Journal Pre-proof

Alfalfa yield estimation based on time series of Landsat 8 and PROBA-V images: An investigation of machine learning techniques and spectral-temporal features

Mohsen Azadbakht, Davoud Ashourloo, Hossein Aghighi, Saeid Homayouni, Hamid Salehi Shahrabi, AliAkbar Matkan, Soheil Radiom

PII:	S2352-9385(21)00193-2
DOI:	https://doi.org/10.1016/j.rsase.2021.100657
Reference:	RSASE 100657
To appear in:	*Remote Sensing Applications: Society and Environment*
Received Date:	27 April 2021
Revised Date:	19 October 2021
Accepted Date:	19 October 2021

Please cite this article as: Azadbakht, M., Ashourloo, D., Aghighi, H., Homayouni, S., Shahrabi, H.S., Matkan, A., Radiom, S., Alfalfa yield estimation based on time series of Landsat 8 and PROBA-V images: An investigation of machine learning techniques and spectral-temporal features, *Remote Sensing Applications: Society and Environment* (2021), doi: https://doi.org/10.1016/j.rsase.2021.100657.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2021 Published by Elsevier B.V.

Authors & affiliations

Manuscript: Alfalfa yield estimation based on time series of Landsat 8 and PROBA-V images: an investigation of Machine Learning techniques and spectral-temporal features

Mohsen Azadbakht (Corresponding author): m_ azadbakht@sbu.ac.ir

Center for Remote Sensing and GIS Research, Faculty of Earth Sciences, Shahid Beheshti University, Tehran, Iran

Davoud Ashourloo: d_ ashourloo@sbu.ac.ir

Center for Remote Sensing and GIS Research, Faculty of Earth Sciences, Shahid Beheshti University, Tehran, Iran

Hossein Aghighi: h_ aghighi@sbu.ac.ir

Center for Remote Sensing and GIS Research, Faculty of Earth Sciences, Shahid Beheshti University, Tehran, Iran

Saeid Homayouni: saeid.homayouni@ete.inrs.ca

INRS, Centre Eau Terre Environnement, Québec (Québec) G1K 9A9, Canada

Hamid Salehi Shahrabi: hamidsalehi2007@gmail.com

Center for Remote Sensing and GIS Research, Faculty of Earth Sciences, Shahid Beheshti University, Tehran, Iran

AliAkbar Matkan: a-matkan@sbu.ac.ir

Center for Remote Sensing and GIS Research, Faculty of Earth Sciences, Shahid Beheshti University, Tehran, Iran

Soheil Radiom: soheil.radiom@gmail.com

Applied Remote Sensing department, Iranian Space Research center, Tehran, Iran

Journal Pre-proof

Alfalfa yield estimation based on time series of Landsat 8 and PROBA-V images: an investigation of Machine Learning techniques and spectral-temporal features

November 12, 2021

Abstract

Remote Sensing (RS) technology provides regular monitoring of alfalfa farms, as a major source of forage production worldwide. Phenological characteristics derived from time series of RS imagery provide a valuable information source to estimate crop yield accurately. In this study, we computed spectral vegetation indices (SVIs) from time series of Landsat 8 and PROBA-V images to extract temporal characteristics of alfalfa farms throughout the growth periods in three consecutive years in the Moghan plain, Iran. Then, several new spectral-temporal features were developed based on phenological characteristics of alfalfa during the growing season. Such features particularly describe geometry and variations of the temporal curves and are thus invaluable in describing phenological attributes. We conducted several feature selection methods due to the variety of features. Machine learning (ML) methods, including ridge, lasso, Gaussian Process Regression (GPR), Random Forest Regression (RFR), Boosted Regression Trees (BRT), and Support Vector Regression (ν-SVR) were utilized to build inversion models in order to estimate alfalfa yields, where the results showed satisfactory performance of GPR using the selected features by GS (RMSE=1114.0 kg/ha), RReliefF (RMSE=1157.7 kg/ha) and Boruta (RMSE=1210.2 kg/ha) as compared to the complete feature dataset (RMSE=1237.4 kg/ha). Overall, the developed phenological features coupled with feature selection methods resulted in the appropriate performance of the ML methods in alfalfa yield estimation.

