Monitoring of grassland productivity using Sentinel-2 remote sensing data

Pauline Dusseux, Thomas Guyet, Pierre Pattier, Valentin Barbier, Hervé Nicolas

Monitoring of grassland productivity using Sentinel-2 remote sensing data

Pauline Dusseux[a], Thomas Guyet[b], Pierre Pattier[c], Valentin Barbier[d], Hervé Nicolas[c]

[a] *University Grenoble Alpes, Pacte-UMR 5194 CNRS / Institut Agro, France*
[b] *Inria, Centre de Lyon, France*
[c] *Institut Agro, SAS-UMR 1069 CNRS, France*
[d] *Institut Agro, ENSAI, France*

Abstract

Grasslands are a source of goods and ecosystem services. It would therefore be helpful to monitor grass growth and estimate grass productivity indicators in order to optimize grassland management over time. Until today, farmers have had to cope with a lack of regular assessments of grass availability over time across the whole farm. In order to simplify and automate grass measurements, we propose to develop methods for estimating grassland biomass using remote sensing.

The aim of this study is to assess the ability of Sentinel-2 remotely sensed data to estimate grassland height as measurements in order to provide farmers with information on the quantity of grass available per agricultural plot. We propose a generic data-driven methodology to identify 1) the set of features derived from Sentinel-2 remote sensing images and 2) a regression technique, in order to yield the best performances in estimating grassland height. Before selecting a subset of features, we generated 1,935 partly new but potentially meaningful features derived from the spectral indices available.

The study was conducted between 2017 and 2020 on 18 farms located in France. The model has been tested and evaluated using the data from 2017 to 2019. The average RMSE (resp. R^2) is 1.78 ± 0.30 *cm* (resp. 0.70 ± 0.12) on the test set. The RMSE is lower than 10 percent of the range width of the predicted values, indicating a very good assessment of grassland height and this is consistent with the precision required for the grassland management support service.

The model has also been evaluated on the data from 2020. The correlation between measurements and estimations is encouraging with $R^2 = 0.56$ and RMSE = 2.1 *cm*. The majority of the differences are between -1 *cm* and 2 *cm* which are relevant according to grassland management.

Keywords: Data science, Regression, Feature engineering, Satellite images, Agriculture, Vegetation

1. Introduction

Grasslands, including natural and sown pasture, rangeland and fodder crops, are one of the largest ecosystems of the world, covering approximately 25% of the earth's terrestrial surface and 70% of the Earth's agricultural area (Suttie et al., 2005; White et al., 2000). Grasslands are a source of goods and ecosystem services: many millions of people depend on grasslands as they are an important feed source for livestock contributing to their livelihoods and food security (Lemaire et al., 2011). Moreover, they are important for environmental protection as they provide carbon storage and contribute to biodiversity and wildlife habitats, the regulation of water, nitrogen and pollutant flows, and to water, air and soil quality (Peeters, 2009;

Soussana and Lüscher, 2007). However, grassland management and pressures are having an impact on the functions performed by grasslands and may lead to a reduction in their productivity and the ecosystem services they provide (Suttie et al., 2005). Moreover, climate change, with increasing temperatures and changes in precipitation amounts and yearly distribution, may have consequences on grassland productivity. Indeed, a reduction in the amount of grass available can be observed, particularly in the summer period, which can lead to difficulties in feeding livestock and thus to economic losses (Soussana, 2013).

A grass-based production system could be helpful to protect the environment and it is the cheapest feed for livestock in farming systems. Yet grass productivity and quality, and thus farming production, are dependent on grassland management (and climatic and soil conditions) (Lemaire et al., 2011). It would therefore be helpful to monitor grass growth and estimate grass productivity through indicators in order to optimize grassland management over time. In particular, this information could help prioritize fields for grazing and/or identify mowing to provide winter fodder (Seuret et al., 2014; Lemaire et al., 2005). As grass growth can be fast depending on the season and as rotational grazing systems can be practiced, regular measurements of biomass and/or grass height are required to optimise the grasslands productivity. However, a lot of time would be required to collect this information on all the fields of a farm, even visually by a simple estimation of the grass height, or with simple and non-destructive assessment equipment such as a plate meter (Herbomètre®) which measures the height and density of the grass (Welter and Le Bris, 1992).

To contribute for maintaining or even developing pastureland, a simplification and automation of grass measurements is necessary. Thus, methods for estimating grassland biomass using remote sensing acquisition tools will be developed and operational feasibility assessed in this paper.

Grassland management varies in time and space according to the needs of the farm operators, the type of grass, soil and climate conditions. Moreover, grasslands are very diverse, with a wide distribution in quite different situations (high or low, sloping or flat, dry or wet, natural or sown). So, to characterize and study them at different scales from a local to a regional scale, high spatial, temporal and spectral resolution remote sensing data are required to implement a weekly grass biomass estimation model and develop an innovative and operational management decision-support service (Kumar and Mutanga, 2017; Pottier et al., 2017; Grant et al., 2015).

Recent major progress in satellite remote sensing, in terms of precision (spatial and spectral), revisit frequency and access to data, as the Sentinel missions from the European Space Agency[1], offer interesting opportunities for precision farming and monitoring grassland vegetation biomass at a regional scale (Weiss et al., 2020; Reinermann et al., 2020).

In the agricultural context, crops are regularly and precisely inventoried and monitored with remote sensing data and with well-established decision-support tools (Bégué et al., 2018). Grassland and pastures in agricultural production systems are still less studied mainly because grass production is not linear during the growing season: rather it is a dynamic system with

[1] `https://sentinel.esa.int/web/sentinel/home`

several growth periods dependent on grassland management that are highly diversified and changing throughout the season.

Several studies show the benefits of remote sensing data for grassland monitoring (Ali et al., 2016; Zhang and Guo, 2008). Mainly, the normalized difference vegetation index (NDVI) is used to estimate pasture biomass and grass growth rate (Edirisinghe et al., 2012; Hill et al., 2004). Vegetation indices improve the spectral characteristics sensitive to plant properties while reducing the disturbances by combining certain spectral bands into an index (Glenn et al., 2008; Clevers, 2014). Most of the conventional vegetation indices used in remote sensing are formed by the combination of two spectral bands (the red visible and near-infrared), as is the case for the NDVI index (Rouse et al., 1973). However, the common problem of these indices, based on the near-infrared spectral band, is the rapid saturation as the estimated or measured magnitudes increase, together with their insensitivity to changes over dense vegetation (Edirisinghe et al., 2011; Glenn et al., 2008; Gitelson et al., 2002). Research on hyperspectral data was conducted to identify alternative vegetation indices to overcome the saturation effect and to ensure better estimates and monitoring of photosynthetically active and senescent vegetation. Narrow bands in the red-edge and infrared domains have been identified as interesting to study vegetation (Mutanga et al., 2012; Gitelson et al., 2006; Mutanga and Skidmore, 2004a,b; Le Maire et al., 2004). Thus, some of these alternative vegetation indices for vegetation biomass assessment can be evaluated using Sentinel-2 data as it presents these spectral bands and they are yet under-used (Verrelst et al., 2015b; Delegido et al., 2011).

Therefore, the use of statistics based on vegetation indices is one of the most used and simplest methods to estimate biomass (Verrelst et al., 2015a). However, the accuracy of the estimation, based on these indices, depends strongly on the choice of the index formula and the spectral bands selected (Rivera et al., 2014). In order to determine spectral bands and index formulas that perform well, Verrelst et al. (2015b) propose to calculate all possible band combinations according to different index formulas. Here, indices constructed from two or three spectral bands have been proposed to limit saturation problems (Verrelst et al., 2015a) and evaluate all Sentinel-2 bands (Verrelst et al., 2015b; Delegido et al., 2011; Wanga et al., 2013). Finding the Sentinel-2 bands thereof that are the most useful to accurately assess grassland biomass was thus an issue addressed in these studies.

The aim of this study is to assess the ability of Sentinel-2 remotely sensed data to estimate dry grassland biomass (using grassland height as measurements) in order to provide users (agricultural organizations, farmers) with information on the quantity of grass available per agricultural plot on a weekly basis. In this context, an operational spatial service for precision agricultural applications will be developed. This research is part of the CASDAR HERDECT project, funded by the French Ministry of Agriculture, Agri-Food and Forestry.

In this paper, we present the work that led us to propose a grassland height assessment model. This model is based on a small-sized set of features derived from Sentinel-2 images and a regression model. One of the challenges addressed in the article is to set up a model that would yield accurate predictions. More specifically, we need both to identify the best set

of features and to select a suitable regression technique. We assume that the richness of the Sentinel-2 spectral bands (especially the Red-edge, Near Infra-Red and Short Wave Infra-Red bands) will allow for the correct estimation of grass height from standard regression methods.

The contributions of this article are therefore twofold:

- the proposal of a generic data-driven methodology to identify 1) the set of features derived from remote sensing images and 2) a regression technique, in order to yield the best performances in a regression task.
- the application of this methodology to grassland height estimation.

In the following, section 2 presents the data that have been collected and our data science processing chain to build a model. Section 3 presents results of the models learned and an application of one model on a new dataset.

2. Material and methods

2.1. Study sites and field data collection

The study was conducted between 2017 and 2020 on 18 farms located in France. The farms are located in different parts of France (Figure 1), with different climates and soils, in order to cover different types and uses of grassland (grazing, mowing, suckling cattle, dairy cattle, sheep, different grazing severities, several botanical compositions and ages of the grasslands).

A plate meter (Herbomètre®), a non-destructive technique for the estimation of grassland biomass, was used due to its facility and reasonable accuracy (Seuret et al., 2014; Welter and Le Bris, 1992). The plate meter has a tray and a measuring ruler (Figure 1). To estimate the biomass of the

grassland, the tray is lowered into the grass along the ruler in the center. At the point where the tray is fully retained by the grass, the height of the grass can be measured on the ruler.

As direct measurement of biomass is very costly in terms of sampling and processing time, the measurement of grass height, based on the plate meter, was used as a reference to estimate the biomass of grasslands.

Each week between February and October, measurements were conducted in all the fields studied in the project to record grass height and density in order to obtain a more accurate estimation of dry matter yield. Thus, more than $2,000$ measurements were acquired during the project. Some of these measurements will ultimately not be used because no cloud-free satellite images will have been acquired at or close to the date. In total, 738 ground measurements were thus used for the modelling in this study (acquired during 2017, 2018 and 2019, on 103 fields spread between 9 farms) and 473 measurements, acquired during 2020 (on 85 fields spread over the initial 9 farms and 9 other farms), were used for the evaluation of the model.

