Mowing Events for Germany by Combining Sentinel-1 and Sentinel-2 Time Series. *Remote Sens.* **2022**, *14*, 1647. [CrossRef]
19. Garioud, A.; Giordano, S.; Valero, S.; Mallet, C. Challenges in Grassland Mowing Event Detection with Multimodal Sentinel Images. In Proceedings of the 2019 10th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp), Shanghai, China, 5–7 August 2019; pp. 1–4.
20. Lobert, F.; Holtgrave, A.-K.; Schwieder, M.; Pause, M.; Vogt, J.; Gocht, A.; Erasmi, S. Mowing Event Detection in Permanent Grasslands: Systematic Evaluation of Input Features from Sentinel-1, Sentinel-2, and Landsat 8 Time Series. *Remote Sens. Environ.* **2021**, *267*, 112751. [CrossRef]
21. Holtgrave, A.-K.; Lobert, F.; Erasmi, S.; Röder, N.; Kleinschmit, B. Grassland Mowing Event Detection Using Combined Optical, SAR, and Weather Time Series. *Remote Sens. Environ.* **2023**, *295*, 113680. [CrossRef]
22. Komisarenko, V.; Voormansik, K.; Elshawi, R.; Sakr, S. Exploiting Time Series of Sentinel-1 and Sentinel-2 to Detect Grassland Mowing Events Using Deep Learning with Reject Region. *Sci. Rep.* **2022**, *12*, 983. [CrossRef]
23. De Vroey, M.; de Vendictis, L.; Zavagli, M.; Bontemps, S.; Heymans, D.; Radoux, J.; Koetz, B.; Defourny, P. Mowing Detection Using Sentinel-1 and Sentinel-2 Time Series for Large Scale Grassland Monitoring. *Remote Sens. Environ.* **2022**, *280*, 113145. [CrossRef]
24. Wang, Z.; Ma, Y.; Zhang, Y.; Shang, J. Review of Remote Sensing Applications in Grassland Monitoring. *Remote Sens.* **2022**, *14*, 2903. [CrossRef]
25. Tiscornia, G.; Baethgen, W.; Ruggia, A.; Do Carmo, M.; Ceccato, P. Can We Monitor Height of Native Grasslands in Uruguay with Earth Observation? *Remote Sens.* **2019**, *11*, 1801. [CrossRef]
26. Cimbelli, A.; Vitale, V. Grassland Height Assessment by Satellite Images. *Adv. Remote Sens.* **2017**, *6*, 40–53. [CrossRef]
27. Nickmilder, C.; Tedde, A.; Dufrasne, I.; Lessire, F.; Tychon, B.; Curnel, Y.; Bindelle, J.; Soyeurt, H. Development of Machine Learning Models to Predict Compressed Sward Height in Walloon Pastures Based on Sentinel-1, Sentinel-2 and Meteorological Data Using Multiple Data Transformations. *Remote Sens.* **2021**, *13*, 408. [CrossRef]
28. Wrage, N.; Şahin Demirbağ, N.; Hofmann, M.; Isselstein, J. Vegetation Height of Patch More Important for Phytodiversity than That of Paddock. *Agric. Ecosyst. Environ.* **2012**, *155*, 111–116. [CrossRef]
29. Nickmilder, C.; Tedde, A.; Dufrasne, I.; Lessire, F.; Glesner, N.; Tychon, B.; Bindelle, J.; Soyeurt, H. Creation of a Walloon Pasture Monitoring Platform Based on Machine Learning Models and Remote Sensing. *Remote Sens.* **2023**, *15*, 1890. [CrossRef]
30. Obermeyer, K.; Komainda, M.; Kayser, M.; Isselstein, J. Exploring the Potential of Rising Plate Meter Techniques to Analyse Ecosystem Services from Multi-Species Grasslands. *Crop Pasture Sci.* **2022**, *74*, 378–391. [CrossRef]
31. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2020.
32. Posit Posit. Available online: https://www.posit.co/ (accessed on 20 July 2023).
33. van Rossum, G. *Python Reference Manual*; Department of Computer Science [CS]; CWI: Amsterdam, The Netherlands, 1995.
34. Spyder Welcome to Spyder's Documentation—Spyder 5 Documentation. Available online: https://docs.spyder-ide.org/current/index.html (accessed on 20 July 2023).
35. OS Os—Miscellaneous Operating System Interfaces. Available online: https://docs.python.org/3/library/os.html (accessed on 20 July 2023).