Precision agriculture, alfalfa yield estimation, Machine Learning, Time series images, Feature selection.

1 Introduction

Food security is a critical issue in both developing and developed countries (Matton et al., 2015). Accurate crop yield estimation plays an important role in ensuring food security and agricultural management at farm level. Crop yield estimation has been emphasized as a critical component of remote sensing applications in agricultural studies (Atzberger, 2013; Aghighi et al., 2018).

Several techniques have been utilized for yield forecasting and estimation (Sakamoto et al., 2013; Guan et al., 2017; Azzari et al., 2017; Huang et al., 2016). Remotely sensed yield estimation techniques offer advantages over other methods such as ground surveys by providing broad coverage, reasonable accuracy, and less expense.

Remote sensing data have been repeatedly adapted to estimate crop yields at local and regional scales (Azzari et al., 2017; Aghighi et al., 2018; Lobell et al., 2015; Lambert et al., 2018). Yield prediction using optical remote sensing employs methods based on spectral vegetation indices (SVIs), machine learning (ML) techniques and crop simulation models (Chahbi et al., 2014). For instance, SVIs extracted from remote sensing data have been utilized for forecasting maize biomass and yield at local and regional scales with a correlation coefficient of more than 80% (Battude et al., 2016) as well as to predict the yield of wheat using Sentinel-2 (Zhao et al., 2020). Although SVIs derived from one or more images show acceptable results during particular stages of the growth period, they only use few spectral bands for yield estimation, and they may not be efficient enough when many factors and relationships between bands must be considered. Moreover, SVIs usually suffer from saturation effects (Xing et al., 2020).

Crop simulation models usually combine growth models, climate data, and SVIs to estimate crop yield. For example, different crop simulation models, e.g. the Simple Algorithm For Yield estimate (SAFY) (Duchemin et al., 2008) and Aquacrop (Steduto et al., 2009), coupled with time series of SVIs derived from satellite images were employed for winter wheat yield estimation (Silvestro et al., 2017) as well as to predict the yield and biomass of maize and sunflower (Claverie et al., 2012). Such models demand a large amount of ground truth data, in addition to numerous input parameters that have to be optimized and making it a time-consuming task that adversely influences models' performance, especially when there is a lack of adequate field data.

On the other hand, ML techniques can employ full-spectrum simultaneously, and they have been known as efficient methods in vegetation studies (Azadbakht et al., 2019). ML methods have also been used to predict the yield of several crops such as maize (Aghighi et al., 2018), wheat (Pantazi et al., 2016) and alfalfa (Feng et al., 2020). Among various ML methods, Support Vector Machine (SVM) and Random Forest (RF) have shown promising results in processing large amounts of data (Ebrahimy and Azadbakht, 2019). They can capture multivariate and nonlinear relationships between dependent and independent variables (Azadbakht et al., 2018; Ghaseminik et al., 2021). These two ML techniques have been employed successfully for estimating biochemical and biophysical properties of various crops (Verrelst et al., 2012; Durbha et al., 2007). However, less attention has been paid to alfalfa yield estimation. ML methods may overfit due to the curse of dimensionality in cases where the variable-to-observation ratio is large (Aghighi et al., 2018; Van der Walt and Barnard, 2006). Additionally, combination of SVIs and ML techniques can significantly improve remote sensing yield estimation. For instance, Aghighi et al. (2018), Johnson et al. (2016), Pantazi et al. (2016), Feng et al. (2020), and Yu and Shang (2018) emphasized the potential of ML techniques in combination with SVIs for yield estimating of various crops.