Figure 2 shows the histogram of grass height measurements for the years 2017 and 2018 (data used for the modelling). We have less data for 2018 than for 2017 due to the less favorable conditions (less cloud-free image). This explains the differences in level between the two histograms, but we observe that the two distributions are similar. The values range from 2 to 19.5 cm. The overall mean height is 8.26 ± 3.23 cm. The distribution fits a Gamma distribution well ($\Gamma(k = 0.79, \theta = 6.56)$). The quality of fit is a good indicator of the dataset quality.

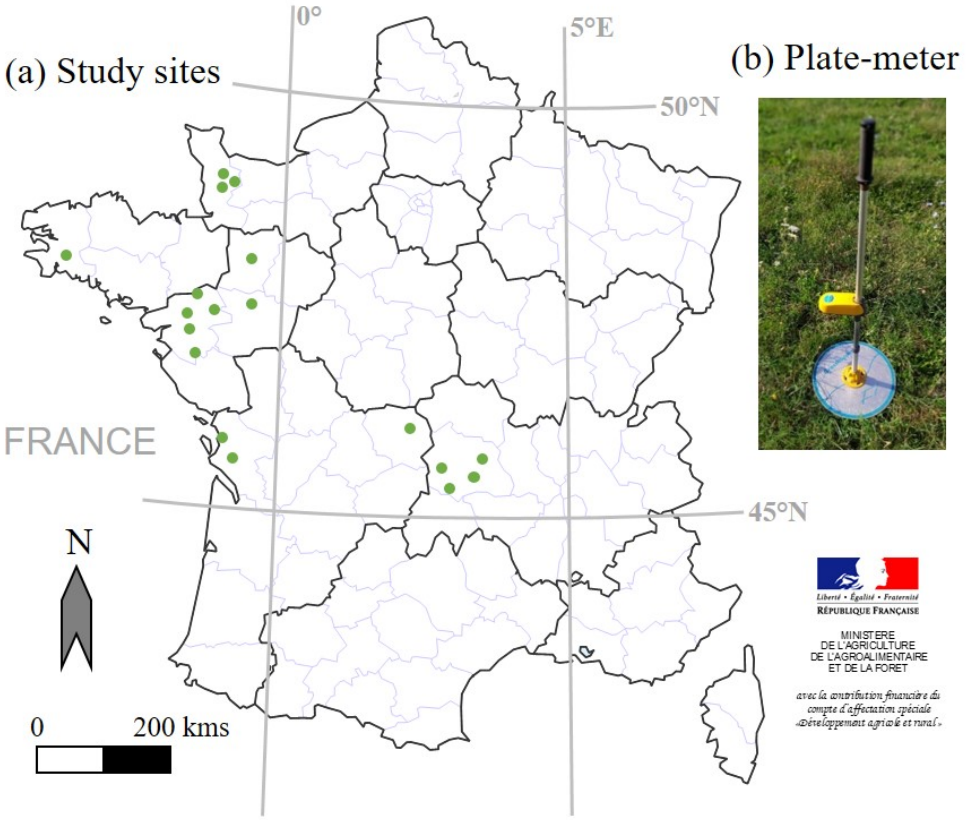


Figure 1: Study sites (a) and illustration of the plate meter used in this study to estimate on-site grassland biomass (b). Each green dot locates a farm where measurements were conducted over several fields.

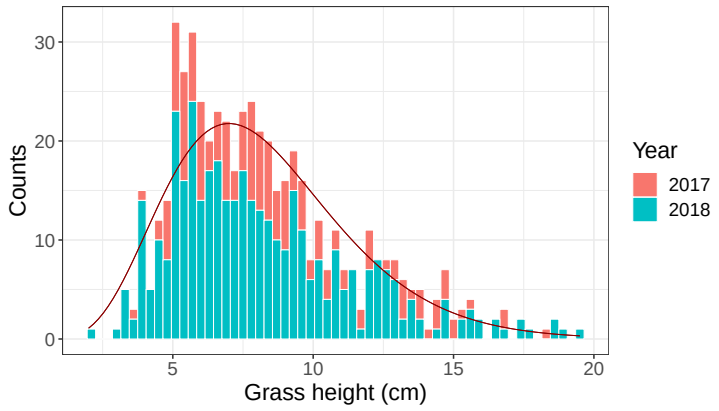


Figure 2: Histogram representing the grass height distribution in the dataset (for the years 2017 and 2018).

2.2. Sentinel-2 remote sensing data

For this study, all Sentinel-2 images were acquired between the years 2017 and 2020 during the grass growing season (between March and November). Sentinel-2 images were downloaded from the Theia website[2]. These Sentinel-2 images are composed of the three classic visible bands and a Near Infra-Red (NIR) band (B2, B3, B4, and B8, respectively) at a 10-meter spatial resolution and the six spectral bands at 20 meters, from the Red-edge (Re) and NIR spectral domains (B5, B6, B7 and B8a, respectively) to the Short Wave Infra-Red (SWIR) spectral domain (B11 and B12) (see Appendix A).

First, the 20-meter bands of each image were subsampled to a 10-meter spatial resolution using the nearest neighbor approach. Cloud and cloud shadow masking was then applied on each image.

From each cloud-free image, spectral bands were gathered at the pixel

[2] `https://www.theia-land.fr/`

scale. Pixel values, within each survey plot, were then averaged in order to characterize grassland productivity at plot scale. Thus, each data point in our dataset corresponds to a plot at a date for which we have both a cloud-free image and a ground measurement.

The average difference between an image and its ground measurement is 2 days over the entire study period (with a maximum of 9 days in February at the very beginning of the growth period) and 1.5 days during the period of rapid grass growth (with a maximum of 4 days).

2.3. Feature engineering

In this work, we investigated how to automatically identify a small subset of features suitable for a prediction task, e.g. the prediction of the quantified vegetation productivity. Before selecting a subset of features, we generated numerous potentially meaningful features derived from the spectral indices available.

Firstly, 13 indices from the literature were calculated (Table 1). Some of them, the $NDVI$ (or $NDVI2$) and the $GNDVI$, are widely known and commonly used for their relation with biomass (Gitelson et al., 2006, 2002). Moreover, as several studies show the important relationship between the Red-Edge reflectance and chlorophyll and nutrients present in the plant cell structure, other indices ($NDI45$, $IRECI$, $NREDI1$, $NREDI2$, $NREDI3$, $PSRI$, $MTCI$ and $MCARI$), using the Red-edge spectral domains, were tested (Frampton et al., 2013; Delegido et al., 2011; Le Maire et al., 2004). Finally, two indices, the $NDII$ and the MSI, used for their relation with vegetation water content and canopy water thickness were added (Yilmaz et al., 2008; Fang et al., 2017).

Index *	Formula and Sentinel-2 bands used	Original author
$NDVI$	$(R_{NIR} - R_R)/(R_{NIR} + R_R)$	Rouse et al. (1973)
	$(R_{B8} - R_{B4})/(R_{B8} + R_{B4})$	
$NDVI2$	$(R_{NIR} - R_R)/(R_{NIR} + R_R)$	
	$(R_{B8a} - R_{B4})/(R_{B8a} + R_{B4})$	
$GNDVI$	$(R_{NIR} - R_G)/(R_{NIR} + R_G)$	Gitelson et al. (1996)
	$(R_{B8} - R_{B3})/(R_{B8} + R_{B3})$	
$NDI45$	$(R_{Re} - R_R)/(R_{Re} + R_R)$	Delegido et al. (2011)
	$(R_{B5} - R_{B4})/(R_{B5} + R_{B4})$	
$IRECI$	$(R_{Re3} - R_R)/(R_{Re1}/R_{Re2})$	Frampton et al. (2013)
	$(R_{B7} - R_{B4})/(R_{B5}/R_{B6})$	
$NREDI1$	$(R_{Re2} - R_{Re1})/(R_{Re2} + R_{Re1})$	Gitelson and Merzlyak (1994)
	$(R_{B6} - R_{B5})/(R_{B6} + R_{B5})$	
$NREDI2$	$(R_{Re3} - R_{Re1})/(R_{Re3} + R_{Re1})$	
	$(R_{B7} - R_{B5})/(R_{B7} + R_{B5})$	
$NREDI3$	$(R_{Re3} - R_{Re2})/(R_{Re3} + R_{Re2})$	
	$(R_{B7} - R_{B6})/(R_{B7} + R_{B6})$	
$PSRI$	$(R_R - R_G)/R_{Re1}$	Merzlyak et al. (1999)
	$(R_{B4} - R_{B3})/R_{B5}$	
$MTCI$	$(R_{NIR} - R_{Re})/(R_{Re} - R_R)$	Dash and Curran (2004)
	$(R_{B8} - R_{B5})/(R_{B5} - R_{B4})$	
$MCARI$	$[(R_{Re} - R_R) - 0.2(R_{Re} - R_G)] * (R_{Re} - R_R)$	Wu et al. (2009)
	$[(R_{B5} - R_{B4}) - 0.2(R_{B5} - R_{B3})] * (R_{B5} - R_{B4})$	
$NDII$	$(R_{NIR} - R_{SWIR})/(R_{NIR} + R_{SWIR})$	Hardisky et al. (1983)
	$(R_{B8} - R_{B11})/(R_{B8} + R_{B11})$	
MSI	R_{SWIR}/R_{NIR}	Hunt and Rock (1989)
	R_{B11}/R_{B8}	

Index * : $NDVI$ = Normalized Difference Vegetation Index, $GNDVI$ = Green Normalized Difference Vegetation Index, $NDI45$ = Normalized Difference Index, $IRECI$ = Inverted Red-Edge Chlorophyll Index, $NREDI$ = Normalized Red-edge Index, $PSRI$ = Plant Senescence Reflectance Index, $MTCI$ = Meris Terrestrial Chlorophyll Index, $MCARI$ = Modified Chlorophyll Absorption in Reflectance Index, $NDII$ = Normalized Difference Infrared Index, MSI = Moisture Stress Index

Table 1: Vegetation indices calculated from Sentinel-2 images (R_* = Sentinel-2 reflectance, $_G$ = Green, $_R$ = Red, $_{Re}$ = Red-edge, $_{NIR}$ = Near Infra-Red, $_{SWIR}$ = Short Wave Infra-Red) ; more information on the spectral bands in the table A.5

Despite all these indices, some spectral bands are rarely or never tested. Also, some spectral bands are never used together. By testing and evaluating all possible band combinations according to formulas of common indices, it will become feasible to identify optimal band combinations among all available Sentinel-2 bands (Verrelst et al., 2015b; Thenkabail et al., 2000). Thus, a large number of indices (or features) was generated based on two-band and three-band common types of published indices (Verrelst et al., 2015b; Xue and Su, 2017; Le Maire et al., 2004) as described in Table 2. We denote by $(b_i)_{i\in 1..10}$ the spectral image bands. Two two-band indices were calculated: the Normalized Difference (ND) index and the Simple Ratio (SR) index, and four three-bands indices: the modified SR (mSR) index, the three-Band Spectral Index ($3BSI$), the three-Band Spectral Index Tian ($3BSITian$), and the Chlorophyll Vegetation Index (CVI). These indices were more or less correlated and influenced by canopy leaf area index, vegetation coverage or canopy leaf nitrogen concentration.