36. Hijmans, R.J.; van Etten, J.; Sumner, M.; Cheng, J.; Baston, D.; Bevan, A.; Bivand, R.; Busetto, L.; Canty, M.; Fasoli, B.; et al. Raster: Geographic Data Analysis and Modeling. Available online: https://cran.r-project.org/web/packages/raster/ (accessed on 20 July 2023).
37. RE 6.2. Re—Regular Expression Operations—Python 3.6.15 Documentation. Available online: https://docs.python.org/3.6/library/re.html (accessed on 20 July 2023).
38. Pebesma, E.; Bivand, R.; Racine, E.; Sumner, M.; Cook, I.; Keitt, T.; Lovelace, R.; Wickham, H.; Ooms, J.; Müller, K.; et al. Sf: Simple Features for R. Available online: https://cran.r-project.org/web/packages/sf/ (accessed on 20 July 2023).
39. Pandas Pandas Documentation—Pandas 2.0.3 Documentation. Available online: https://pandas.pydata.org/docs/index.html (accessed on 20 July 2023).
40. Dowle, M.; Srinivasan, A.; Gorecki, J.; Chirico, M.; Stetsenko, P.; Short, T.; Lianoglou, S.; Antonyan, E.; Bonsch, M.; Parsonage, H.; et al. Data.Table: Extension of "Data.Frame". Available online: https://cran.r-project.org/web/packages/data.table/ (accessed on 20 July 2023).
41. Harris, C.R.; Millman, K.J.; van der Walt, S.J.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N.J.; et al. Array Programming with NumPy. *Nature* **2020**, *585*, 357–362. [CrossRef]
42. Kuhn, M.; Wing, J.; Weston, S.; Williams, A.; Keefer, C.; Engelhardt, A.; Cooper, T.; Mayer, Z.; Kenkel, B. Caret: Classification and Regression Training. Available online: https://cran.r-project.org/web/packages/caret/ (accessed on 20 July 2023).
43. Dask Dask—Dask Documentation. Available online: https://docs.dask.org/en/stable/ (accessed on 20 July 2023).

44. Scikit-Learn Scikit-Learn: Machine Learning in Python—Scikit-Learn 1.3.0 Documentation. Available online: https://scikit-learn.org/stable/index.html (accessed on 20 July 2023).
45. ESA Open Access Hub. Available online: https://scihub.copernicus.eu/ (accessed on 16 January 2024).
46. ESA Copernicus Data Space Ecosystem | Europe's Eyes on Earth. Available online: https://dataspace.copernicus.eu/ (accessed on 16 January 2024).
47. CRA-W Agromet.Be. Available online: https://agromet.be/fr/pages/home/ (accessed on 16 January 2024).
48. Liu, F.T.; Ting, K.M.; Zhou, Z.-H. Isolation Forest. In Proceedings of the 2008 Eighth IEEE International Conference on Data Mining, Pisa, Italy, 15–19 December 2008; pp. 413–422.
49. Forton, F.; Meniger, G. Cours Sur Les Prairies. 2019. Available online: https://www.fourragesmieux.be/Documents_telechargeables/ (accessed on 13 March 2023).
50. [Chiffres-clé] Quelles Sont les Céréales et Cultures Cultivées en Wallonie? *Celagri* **2019**. Available online: https://www.celagri.be/quelles-cereales-et-cultures-cultive-t-on-en-wallonie/ (accessed on 20 July 2023).
51. Herman-Saffar, O. An Approach for Choosing Number of Clusters for K-Means. Available online: https://towardsdatascience.com/an-approach-for-choosing-number-of-clusters-for-k-means-c28e614ecb2c (accessed on 20 July 2023).
52. Grelet, C.; Larsen, T.; Crowe, M.A.; Wathes, D.C.; Ferris, C.P.; Ingvartsen, K.L.; Marchitelli, C.; Becker, F.; Vanlierde, A.; Leblois, J.; et al. Prediction of Key Milk Biomarkers in Dairy Cows through Milk MIR Spectra and International Collaborations. *J. Dairy Sci.* **2023**, *in press*. [CrossRef] [PubMed]
53. Vogel, A.; Scherer-Lorenzen, M.; Weigelt, A. Grassland Resistance and Resilience after Drought Depends on Management Intensity and Species Richness. *PLoS ONE* **2012**, *7*, e36992. [CrossRef] [PubMed]
54. Moinet, G.Y.K.; Midwood, A.J.; Hunt, J.E.; Rumpel, C.; Millard, P.; Chabbi, A. Grassland Management Influences the Response of Soil Respiration to Drought. *Agronomy* **2019**, *9*, 124. [CrossRef]