As one of the significant sources of forage production globally, accurate alfalfa yield estimation is of paramount importance. The growth stages of alfalfa and the harvest time largely depend on several factors, such as alfalfa varieties, production year of the given field, crop implantation and accessibility to harvest machines (Pittman et al., 2015; Ashourloo et al., 2018; Feng et al., 2020). An inherent characteristic of alfalfa fields is that they are generally harvested three to seven times a year, which is not common with other crops and makes alfalfa yield estimation challenging. To the best of our knowledge, there are few studies about alfalfa yield estimation using remotely sensed data because of the aforementioned specific characteristics of alfalfa during the growing season. On the one hand, continuous monitoring of crop phenology, crop growth and disturbances to crop growth are valuable information for yield estimation and managing the crop production risks. This can be performed using time series of remote sensing data. Hence, it is crucial to study alfalfa yield estimation based on the time series of remote sensing data. On the other hand, due to the high performance of ML techniques in handling data from different sources, they have been effectively used to predict biochemical and biophysical properties of crops (Aghighi et al., 2018). Therefore, this research aims to extract new features from the time series of SVIs to describe alfalfa's phenological characteristics by utilizing several ML methods for alfalfa yield estimation using Landsat 8 and PROBA-V images.

2 Materials and Methods

In this section, the study area of the current research is described, and then ground truth data collection and the employed satellite imagery are presented. In the following, the feature dataset developed over time series of satellite images is presented. Large feature datasets may result in the course of dimensionality and overfitting ML methods (Cawley and Talbot, 2010). Therefore, a set of feature selection methods is introduced to select the most informative subset of features while reducing the computational burden. The adopted ML methods, namely Gaussian Process Regression (GPR), Random Forest Regression (RFR), Boosted Regression Trees (BRT), and ν-SVR, for alfalfa yield estimation are then presented and the evaluation measures are described.

Figure 1 shows the adopted workflow for the present study, where initially Landsat 8 and PROBA-V satellite images are acquired across the three

Figure 1:Workflow of alfalfa yield estimation using the machine learning methods and time series of Landsat 8 and PROBA-V images.

Figure 2:Map of Moghan study area, and a true color composite (red, near-infrared and blue, respectively) of Landsat 8 satellite image of the study area.

consecutive years, and cloud-free images are selected to build three time series. Then, common SVIs are calculated from the given time series, and feature datasets are generated to describe the spectral-temporal characteristics of alfalfa fields. A set of feature selection methods is then employed to identify the most essential features. Adopted ML models are subsequently built using both the entire feature dataset and the selected feature subsets using the feature selection methods. Alfalfa yields of the combined dataset across the three years are predicted, and the ML methods' performance is evaluated under different circumstances. The best subsets of features are finally introduced, and inter-comparison among them are conducted.

2.1 Study area

Moghan Agro-industrial & Animal Husbandry Company is one of the largest agricultural companies in the North-West of Iran with over 30 thousand hectares of agricultural lands. It is located between the northern latitudes of 39.465 and 39.615 and the eastern longitudes of 47.548 and 48.009 (see Figure 2). The climate of this area is semi-arid with an average annual rainfall of about 310 mm (Hamdi Ahmadabad et al., 2021). Soils of the area are loam, silt, and clay. Above 90% of the lands in this area are irrigated, and the rest are rain-fed lands. The study area's main crops are wheat (Triticum aestivum), barley (Hordeum vulgare), alfalfa (Medicago sativa), canola (Brassica napus), cotton (Gossypium herbaceum Linnaeus), corn (Zea mays), and sugar beet (Beta vulgaris). This area is one of Iran's most advanced agricultural areas, with advanced machinery, agricultural irrigation and drainage network, and advanced harvesting machines.

2.2 Data used in the study

2.2.1 Ground data of alfalfa yield

The yields of 86, 88, and 88 alfalfa fields were respectively collected in 2014, 2015, and 2016. The field areas in the study area vary from 4 to 20 hectares, and their alfalfa yield values range from 2,152 and 20,870 kilograms per hectare (kg/ha). The crops were harvested by alfalfa harvesters, and the net yields were recorded using a digital weighbridge. The yield measurements were relatively accurate with an accuracy of 1 kg/ha. The final yields were obtained by normalization of the accumulated yields (in *kg*) per area of each field (in *ha*).

Table 1 Number of available cloud free satellite images in each year.