The symmetries in the feature formula may generate linearly correlated variables (for instance, $ND(b_i, b_j) = -ND(b_j, b_i)$ for all $i \in [1, 10]$ and $j \in [1, 10]$). The generation process prevents these redundant features from being generated. Table 2 also indicates the number of features generated by each formula (after removal of redundant features), totaling $1,935$ features.

Finally, each ground measurement, *i.e.* a plot and a date, has its associated cloud-free image and associated features. It composes the dataset we analyzed for this study. The dataset is composed of $1,211$ points (a ground measurement associated with a cloud-free and usable image – 738 points for

Combination	Formula	Number of combinations
$ND(b_i, b_j)$	$\frac{b_i - b_j}{b_i + b_j}$	$\frac{1}{2} \times \frac{10!}{(10-2)!} = 45$
$SR(b_i, b_j)$	$\frac{b_i}{b_j}$	$\frac{10!}{(10-2)!} = 90$
$mSR(b_i, b_j, b_k)$	$\frac{b_i - b_k}{b_j - b_k}$	$\frac{1}{2} \times \frac{10!}{(10-3)!} = 360$
$3BSI(b_i, b_j, b_k)$	$\frac{b_i - b_k}{b_j + b_k}$	$\frac{10!}{(10-3)!} = 720$
$3BSITian(b_i, b_j, b_k)$	$\frac{b_i - b_j - b_k}{b_i + b_j + b_k}$	$\frac{1}{2} \times \frac{10!}{(10-3)!} = 360$
$CVI(b_i, b_j, b_k)$	$\frac{1}{2} \times \frac{b_i \times b_j}{b_k^2}$	$\frac{10!}{(10-3)!} = 360$

Table 2: Formulas to generate features from spectral bands. b_i, b_j and b_k are three different spectral bands among the 10 Sentinel-2 bands used. Factors $\frac{1}{2}$ indicate symmetry breaking to prevent linearly correlated features.

the model construction and 473 points for the application). Each row of the dataset, *i.e.* a plot and a date, is made of one ground measurement of grassland height and $1,958$ features computed by the average indices for all valid pixels of the plot (10 bands, 13 common indices and $1,935$ band combinations).

2.4. Model construction

This section presents the proposed methodology to fit a frugal but accurate regression model from large remote sensing features. In the above section, we describe how to create a large number of potentially interesting features, but a model using all these features does not match the needs for routine exploitation. We need a model requiring few features to be computed from images to propose a space- and time-efficient grassland height assessment service.

The model we propose is therefore both a subset of features and a regression model built upon the selected features. The remainder of this section presents the proposed methodology to learn such a model. Note that this methodology does not rely on specific tools or implementation. The implementation details are given at the beginning of Section 3.

2.4.1. Overall methodology

The input of this process is the dataset described in previous section. It is split into two subsets: a training set and a test set. The main steps of the methodology are illustrated in Figure 3 and are detailed in the following:

1. All descriptive features (spectral bands, common indices and band combinations) are z-normalized using the training data set. The target feature (height of grassland) is not normalized.
2. Feature selection identifies different feature sets. The feature selection process is repeated several times on different subsets of the training set and the best model is selected.
3. Different regression models are fitted from the dataset with selected features. Each model is trained using a leave-one-out procedure. The model with the lowest mean prediction error is the selected model.
4. Fitted models (normalization model, feature selection model and prediction model) are evaluated on the test set.

The prediction error is evaluated by two measurements: the root mean squared error (RMSE) and the coefficient of determination (R^2). The whole process (steps 1 to 4) is repeated ten times to evaluate the whole learning chain (surrounded with dashed line in Figure 3) on random training sets.

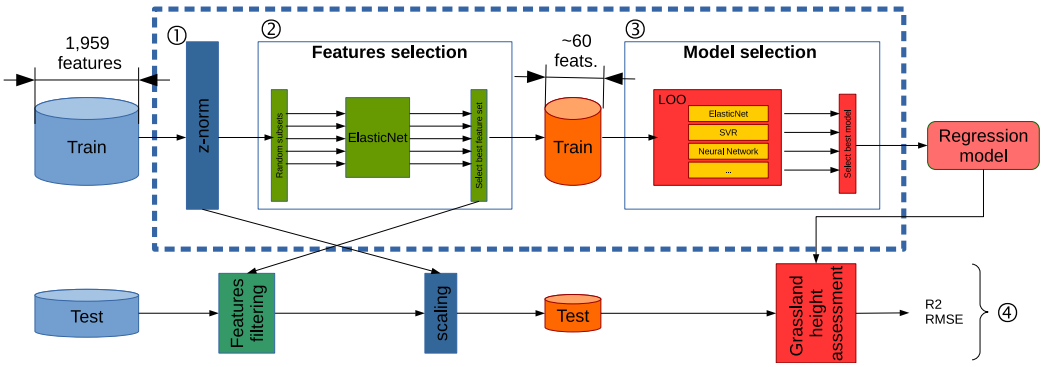


Figure 3: Overall methodology for regression model learning and evaluation. The dashed square illustrates the training phase to fit the regression model in two main steps. The overall model is composed of: a selection of features (green square), the z-normalization coefficients (blue square) and the regression model (red square). Rounded figures match the step numbers detailed in the text.

The final model combines three elements: the scaling coefficients of the z-normalization, the subset of features and the regression model. It is applied on the test set in three steps: select the features of the test set, apply the scaling for each feature and finally, apply the trained regression model. Features are filtered before scaling to prevent scaling features that are not used in the regression model.

2.4.2. Feature set selection

Too large a set of features can negatively impact the prediction performances of a model and lead to a lack of interpretability to qualitatively evaluate the model (Guyon and Elisseeff, 2003; Blum and Langley, 1997). In addition, a large number of features would lead to larger resource requirements for our service. Our objective is therefore to select a small subset of features that leads to accurate target prediction. Feature selection is the pro-

cess that automatically selects the optimal subset of features for the machine learning task.

For a regression task, there are three types of methodology for feature selection (Brownlee, 2020):

- Filter method: select a subset of features based on the relationship with the target. Statistical test (F-test or Mutual Information) ranks features by the measured correlation with the target variable. Then, it selects the desired number of the most correlated variables as a feature subset. The major issue with this method is that the features are selected independently.

- Wrapper method: search for well-performing subset of features. The recursive feature elimination recursively selects a feature to eliminate until the desired number of features is reached. The feature that is eliminated is the one for which the model trained without this feature is better. Thus, it takes into account the remaining features to select which one to discard. The progressive selection of features may miss subsets of features that are interesting together but not necessarily at the beginning of the elimination process.

- Intrinsic: perform automatic feature selection during training. For instance, penalized regression models (*e.g.*, Ridge, Lasso (Tibshirani, 1996) or ElasticNet (Zou and Hastie, 2005)) have been proposed to train a model with a sparsity constraint. A sparsity constraint favors models with small (Ridge) or zero (Lasso) feature weights. Zero weights designate features to discard. Contrary to the previous methods, the

number of desired features can not be directly specified.

We use ElasticNet (Zou and Hastie, 2005) to select features. Through optimization strategies, this technique enables us to identify cohesive subsets of features for optimizing regression performances. Such a cohesive subset of features is compatible with our objective of using this subset in a regression model thereafter. ElasticNet is a penalized linear regression method. It combines ℓ_1 and ℓ_2 penalties for learning a sparse model in which the number of non-zero weights can be controlled by the balance between penalties. The objective function to minimize is:

$$\min_w \frac{1}{2n}||Xw - y||_2^2 + \alpha\lambda||w||_1 + \frac{\alpha(1-\lambda)}{2}||w||_2^2$$

where n is the number of samples in the dataset, w is the weights of the linear regression model, X is the dataset, y is the height of grassland and α, λ are the parameters of ElasticNet.

ElasticNet parameters are fitted in a two-step process:

1. the value of λ is set: ElasticNet is run on the training set with 10 values of $\lambda \in [0, 1]$ and $\alpha \in [0, 1]$ chosen by cross-validation. The model with the closest number of features to the desired one designates the λ value for the next step.
2. ElasticNet, set up with the λ value, is run several times with random subsets of 66% of the training set. Then, the error of each regression model is evaluated on the remaining part of the training set. The evaluation measurement is the coefficient of determination of each ElasticNet regression model computed on the test set.

The feature selection process may be sensitive to the training dataset. This justifies the cross-validation setting of feature selection. The evaluation of the selected features on a test set makes the results generalizable on new data. The repetition of the process reduces the dependency on the input dataset. The overall feature selection is thus more stable and reliable.

2.4.3. Regression model learning

At this stage of the process, a regression model is learned from a training dataset with only the selected features. We assume that the use of ElasticNet for feature selection is not necessarily the best approach to obtain an accurate regression model. Furthermore, there are plenty of different regression techniques that suit our problem. This stage of the process experimentally selects the best one.

In this work, we compare the following regression techniques which have been selected for their diversity:

- Linear regression: ElasticNet, Ridge, Lasso. The parameters of these models are selected by cross-validation.

- Bayesian regression: BayesianRidge (BR). BR is a linear model. It assumes that the output is Gaussian distributed around the predicted value. The learning procedure is parameter-free.

- Non-linear regression: we experiment with support vector regression (SVR) with a polynomial kernel of degree 2. Three penalty parameters C are evaluated: 10^{-3}, 10^{-2} and 10^{-1}.

- Neural networks regression: a conventional neural network is also tested.

The architecture of the neural network is made of three fully connected layers with a ReLU activation function. The size of the input layer is the number of selected features, the hidden layer is of size 60 and the output layer is of size 1.

3. Experiments and results

This section presents the results we obtain by applying the method proposed in the previous section to the dataset of $1,959$ features. The objective is to identify a regression model based on a few features which can accurately estimate the ground height of grassland. The desired number of features is around 60. This number has been estimated to be both potentially representative of the diversity of the whole set of features and low enough to ensure the memory and time-efficiency of a large scale operational service.

The experiments were conducted on a desktop computer. Our implementation is in Python and uses the *sklearn* library (Pedregosa et al., 2011) for machine learning algorithms and *Keras/Tensor flow* (Chollet et al., 2015) for neural networks.[3] For the sake of reproducibility, we manually set up the random seed for the experiments presented in this section. With this setting, the overall process takes about 30 minutes to select the best feature sets and models. More details about requiring computing resources are provided in Appendix B. We use the `R` software to conduct statistical analysis.

[3]Note that our methodological framework may be reimplemented and enhanced with other data science tools, e.g. in `R`.

Feature	Best bands	R^2	RMSE
Spectral bands b_i	B7	0.45	2.45
Common indices	$NREDI3$	0.54	2.23
$ND(b_i, b_j)$	b_i: B6, b_j: B7	0.54	2.23
$SR(b_i, b_j)$	b_i: B6, b_j: B7	0.54	2.23
$mSR(b_i, b_j, b_k)$	b_i: B8a, b_j: B3, b_k: B11	0.54	2.25
$3BSI(b_i, b_j, b_k)$	b_i: B7, b_j: B11, b_k: B6	0.58	2.13
$3BSITian(b_i, b_j, b_k)$	b_i: B7, b_j: B6, b_k: B8	0.49	2.35
$CVI(b_i, b_j, b_k)$	b_i: B7, b_j: B8a, b_k: B11	0.56	2.19

Table 3: Cross-validation results for the best performing feature per type and formulation

3.1. Correlation of grass height with individual spectral bands, common indices or band combinations

Before presenting the experiments about our methodology, we provide some insights about the informativeness of generated features. In order to evaluate the correlation between features and the targeted height of grass, a regression was applied on all available S2 spectral bands (Table A.5), computed common indices (Table 1) and S2 band combinations (Table 2). Because of the large number of features (10 spectral bands, 13 common indices, $1,935$ band combinations) only the best performing feature per type and formulation is listed in Table 3 but several features per type and formulation may perform almost equally well.

Table 3 shows that band combinations perform slightly better than common indices and spectral bands but R^2 values are generally not very high (below 0.6). The selected spectral bands are encountered as well in other optimized features. The Red-edge B7 ($Re3$) spectral band has been selected in 5 of the 6 band combinations. The Red-edge B6 ($Re2$) spectral band has been selected in 4 of the 6 band combinations. Also, the Short Wave Infra-Red band B11 ($SWIR1$) has been selected in 3 of the 4 three-band combinations. The scatter plot of estimated versus measured grass height values using the best performing feature is shown in Figure 4. Saturation and underestimation can be observed at high grass height values, as well as a rather large dispersion around the regression line explaining the relatively low R^2 (0.58).

This result highlights the limited ability of the two- and three-band indices individually to precisely estimate vegetation cover.

From the spectral signatures (Figure 5), organised by grass height classes, we can see an increase in the Red-edge and Near Infra-Red spectral band values and a decrease in the Red spectral band values with increasing biomass. However, the distinctions can sometimes be quite small depending on the measurement points. Thus, mobilising a set of indices could allow a better estimation of the height of the grass. Finally, we can see a good distinction between plots with heights below 9 cm and plots with heights above 12 cm. However, within these two groups, similarities can be observed. Also we can see quite clearly the saturation of the infrared values as the grass height increases.

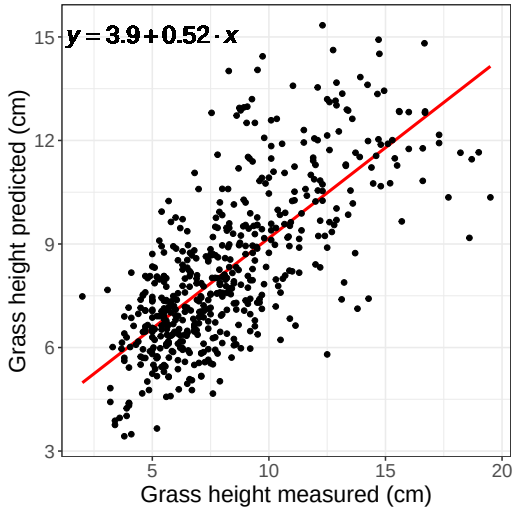


Figure 4: Measured versus estimated grass height values of the best performing feature 3BSI (B7, B11, B6)

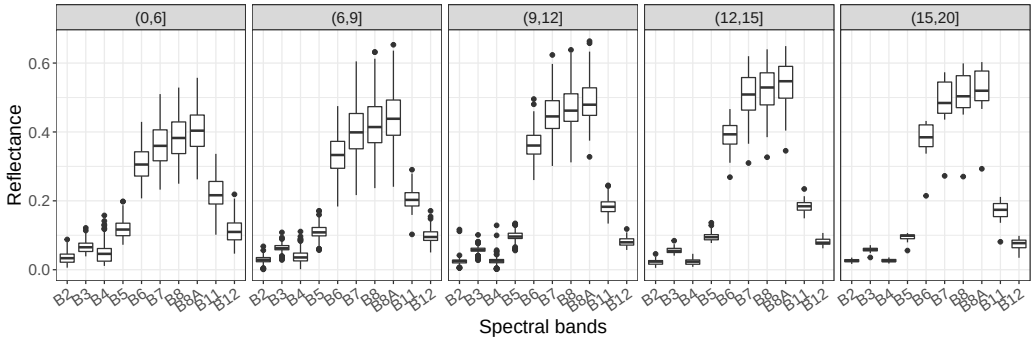


Figure 5: Spectral signatures with respect to height of grass classes, from left to right: $(0, 6]$ cm, $(6, 9]$ cm, $(9, 12]$ cm, $(12, 15]$ cm and $(15, 20]$ cm.

3.2. Regression based on subsets of features

3.2.1. Feature set analysis

175 feature sets were generated by applying ElasticNet to random subsets of the training set and were tested on the remaining examples (35 runs of the process with 5 repetitions of feature selection). For one run of ElasticNet, 66% of the training set is used to select features. The remaining examples are used to evaluate the regression model obtained by ElasticNet.

To facilitate the reproducibility of the method and to limit the calculation time and storage, a limited number of indices is desired. Several tests have been performed by requesting around 10, 30, 60, 100 and 200 features (i.e. 1 to 10% of the initially calculated feature set). On average, 50 to 60 indices are selected by requesting 10 to 100 features and then a hundred for a request of around 200. Moreover, the estimation errors are stable despite the increase in the number of features considered. Thus, the desired number of features is set at around 60 for this study and the average number of selected features is 58.75 ± 10.3.

If the features were randomly drawn from the whole feature set, then they would have $58.75/1,658$ chances of being selected. This corresponds to an average number of selections of 6.20 times. This means that features that have been selected at most 6 times occur only marginally in the feature selections. This represents $1,740$ features (including $1,501$ features that have never been selected). The 60 most selected features have been selected in at least 34.2% of the feature sets.

Figure 6, on the left, gives the distribution of numbers of features with respect to their number of selections. It illustrates that the majority of the

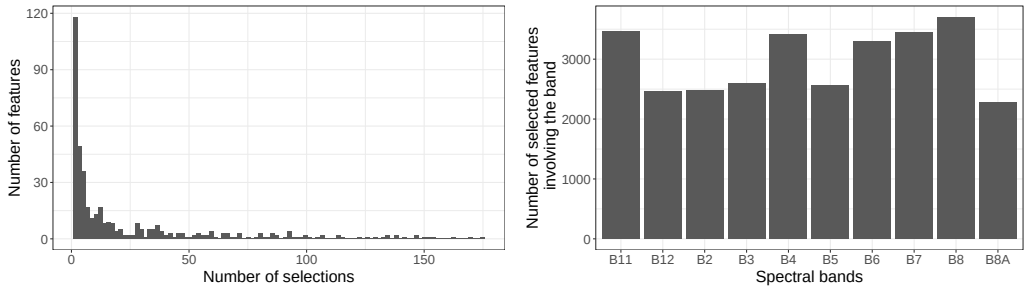


Figure 6: On the left, distribution of 421 features w.r.t. the number of selections. On the right, average number of involvements of a spectral band in selected features.

features are selected less than 6 times and that few features are selected more than 120 times. This indicates that there is actually a core set of features that are more likely to be selected, but the majority of the features changes a lot with the subset of examples. Nonetheless, among the initial set of $1,958$ features, the set of features effectively used (421) is relatively small: about four fifths of the features are discarded.

The two features that appear the most are $CVI\_B3\_B8\_B11$ (174 times) and $CVI\_B3\_B7\_B11$ (170 times). A total of 421 features has been selected at least once.

Figure 6, on the right, illustrates the average number of times a spectral band is used in selected features. The total number of band selections is above the average number of features because each generated feature involves 2 or 3 different spectral bands.

This figure shows that bands $B4$, $B6$, $B7$, $B8$ and $B11$ are the most used. $B6$ and $B7$ are new spectral bands for Sentinel-2. They correspond to the red-edge and are closely related to vegetation density. These bands are very interesting for accurately discriminating grassland biomass. $B12$

corresponds to short-wave infrared, a wavelength closely related to the water content present in the canopy. $B4$ and $B8$ are classic bands on satellite and correspond to the red and the near-infrared bands, respectively, of the visible spectrum. They are widely used to discriminate vegetation from the ground, thus they are also closely linked to the vegetation density.

Then, the subset of features with the lowest RMSE on its test set is selected for the regression model selection phase. The average RMSE (resp. R^2) is 1.61 ± 0.09 cm (resp. 0.74 ± 0.02). These values will be compared to the RMSE and R^2 obtained by the regression models trained in the next step of the process.