Satellite	2014	2015	2016
Landsat 8	5	11	8
PROBA-V	37	26	20

Table 2 Spectral Bands of PROBA-V.

Band Number	Band Name	Wavelength Center (μm)	Spectral Range (μm)
1	Blue	0.464	0.440 – 0.487
2	NIR	0.837	0.772 – 0.902
3	Red	0.655	0.614 – 0.696
4	SWIR	1.603	1.570 – 1.635

The alfalfa fields' location was recorded during a field campaign using a handheld global positioning system (GPS) receiver with a positional error of less than 2 m.

2.2.2 Satellite data

In this study, Landsat 8 OLI and PROBA-V (see Table 2) satellite images were employed to create the time series of various SVIs. The Landsat images were downloaded from the US Geological Survey (USGS) LSDS Science Research and Development (LSRD) website (`https://espa.cr.usgs.gov/`) and the PROBA-V data was obtained from the ESA Product Distribution Portal (`https://www.vito-eodata.be/PDF/portal/Application.html#Home`). The spatial resolution of Landsat 8 and PROBA-V are respectively 30 and 100 m, with the temporal resolutions being 16 and 5 days, respectively. The reason for including PROBA-V images, in addition to the Landsat images, was that due to the weather conditions and sky cloudiness, there were large temporal gaps between some Landsat 8 images during the growing season. Cloud-free surface reflectance images of the blue, green, red, Near Infrared (NIR), and Short-Wave Infrared (SWIR) bands were used during the alfalfa growth period. All images were obtained in geometrically and atmospherically corrected formats. In this research, we used 42, 37 and 28 images in 2014, 2015 and 2016, respectively (Table 1).

Figure 3 shows the acquisition times of the satellite images in the three years. As can be seen, there is a limited number of available Landsat 8 images, and also some gaps exist between the acquisition times of all three year datasets. This might cause significant errors in alfalfa yield estimation, as such gaps may occur at peaks of greenness, and thus no information can then be recorded. An available precise map of the field boundaries was utilized to calculate field-based spectral reflectance values from the satellite images. The mean spectral reflectance values of all pixels located within a given field were considered for feature development to estimate alfalfa yields.

Figure 3:Acquisition times of the available Landsat 8 and PROBA-V satellite images used in this study.

Figure 4:Color composites (R:Red, G:Near-infrared, B:Blue) of alfalfa farms with various patterns on Julian Days (a) 150, (b) 170, and (c) 200 in 2014. (d) Temporal NDVI profiles of the four alfalfa fields depicted in (a-c).

2.2.3 Developing the feature dataset

One of the main objectives of this study is to develop a new spectral-temporal feature dataset based on alfalfa's specific spectral characteristics during the growing season. As previously stated, alfalfa fields are harvested periodically (Tang et al., 2018), depending on many factors, including the cultivation year and growing conditions. The harvesting frequency of two-/three-year-old alfalfa fields is higher than for other fields, resulting in higher yields. These intrinsic properties of alfalfa fields result in periodic variations of their reflectance values throughout the cultivation years (see Figure 4), which can be illustrated using various SVIs (Ashourloo et al., 2018). Therefore, in this research, we firstly compute prevalent SVIs from the time series of remote sensing images. Secondly, new spectral-temporal features are suggested based on the calculated SVIs.

Several SVIs have been developed and applied to crop yield estimation (Bolton and Friedl, 2013; Panda et al., 2010). However, in this study, from the spectral reflectance data obtained at different wavelengths, the commonly used SVIs as shown in Table 3 were calculated. These include the Normal Difference Vegetation Index (NDVI), Enhanced Vegetation Index-2 (EVI2), Optimized Soil Adjusted Vegetation Index (OSAVI), Land Surface Water Index (LSWI), and Simple Ratio (SR). Therefore, there are five temporal profiles called curve henceforth, for each farm representing the indices across time. Although these SVIs are mostly based on similar spectral bands, some differences have been reported in the literature. For example, NDVI is saturated at high leaf area index (LAI) levels, while EVI2 is not saturated rapidly (Gerstmann et al., 2016; Tang et al., 2018).

An imaginary NDVI curve of an alfalfa field with three temporal peaks is

Table 3 Spectral vegetation indices used in this study.

SVI	Description	Reference
SR	ρ_{NIR}/ρ_{Red}	(Tucker and Sellers, 1986)
NDVI	$\frac{\rho_{NIR}-\rho_{Red}}{\rho_{NIR}+\rho_{Red}}$	(Tucker and Sellers, 1986)
EVI2	$2.5 * \frac{\rho_{NIR}-\rho_{Red}}{(\rho_{NIR}+2.4*\rho_{Red}+1)}$	(Gitelson et al., 2003)
OSAVI	$1.16 * \frac{\rho_{NIR}-\rho_{Red}}{(\rho_{NIR}+\rho_{Red}+0.16)}$	(Huete et al., 1994)
LSWI	$\frac{\rho_{NIR}-\rho_{SWIR}}{\rho_{NIR}+\rho_{SWIR}}$	(Xiao et al., 2002)

Figure 5 A typical synthetic curve of a given alfalfa farm.

shown in Figure 5. The three temporal peaks are located at times t_i, t_j, t_k, and the first peak is marked at $(t_i, \boldsymbol{NDVI}(t_i))$. Positive slopes of the second temporal peak, for example, are located within the interval of $[t_{j-u}, t_j]$, while negative slopes of the given peak span from t_j to t_{j+1}. According to Figure 5, for a given curve $\boldsymbol{I}(t)$, a set of features is extracted. These include the summation of the curve values from the start (SOY) to the end (EOY) of a cultivation year (Eq. 1), number of peaks within the temporal profile (Eq. 2), summation of peak values (Eq. 3), summation of the absolute values of slopes (Eq. 4), summation of positive slopes (Eq. 5) and summation of areas under the entire curve (Eq. 6). Therefore, the complete feature dataset is comprised of 30 features (i.e. 5 SVIs×6 features).

$$\boldsymbol{Sum} = \sum_{t=SOY}^{EOY} \boldsymbol{I}(t) \tag{1}$$

$$\boldsymbol{\#Peaks} = \boldsymbol{Number\ of\ peaks}(\boldsymbol{I}(t)) \tag{2}$$

$$\boldsymbol{Sum}_{Peaks} = \sum_{m=1}^{\#Peaks} \boldsymbol{I}(t_{pk}(m)) \tag{3}$$

$$\boldsymbol{Sum}_{|Slopes|} = \sum_{t=SOY}^{EOY} \mid Slope(\boldsymbol{I}(t)) \mid \tag{4}$$

$$\boldsymbol{Sum}_{+Slopes} = \sum_{t=SOY}^{EOY} Slope(\boldsymbol{I}(t)) > 0 \tag{5}$$

$$\boldsymbol{AUC} = \int_{t=SOY}^{t=EOY} \boldsymbol{I}(t)dt \tag{6}$$

$\boldsymbol{Sum}_{Peaks}$ of each curve is calculated subsequent to identify all the peaks located on the corresponding curve's temporal profile. To this end, the *findpeaks* function in Matlab (The Mathworks, Inc., Natick, MA, USA), with a prominence of at least 0.1, was employed. Since the time interval between the available satellite images is not equal, a 30-day interval between the peaks was considered as the minimum time interval through comparison of the Julian days. $\boldsymbol{Sum}_{|Slopes|}$ considers both positive and negative slope values of the temporal curves and, then, calculates the summation of absolute values of the slopes, whiles $\boldsymbol{Sum}_{+Slopes}$ only takes positive slopes located on the leading edge (e.g., $[t(j-u), t(j)]$ in Figure 5) of each peak on the temporal profile into account. The latter ($\boldsymbol{Sum}_{+Slopes}$) represents the growth rate of alfalfa farms and can be linked to the health status and phenological characteristics of a given farm. For example, frequent high positive slopes in a time series is an indication of a healthy alfalfa farm, and therefore a higher yield is expected compared to a farm with less frequrnt high positive slopes in the time series. AUC is the sum of areas under the temporal SVI and is calculated from SOY to EOY. As another feature, Sum simply calculates the summation of a given SVI value across a cultivation year.