3.2.2. Analysis of regression model errors

Figure 7, at the top, illustrates the RMSE and R^2 for the six regression methods for one run of the model regression selection (step 3 of the processing chain). Table 4 gives mean values for RMSE (in cm) and the determination coefficients (R^2).

We first notice that linear regression models outperform non-linear models (Neural Networks and SVR with polynomial kernel). This can be explained by the feature selection step that is based on a linear regression model. However, our three-layer Neural Network is surprisingly close to linear models and much better than a two-layer model. Moreover, the four linear regression models perform equally well. Both RMSE and R^2 are very close for the test set. In our experiments, Bayesian Ridge is more often selected as the best model. SVR has very poor performances. It is probably due to limited modelling capability of our SVR. In our experiments, we used polynomial kernels of degree 2 (and we evaluate different values of the C parameter).

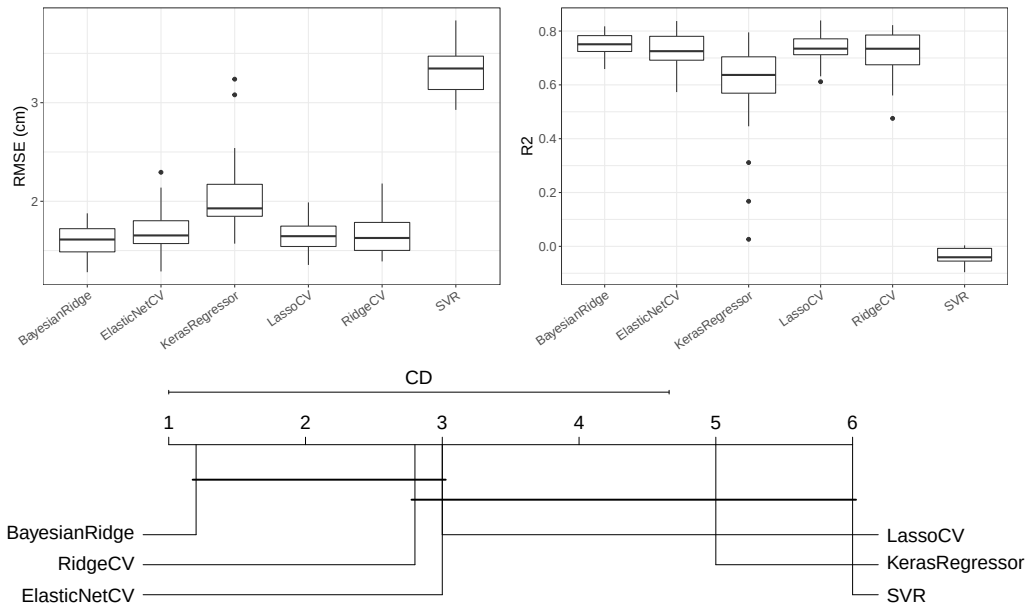


Figure 7: Top-left: Boxplots of RMSE on the test sets for best regression model (the lower, the better). On the top-right: Boxplots of determination coefficients (R^2) on the test sets for best regression model (the higher, the better). Bottom: Critical difference diagram comparing the different regression models.

This means that the function is locally approximated by a polynomial function of 2nd degree. The modelling capability seems to not fit the data. Note that alternative kernels and setting may be investigated, but SVR is long to train and exploring all the possible setting seems not reasonable. For these different reasons, we prefer to not tune finely SVR, even if a better kernel choice may leads to better performances.

To confirm these results, Figure 7 shows a critical difference diagram (Demšar, 2006). A critical difference diagram represents the mean ranks of the methods that have been obtained for each prediction of the test sets. The lower the rank, the better the method. In addition, the representation shows hori-

Table 4: Mean RMSE and R^2 on test set with respect to the regression model

method	RMSE (cm)	R^2
BayesianRidge	1.60 ± 0.16	0.75
ElasticNetCV	1.69 ± 0.21	0.73
KerasRegressor	2.03 ± 0.37	0.60
LassoCV	1.66 ± 0.15	0.73
RidgeCV	1.66 ± 0.20	0.72
SVR	3.33 ± 0.23	-0.03

zontal bars that group some methods. Within the same group, the methods are not statistically different according to the Nemenyi test. Figure 7 at the bottom shows that Bayesian Ridge has a better rank, but the four linear models are not significantly different. On the contrary, the SVR and Neural Network are significantly different.

The average RMSE (1.60 cm) and R^2 (0.75) values are comparable to those obtained by feature selection. These values are reasonably low, indicating a good assessment of grassland height. As a reminder, the measurements were between 2 cm and 19.5 cm. Thus, the RMSE is lower than 10 percent of the range width of the predicted values. In addition, our R^2 value is much higher than for the model developed by Cimbelli and Vitale (2017) ($R^2_{Cimbelli} = 0.63$).

Figure 8 illustrates regression errors (predicted value vs true value) and residuals (difference between the predicted and true value with respect to the true value) for four regression models. The residual indicates that the model

has difficulties in accurately estimating high grassland height values. This can be explained by the lack of examples for such values but also by possibly less accurate measurements. Plate meters are known to be less accurate when grass is high.

Note that the scaling factor and selected features were computed from the dataset that was used to evaluate the RMSE of Figure 7. The next section presents an unbiased estimation of the RMSE of the best model on a separate test set.

3.2.3. Comparisons with and without feature selection

Figure 9 shows the RMSE and R^2 achieved by the same regression models as before, but trained on a dataset with the whole set of features.

3.2.4. Evaluation of the learning procedure

Finally, each of the 35 models (scaling, best feature set and best regression model) is applied to the test set. This experiment evaluates our overall methodology in terms of its ability to build a model that generalizes the grassland assessment with new images or new grasslands. The difference with the previous setting is that in the previous setting the test set of the regression model is not independent of the feature selection step.

The RMSE on the test set is 1.78 ± 0.30 *cm* and the R^2 is 0.70 ± 0.12 (see Figure 10). These average values are close to the values obtained on the training set (Figure 7) showing that our processing chain built a model that generalizes the training examples well. Compared to the approach of Cimbelli et *al.*, with $R^2 = 0.63$, our training procedure extracts linear regression models with lower prediction errors while keeping the number of features

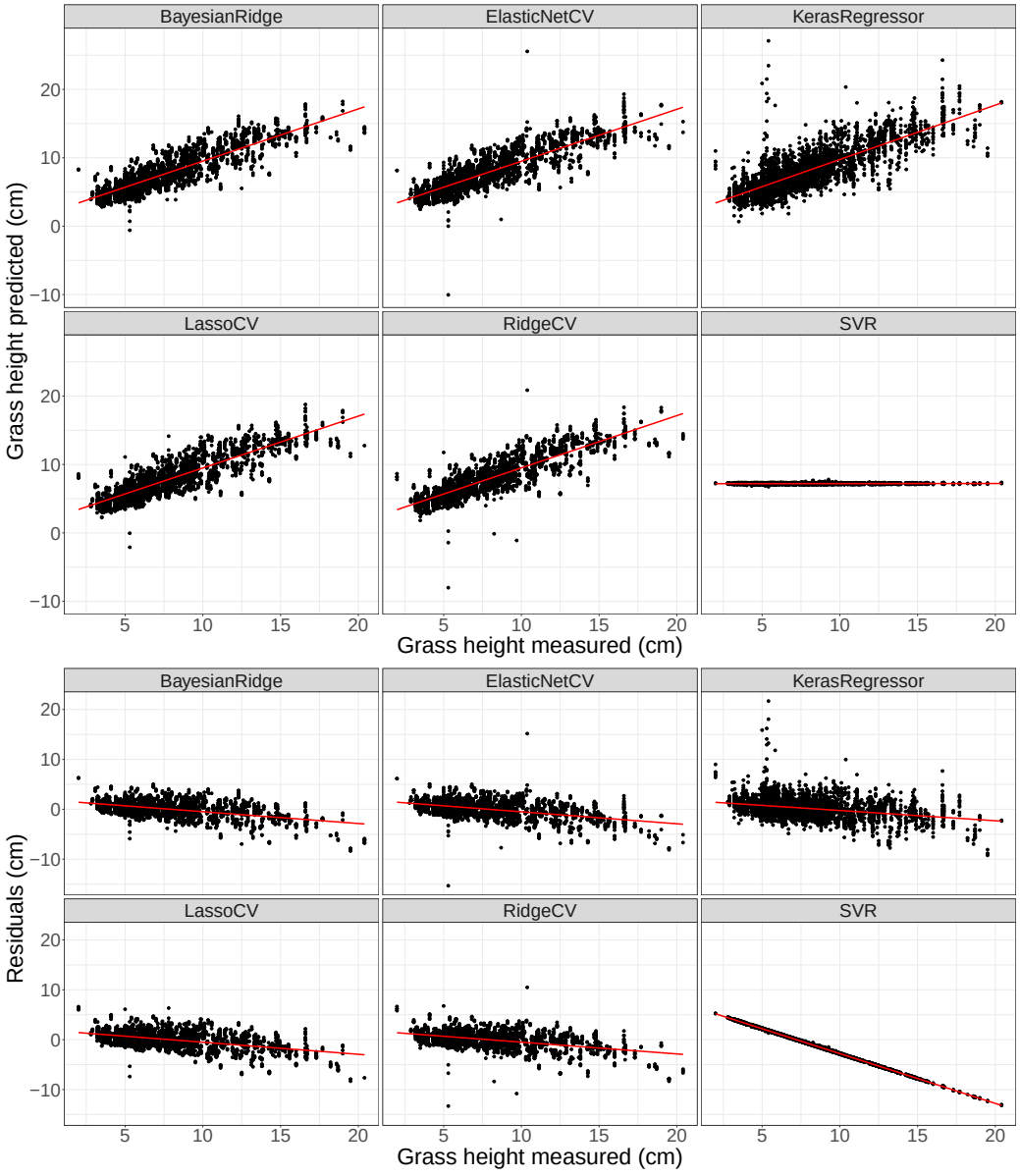


Figure 8: From left to right: ElasticNet, Ridge, SVR and Neural Network. Top: prediction vs actual value. Bottom: residual.