2.3 Feature Selection Methods

In most applications, including far too many features (input variables) does not guarantee a learning algorithm's success and may lead to over-fitting in model training and, therefore, poor prediction ability. On the other hand, it is proved that only a subset of the complete set of features can be sufficient, because exclusion of irrelevant features that are not informative reduces the data complexity and the computational cost on subsequent modeling steps (Guyon and Elisseeff, 2003).

There are three categories of feature selection methods: filters, wrappers, and embedded methods (Stańczyk, 2015). Filter methods score features prior to modeling via considering intrinsic characteristics of features independent of a learning method. The selected features can therefore be served as inputs to any learning model (Roffo, 2016). Wrapper methods employ a particular learning algorithm to evaluate the importance of a feature subset, leading to better outcomes for the chosen algorithm but not necessarily for others. Embedded methods are constructed by combining filter and wrapper methods and interact with learning methods, making the selected features suitable for that learning method (Guyon and Elisseeff, 2003).

RRelief-F is a simple and widely used feature selection method that belongs to the filter methods and provides each feature's weight using the nearest neighbor approach (Stańczyk, 2015). In regression problems, the response values are continuous, and therefore, a probability value based on the relative distance between the predicted values of a pair of samples is adopted to assign weight to features (Robnik-Šikonja and Kononenko, 1997).

A typical example of wrapper methods is forward feature selection based on an iterative manner, with a bottom-up feature selection scheme, starting with no feature and sequentially adding features that improve a learning model until no further improvements are achieved (Marcano-Cedeno et al., 2010). In this study, the Gram-Schmidt (GS) Orthogonalization is used in order to project the feature space onto the response vector. After projection, the features are ranked based on the decrease of their relevance to the response vector. The smallest angle to the direction of maximum projection of input vectors onto the response vector is regarded as the most relevant feature. This process is sequentially employed until a specified number of features are selected (Stoppiglia et al., 2003; Liu et al., 2018).

Boruta is a wrapper algorithm that trains random forest and recursively eliminates unimportant features to find strongly or weakly important features (Kursa et al., 2010). This algorithm generates copies of features and then shuffles their values to remove their correlations with the response variable. A random forest classifier is then implemented on the combination of the shuffled copies with the original data. The original features' importance is assessed against the randomized features, and only features of higher importance than the randomized ones are reflected as important (Kursa et al., 2010).

In Recursive Feature Elimination (RFE), the best subset of features is constructed by eliminating features recursively and then ranking them based on evaluating the cost function changes. As a result, features with the minimum ranking scores are eliminated. This process is repeated until either a predefined number of features is selected or no additional features have remained (Ebrahimy and Azadbakht, 2019; Guyon et al., 2002).

2.4 Alfalfa yield prediction using ML techniques

Advanced machine learning techniques, namely GPR, RFR, BRT, and $\nu-$SVR were utilized to estimate alfalfa yield across three consecutive cultivation years of 2014–2016. Data samples were initially shuffled as a pre-processing step before running the ML methods 100 times to reduce any bias regarding the order of observations. Prior to each run, the observations were divided into training and test data by adopting 5-$fold$ cross-validation.

2.4.1 Ridge and Lasso Regression

Ridge regression (Hoerl and Kennard, 1970) shrinks the regression coefficients toward zero by imposing a penalty on their size. In lasso regression (Tibshirani, 1996), however, some coefficients are exactly zero. In this study, ridge and lasso regression were considered as benchmark. The Lagrangian forms of ridge and lasso are as follows (Hastie et al., 2013):

$$\hat{\beta}_{ridge} = \underset{\beta}{\operatorname{argmin}} \left\{ \sum_{i=1}^{M} \left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j \right)^2 + \lambda \sum_{j=1}^{p} \beta_j^2 \right\} \tag{7}$$

$$\hat{\beta}_{lasso} = \underset{\beta}{\operatorname{argmin}} \left\{ \frac{1}{2} \sum_{i=1}^{M} \left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j \right)^2 + \lambda \sum_{j=1}^{p} \mid \beta_j \mid \right\} \tag{8}$$

Here, $\lambda \geq 0$ is a regulization parameter that controls the shrinkage. Noticeably, the L_2 ridge penalty is replaced by the L_1 lasso, resulting in nonlinear solutions in y_i.