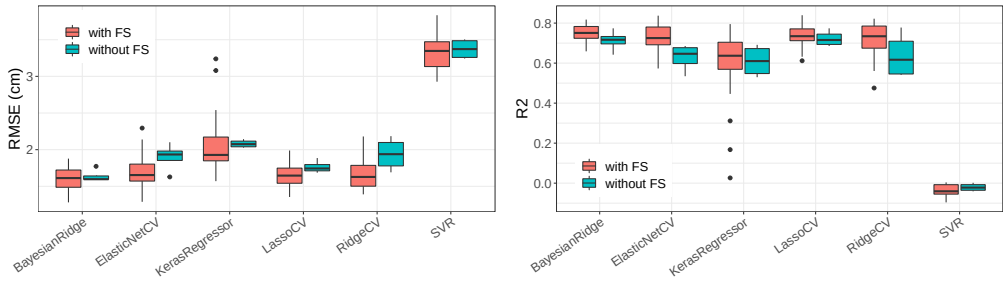


Figure 9: RMSE (on the left) and R^2 (on the right) for best regression model selection with feature selection, in red, and without feature selection, in blue.

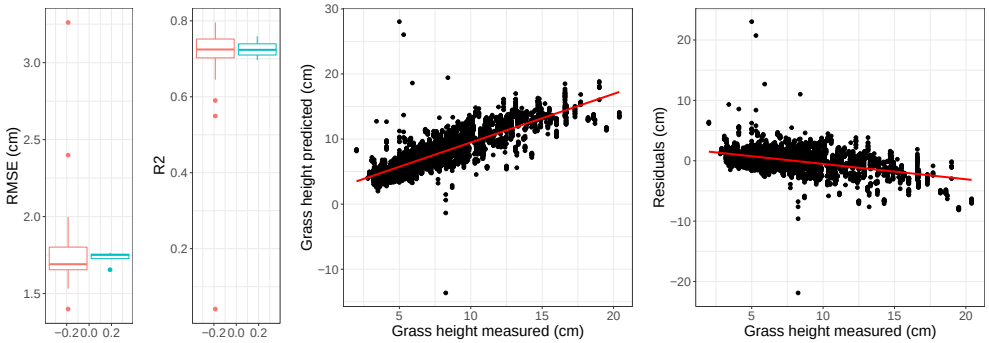


Figure 10: From left to right: Boxplots of RMSE in cm, R^2 computed on test set, prediction error and residual on test set. The boxplots on the left give the RMSE (resp. R^2) with feature selection in red on the left and without feature selection in blue on the right.

low.

Figure 10 on the left compares the RMSE and R^2 obtained by our pipeline with or without feature selection. Without feature selection, the RMSE is 1.73 ± 0.05 cm and the R^2 is 0.72 ± 0.02. This shows that a model trained on selected features is slightly more accurate on average than a model trained on the complete set of features (although not significantly). It is interesting to note that Bayesian Ridge is also the most selected regressor by the process without feature selection. As a reminder, our objective is to propose a grass-

land height assessment model that is computationally light to be embedded in a practical pipeline. The less image pre-processing, the lighter the model will be. This comparison shows that a small subset of features is sufficient to achieve the same accuracy as a model requiring a large number of features.

4. Application

This section presents the results of an external evaluation: it applies the model trained on the 2017 and 2018 data to new data collected in 2020.

The approach for the operational mapping of grassland productivity was applied to the 18 farms for the year 2020. Note that our objective is not to forecast the height of grassland, but to estimate grassland biomass from Sentinel-2 images in order to develop an operational spatial service for precision agricultural applications. Grass heights estimated from satellite images were compared to field data in order to test the robustness of the grass growth estimation model. Note that 2020 field data were not integrated in the model construction. At the scale of the 18 farms monitored in 2020, 473 estimates of grass height from satellite images were coupled with heights measured in the field (only points with a delta of $\pm$ 3 days between the measurement and the estimation have been kept here). The data was organized by height class following discussions with the farmers. Indeed, for grassland management, only height class is required and not the precise height. We observe that the estimation model based on satellite images tends to underestimate grass heights (Figure 11a). 40% of the estimates show a difference of between -1 and 1 centimeter with the ground measurements. More than 80% of these points belong to the two main height classes of 6 to 9 cm and 9 to 12 cm,

which are key heights for optimal grazing management. 10% of the estimates show a difference between 1 and 2 cm and 9% of the points show a difference greater than 2 cm, while 20% of the points show a difference between -1 and -2 cm and 21% of the points show a difference greater than -2 cm.

The correlation between measurements and estimations is encouraging with $R^2 = 0.56$ and RMSE = 2.1 cm (Figure 11b). The overall mean signed error between the estimates and the measurements is -0.56 cm which confirms the underestimation of the heights by satellite (Figure 11c) compared to the plate meter estimates. For low grass heights (< 6 cm) we observe a slight overestimation by satellite ($+0.64$ cm on average). For heights between 6 and 9 cm, we observe a very slight underestimation on average (-0.16 cm) as well as for heights between 9 and 12 cm (-0.76 cm). The underestimation increases slightly for heights between 12 and 15 cm (-1.16 cm on average) while it increases significantly to -3.14 cm for grass heights above 15 cm (Figure 11d).

There are several possible explanations for these differences: firstly, the reliability of the model and secondly, the accuracy and representativeness of the height measurements taken with the plate meter. As can be seen in Figure 11, the greater the grass height, the more the satellite seems to underestimate it. However, beyond $14 - 15$ cm measured with the plate meter, the comparison of estimated and measured heights must be qualified because the reliability of the plate meter measurements is lower. This trend of height underestimation using drone images was also highlighted in the publication of Surrault et al. (2018). Moreover, as illustrated in Figure 11, the $6 - 9$ cm and $9 - 12$ cm height classes are well represented while the

$12 - 15\ cm$ and $> 15\ cm$ height classes include little data. Two Gaussian curves are observed for the $6 - 9\ cm$ and $9 - 12\ cm$ classes, centered on a difference of between $-1; 1\ cm$.

The majority of the differences are between $-1\ cm$ and $2\ cm$, which indicates a certain relevance of the model. These estimates will be produced annually and sent to partner farmers. This will allow us to better quantify the errors and possibly refine the model after several years of estimations.

5. Discussion & Perspectives

Several prospects for improvement can be envisaged, notably on the basis of the work of Verrelst et al. (2015a). We have developed a method based on parametric regression methods and vegetation indices. Other non-parametric regression methods or methods based on physical models can be used.

Machine learning algorithms can be cited for non-parametric regression algorithms. Presented by Verrelst et al. (2015a), one of the advantages here is that they use the spectral bands directly without calculating as many indices. It would be interesting to test this procedure, as it would appear to be less time- and storage-consuming and provide interesting results. It shows that most machine learning algorithms give better results than with parametric regression based on vegetation indices (maximum R^2 of 0.9 *vs.* 0.8 respectively). Also, compared to vegetation index results, it seems that non-parametric regression methods are more stable with lower standard deviation results than parametric regression methods. Finally, this study presents similar results to ours concerning the most relevant spectral bands: the authors noted that the red-edge and the SWIR bands are the most prevalent in the

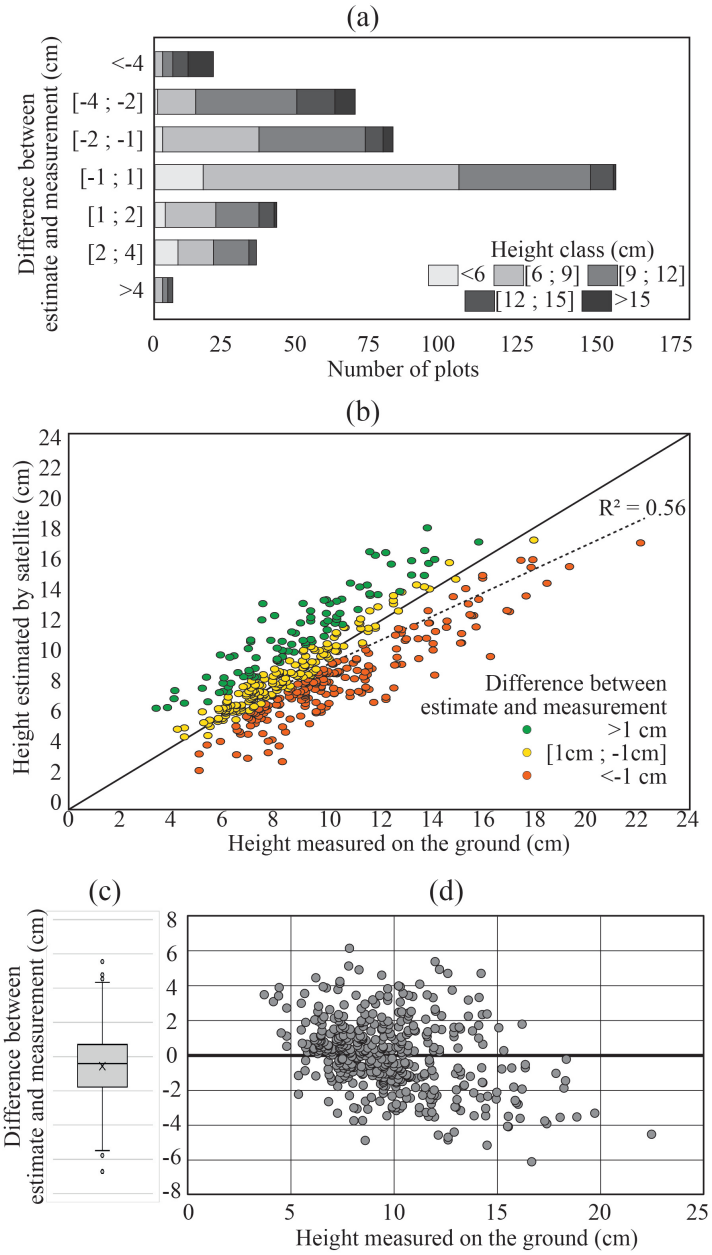


Figure 11: Difference between estimated and measured heights according to height classes (a), correlation between heights measured in the field and estimated by satellite (b), box-plot (c) and cloud point (d) of the difference between estimates and measurements.

different models they were able to establish. This is in agreement with our results and other previous studies (Delegido et al., 2011; Rivera et al., 2014).

Also, several studies have shown that biophysical variables seem to be interesting for monitoring grasslands (Reinermann et al., 2020; Dusseux et al., 2015; Asam et al., 2013; Darvishzadeh et al., 2008). For example, Punalekar et al. (2018) showed the ability of Sentinel-2 images to better capture grassland dynamics and biomass from LAI (Leaf Area Index) than from NDVI. They also showed that, for the estimation of biophysical variables, the use of a physical model like a radiative transfer model is more efficient than using an empirical NDVI-based approach.

However, the application and implementation of these models can be complicated. Also, Sentinel-2 and the Red-edge and Short Wave Infra-Red spectral bands are not yet widely used for grassland studies (Reinermann et al., 2020), so here we wanted to explore their richness and ability to estimate grass height. Finally, the desire to make the methodology and the model re-applicable led us to stick to standard regression methods. We show that with the richness of the sentinel-2 spectral bands, standard methods can be used to obtain valuable results for grassland management.

6. Conclusion

This study demonstrated that Sentinel-2 images provide quantitative information of the biomass status in grasslands. The results presented relatively accurate estimates of grass biomass. The Red-edge, Near Infra-Red and Short Wave Infra-Red spectral bands seem to be rich in information for the estimation of grassland biomass.

The assessment of grassland height with satellite images aims at providing access from the desktop and day by day to the grassland biomass/height assessment for each grassland plot or, in the future, monitoring and forecasting the growth of grasslands to determine optimal cutting dates.

Beyond these new results on the estimate of height of grass, we also presented a generic methodology that can be applied to the estimation of any agro-ecological quantities from Sentinel-2 images. The classical approach in remote sensing uses a small collection of standard or advanced vegetation indices. This study points out that automatically generated features may lead to more accurate models. Our framework addresses the problem of finding a best subset of features and it is end-to-end. The user only has to select the desired number of features and it selects the best feature sets and the best regression model. It will be used in the future on a wide range of applications.

7. Acknowledgements and funding

The authors thank all technical partners (chamber of agriculture, experimental farm ...) for their rigorous contribution to the measurements and field expertise, the French Ministry of Agriculture for financial support and farmers who volunteered to participate in the project.

This research is part of the CASDAR HERDECT project, funded by the French Ministry of Agriculture, Agri-Food and Forestry. Planned over three years, the HERDECT project involves the use of remote sensing to estimate grass biomass in order to improve grazing management on livestock farms. This project relies on highly involved technical partners in the field

of grassland advisory services, experimental sites and scientific structures working with the remote sensing techniques that will allow the elaborated methods to be parameterized. Thanks to the diversity and plurality of the project partners, a large ground data set with a variety of observations to grasslands has been built up.

References

Ali, I., Cawkwell, F., Dwyer, E., Barrett, B., Green, S., 2016. Satellite remote sensing of grasslands: from observation to management – A review. Journal of Plant Ecology , rtw005.

Asam, S., Fabritius, H., Klein, D., Conrad, C., Dech, S., 2013. Derivation of leaf area index for grassland within alpine upland using multi-temporal RapidEye data. International Journal of Remote Sensing 34, 8628–8652.

Blum, A.L., Langley, P., 1997. Selection of relevant features and examples in machine learning. Artificial intelligence 97, 245–271.

Brownlee, J., 2020. Data preparation for machine learning: data cleaning, feature selection, and data transforms in Python. Machine Learning Mastery.

Bégué, A., Arvor, D., Bellon, B., Betbeder, J., De Abelleyra, D., Ferraz, R.P.D., Lebourgeois, V., Lelong, C., Simões, M., Verón, S.R., 2018. Remote sensing and cropping practices: A review. Remote Sensing 10, 99.

Chollet, F., et al., 2015. Keras. `https://keras.io`.

Cimbelli, A., Vitale, V., 2017. Grassland height assessment by satellite images. Advances in Remote Sensing 6, 40.

Clevers, J.G., 2014. Beyond NDVI: Extraction of biophysical variables from remote sensing imagery – a book, in: Land Use and Land Cover Mapping in Europe, pp. 363–381.

Darvishzadeh, R., Skidmore, A., Schlerf, M., Atzberger, C., 2008. Inversion of a radiative transfer model for estimating vegetation LAI and chlorophyll in a heterogeneous grassland. Remote Sensing of Environment 112, 2592–2604.

Dash, J., Curran, P.J., 2004. The MERIS terrestrial chlorophyll index. International Journal of Remote Sensing 25, 5403–5413.

Delegido, J., Verrelst, J., Alonso, L., Moreno, J., 2011. Evaluation of Sentinel-2 red-edge bands for empirical estimation of green LAI and chlorophyll content. Sensors 11, 7063–7081.

Demšar, J., 2006. Statistical comparisons of classifiers over multiple data sets. Journal of Machine Learning Research 7, 1–30.

Dusseux, P., Hubert-Moy, L., Corpetti, T., Vertès, F., 2015. Evaluation of SPOT imagery for the estimation of grassland biomass. International Journal of Applied Earth Observation and Geoinformation 38, 72–77.

Edirisinghe, A., Clark, D., Waugh, D., 2012. Spatio-temporal modelling of biomass of intensively grazed perennial dairy pastures using multispectral remote sensing. International Journal of Applied Earth Observation and Geoinformation 16, 5–16.

731 Edirisinghe, A., Hill, M.J., Donald, G.E., Hyder, M., 2011. Quantitative
732 mapping of pasture biomass using satellite imagery. International Journal
733 of Remote Sensing 32, 2699–2724.

734 Fang, M., Ju, W., Zhan, W., Cheng, T., Qiu, F., Wang, J., 2017. A new
735 spectral similarity water index for the estimation of leaf water content from
736 hyperspectral data of leaves. Remote Sensing of Environment 196, 13–27.

737 Frampton, W.J., Dash, J., Watmough, G., Milton, E.J., 2013. Evaluating the
738 capabilities of sentinel-2 for quantitative estimation of biophysical variables
739 in vegetation. ISPRS Journal of Photogrammetry and Remote Sensing 82,
740 83–92.

741 Gitelson, A., Merzlyak, M.N., 1994. Quantitative estimation of chlorophyll-a
742 using reflectance spectra: Experiments with autumn chestnut and maple
743 leaves. Journal of Photochemistry and Photobiology B: Biology 22, 247–
744 252.

745 Gitelson, A.A., Kaufman, Y.J., Merzlyak, M.N., 1996. Use of a green channel
746 in remote sensing of global vegetation from EOS-MODIS. Remote Sensing
747 of Environment 58, 289–298.

748 Gitelson, A.A., Kaufman, Y.J., Stark, R., Rundquist, D., 2002. Novel al-
749 gorithms for remote estimation of vegetation fraction. Remote Sensing of
750 Environment 80, 76–87.

751 Gitelson, A.A., Viña, A., Verma, S.B., Rundquist, D.C., Arkebauer, T.J.,
752 Keydan, G., Leavitt, B., Ciganda, V., Burba, G.G., Suyker, A.E., 2006.
753 Relationship between gross primary production and chlorophyll content in

crops: Implications for the synoptic monitoring of vegetation productivity. Journal of Geophysical Research: Atmospheres 111.

Glenn, E.P., Huete, A.R., Nagler, P.L., Nelson, S.G., 2008. Relationship between remotely-sensed vegetation indices, canopy attributes and plant physiological processes: What vegetation indices can and cannot tell us about the landscape. Sensors 8, 2136–2160.

Grant, K., Siegmund, R., Wagner, M., Hartmann, S., 2015. Satellite-based assessment of grassland yields. The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 40, 15.

Guyon, I., Elisseeff, A., 2003. An introduction to variable and feature selection. Journal of machine learning research 3, 1157–1182.

Hardisky, M.A., Klemas, V., Smart, R.M., 1983. The influence of soil salinity, growth form, and leaf moisture on the spectral radiance of Spartina alterniflora canopies. Photogrammetric Engineering and Remote Sensing 49, 77–83.

Hill, M.J., Donald, G.E., Hyder, M.W., Smith, R.C., 2004. Estimation of pasture growth rate in the south west of western australia from AVHRR NDVI and climate data. Remote Sensing of Environment 93, 528–545.

Hunt, E.R., Rock, B.N., 1989. Detection of changes in leaf water content using near- and middle-infrared reflectances. Remote Sensing of Environment 30, 43–54.

Kumar, L., Mutanga, O., 2017. Remote sensing of above-ground biomass. Remote Sensing 9, 935.

Le Maire, G., François, C., Dufrêne, E., 2004. Towards universal broad leaf chlorophyll indices using PROSPECT simulated database and hyperspectral reflectance measurements. Remote Sensing of Environment 89, 1–28.

Lemaire, G., Hodgson, J., Chabbi, A., 2011. Grassland productivity and ecosystem services doi:10.1079/9781845938093.0000.

Lemaire, G., Wilkins, R., Hodgson, J., 2005. Challenges for grassland science: managing research priorities. Agriculture, Ecosystems and Environment 108, 99–108.

Merzlyak, M.N., Gitelson, A.A., Chivkunova, O.B., Rakitin, V.Y., 1999. Non-destructive optical detection of pigment changes during leaf senescence and fruit ripening. Physiologia Plantarum 106, 135–141.

Mutanga, O., Adam, E., Cho, M.A., 2012. High density biomass estimation for wetland vegetation using WorldView-2 imagery and random forest regression algorithm. International Journal of Applied Earth Observation and Geoinformation 18, 399–406.

Mutanga, O., Skidmore, A.K., 2004a. Hyperspectral band depth analysis for a better estimation of grass biomass (cenchrus ciliaris) measured under controlled laboratory conditions. International Journal of Applied Earth Observation and Geoinformation 5, 87–96.

Mutanga, O., Skidmore, A.K., 2004b. Narrow band vegetation indices overcome the saturation problem in biomass estimation. International Journal of Remote Sensing 25, 3999–4014.

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. Journal of Machine Learning Research 12, 2825–2830.

Peeters, A., 2009. Importance, evolution, environmental impact and future challenges of grasslands and grassland-based systems in europe. Grassland Science 55, 113–125.

Pottier, E., Jacquin, A., Roumiguie, A., Fougère, M., 2017. Improving grassland use via new technologies. Fourrages , 161–168.

Punalekar, S.M., Verhoef, A., Quaife, T.L., Humphries, D., Bermingham, L., Reynolds, C.K., 2018. Application of sentinel-2A data for pasture biomass monitoring using a physically based radiative transfer model. Remote Sensing of Environment 218, 207–220.

Reinermann, S., Asam, S., Kuenzer, C., 2020. Remote sensing of grassland production and management: A review. Remote Sensing 12, 1949.

Rivera, J.P., Verrelst, J., Delegido, J., Veroustraete, F., Moreno, J., 2014. On the semi-automatic retrieval of biophysical parameters based on spectral index optimization. Remote Sensing 6, 4927–4951.

Rouse, J.J.W., Haas, R.H., Schell, J., Deering, D., 1973. Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation. Technical Report. NASA.

Seuret, J.M., Theau, J.P., Pottier, E., Pelletier, P., Piquet, M., Delaby, L., 2014. Management tools for optimizing grassland utilization and boosting farmer confidence. Fourrages 218, 191–201. RMT "Prairies demain".

Soussana, J.F., 2013. S'adapter au changement climatique: Agriculture, écosystèmes et territoires. Editions Quae.

Soussana, J.F., Lüscher, A., 2007. Temperate grasslands and global atmospheric change: a review. Grass and Forage Science 62, 127–134.

Surrault, F., Barre, P., Escobar-Gutierrez, A., Roy, E., 2018. Le drone, un nouvel outil au service de la sélection pour estimer la hauteur des plantes fourragères. Fourrages .

Suttie, J., Reynolds, S., Batello, C., 2005. Grasslands of the World.

Thenkabail, P.S., Smith, R.B., De Pauw, E., 2000. Hyperspectral vegetation indices and their relationships with agricultural crop characteristics. Remote Sensing of Environment 71, 158–182.

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

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

Verrelst, J., Rivera, J.P., Veroustraete, F., Muñoz Marí, J., Clevers, J.G., Camps-Valls, G., Moreno, J., 2015b. Experimental Sentinel-2 LAI estima-

tion using parametric, non-parametric and physical retrieval methods – a comparison. ISPRS Journal of Photogrammetry and Remote Sensing 108, 260–272.

Wanga, L., Hunt, E.R., Qu, J.J., Hao, X., Daughtry, C.S.T., 2013. Remote sensing of fuel moisture content from ratios of narrow-band vegetation water and dry-matter indices. Remote Sensing of Environment 129, 103–110.

Weiss, M., Jacob, F., Duveiller, G., 2020. Remote sensing for agricultural applications: A meta-review. Remote Sensing of Environment 236.

Welter, M., Le Bris, X., 1992. Herbometer as a tool for assessment of herb mass of permanent pastures in the lorraine area (France). Fourrages .

White, R., Murray, S., M., R., 2000. Pilote Analysis of Global Ecosystems: Grassland ecosystems. World Resources Institute.

Wu, C., Niu, Z., Tang, Q., Huang, W., Rivard, B., Feng, J., 2009. Remote estimation of gross primary production in wheat using chlorophyll-related vegetation indices. Agricultural and Forest Meteorology 149, 1015–1021.

Xue, J., Su, B., 2017. Significant remote sensing vegetation indices: A review of developments and applications. Journal of sensors 2017.

Yilmaz, M.T., Hunt, E.R., Jackson, T.J., 2008. Remote sensing of vegetation water content from equivalent water thickness using satellite imagery. Remote Sensing of Environment 112, 2514–2522.

Zhang, C., Guo, X., 2008. Monitoring northern mixed prairie health using broadband satellite imagery. International Journal of Remote Sensing 29, 2257–2271.

Zou, H., Hastie, T., 2005. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Methodological) 67, 301–320.

Appendix A. Sentinel-2 remote sensind data

Satellite images from the Sentinel-2 mission were used. This mission, conducted by the European Space Agency and the European Commission, was launched in June 2015 with the aim of monitoring variability in land surface conditions. The Sentinel-2 mission is composed of a constellation of two satellites (Sentinel-2A and Sentinel-2B) with sun-synchronous orbit, 13 spectral bands with various spatial resolutions (10, 20 and 60 m) and a high revisit time (10 days with one satellite and 5 days with the two satellites) (Table A.5).

For this study, all Sentinel-2 images were acquired between the years 2017 and 2020 during the grass growing season (between March and November). Sentinel-2 images were downloaded from the Theia website[4], the French land data center, that provides Sentinel-2 images at level 2A (orthorectified product with top of canopy reflectance), with cloud and cloud shadow masking. These Sentinel-2 images are composed of the three classic visible bands and a Near Infra-Red (NIR) band (B2, B3, B4, and B8, respectively) at a 10-

[4] `https://www.theia-land.fr/`

Spectral bands	Band number	Central wavelength S2A / S2B (nm)	Bandwidth S2A / S2B (nm)	Spatial resolution (m)
Aerosols	B1	442.7 / 442.3	21	60
Water vapour	B9	945.1 / 943.2	20 / 21	
Cirrus detection	B10	1373.5 / 1376.9	31 / 30	
Blue (B)	B2	492.4 / 492.1	66	10
Green (G)	B3	559.8/ 559	36	
Red (R)	B4	664.6 / 665	31	
Near Infra-Red (NIR)	B8	832.8 / 833	106	
Red-edge (Re1)	B5	704.1 / 703.8	15 / 16	20
Red-edge (Re2)	B6	740.5 / 739.1	15	
Red-edge (Re3)	B7	782.8 / 779.7	20	
Near Infra-Red narrow (NIR_n)	B8a	864.7 / 864	21 / 22	
Short Wave Infra-Red (SWIR1)	B11	1613.7 / 1,610.4	91 / 94	
Short Wave Infra-Red (SWIR2)	B12	2,202.4 / 2,185.7	175 / 185	

Table A.5: Characteristics of Sentinel-2 images

meter spatial resolution and the six spectral bands at 20 meters, from the Red-edge (Re) and NIR spectral domains (B5, B6, B7 and B8a, respectively) to the Short Wave Infra-Red (SWIR) spectral domain (B11 and B12). The three spectral bands of the Red-edge (B5, B6 and B7) and the NIR narrow (B8a) are new information proposed by Sentinel-2. They are important and very interesting for the study of vegetation, especially for the retrieval and monitoring of biophysical parameters such as indicators of vegetation health, canopy structure and functioning parameters, and biomass estimation.

Appendix B. Computing resources

In this machine learning problem, the most critical computing resource is time. The framework demands a reduced amount of memory and memory requirements does no constrain it uses on a standard desktop computer. The long execution time is required due to the repetition of each step of the process. As illustrated in Figure 3, two steps are repeated: features selection and learning of each model. Remind that the overall objective if the process is to select the best combination of a features set and a regression model. To automatically select the best in the sound way, we need to repeat the experiment on several subsets of the data.

This particular process is time demanding but can easily be parallelized. The computing time depends on the number of step repetitions and the time of each step. On the other side, all step repetitions are independent from each other. It can be parallelized easily and with a high speedup.

In practice, we use 4 cores of our CPUs to parallelize the independent steps. The the mean computing times of each individual steps are detailed

Process step		Time (s)
Feature selection		≈ 290
Model training	SVR	≈ 1.27
-	Ridge	≈ 1.89
-	Lasso	≈ 3.36
-	Bayesian Ridge	≈ 1.90
-	ElasticNet	≈ 1.12

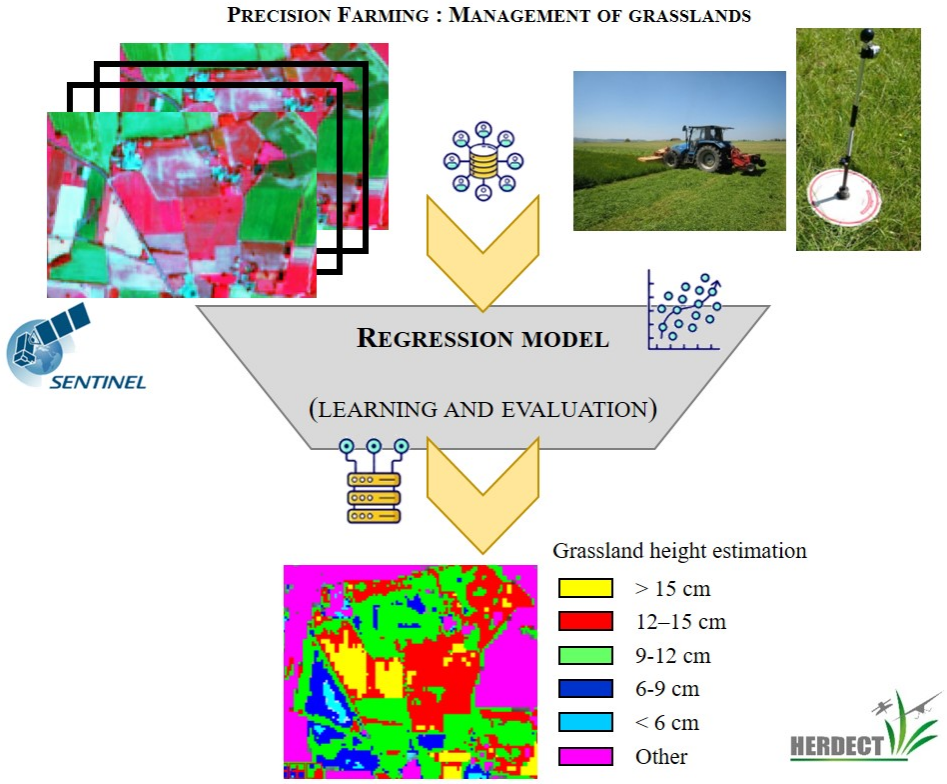
Table B.6: Time (in seconds) of the main processes of our framework. The time for model training correspond to 100 trainings and evaluations (Leave One Out evaluation).

in Table B.6. We repeat feature selection 5 times and we repeat 100 times the training of each model (Leave One Out on 100 random samples). With this setting, the overall running time (user time) is about 30 minutes.

Graphical Abstract

Monitoring of grassland productivity using Sentinel-2 remote sensing data

Pauline Dusseux, Thomas Guyet, Pierre Pattier, Valentin Barbier, Hervé Nicolas



Highlights

Monitoring of grassland productivity using Sentinel-2 remote sensing data

Pauline Dusseux, Thomas Guyet, Pierre Pattier, Valentin Barbier, Hervé Nicolas

- A generic approach is proposed to identify a subset of new features derived from the spectral indices.
- The predicted grassland height average RMSE is 1.78 ± 0.30 *cm* on the test set.
- The model trained on the data collected from 2017 to 2019 achieves similar accuracy with data from 2020.