2.4.2 Support Vector Regression

In order to minimize the loss function, Support Vector (SV) machine considers only residuals larger than a specified threshold and thus finds the corresponding coefficients (Hastie et al., 2013). To this end, the input space is mapped onto a feature space of a higher dimension. Then, linear regression is performed in this new feature space using ϵ-insensitive loss (Vapnik, 2013; Cherkassky and Mulier, 1998). Various kernels can be employed to adjust nonlinear boundaries between the features in the feature space (Hastie et al., 2013). We, implement ν-SVR, where $\nu \in$[0,1] controls the number of support vectors and training errors (Chang and Lin, 2002).

In this study, we use the Radial Basis Function (RBF) as a popular kernel function. The best hyperparameters ν, cost (C) and γ were selected through performing grid search within the intervals [0,1], $[2^{-8}, 2^{+8}]$ and $[2^{-8}, 2^{+8}]$, respectively, using the e1071 library (Meyer et al., 2017) in R environment (R Core Team, 2017).

2.4.3 Random Forest Regression

Random forests (RF) reduces the variance of the final ensemble model through taking subsamples from input features while bootstrapping samples from the observations in building individual weak learners (Hastie et al., 2013). Given B the number of constructed trees $T(x; \theta_b)$, the final predicted value in regression problems is obtained through calculating the average response of the entire trees (Breiman, 2001). The minimum leaf size and the number of features at each node of a total of 500 trees are optimised using the Bayesian optimization algorithm (Snoek et al., 2012).

2.4.4 Boosted Regression Trees

Boosted regression trees (BRT) (Elith et al., 2008) is built through integration of boosting (Schapire, 2003) and regression trees (Breiman et al., 1984). With high statistical interpretability (Friedman et al., 2000), it can handle high nonlinearities between input features. RF is reported to reduce variance in the integrated final model, though due to bootstrapping cannot reduce bias, and the ultimate bias is identical to that of individual trees. However, sequential modeling of residuals throughout all BRT observations reduces both bias and variance possible (Elith et al., 2008). In order to avoid overfitting due to a large number of trees (James et al., 2013), its hyperparameters, namely the number of trees, interaction depth, and learning rate, are optimized using the Bayesian optimization algorithm (Snoek et al., 2012).

2.4.5 Gaussian Process Regression

Gaussian Process Regression (GPR) is a kernel-based non-parametric ML method for regression problems (Lázaro-Gredilla et al., 2014; Rasmussen and Williams, 2006). It generates a prior GPR from the training dataset, and a posterior GPR is then generated from it (Ashourloo et al., 2016; Azadbakht et al., 2019). GPR extracts several relationships between the input and target variables in order to accurately describe their correlations (Williams, 1998). In this study, hyperparameters of the squared-exponential covariance function; namely the length-scale ℓ, the signal variance σ_f and the noise variance σ_n are optimized using the Bayesian optimization algorithm (Snoek et al., 2012).

2.5 Performance Evaluation Measures

Given Table 3 and Eq. 1-6, a feature dataset consisting of 30 features was generated, where the selected ML methods were implemented on the observations of alfalfa yields. This process was repeated 100 times in order to reduce the bias of random splitting. At each of these 100 runs, the whole dataset was initially split into five equally sized folds, where a test fold was sequentially set aside, and the four remained folds were considered to create training datasets iteratively. Then, the test fold is changed, and a similar process is implemented until alfalfa yields are predicted for samples in the five test folds.

Three measures, namely the coefficient of determination ($\boldsymbol{R}^2$), mean absolute error (MAE) and the root mean square error (RMSE) are used for performance evaluation of the ML methods. A one-factor analysis of variance (ANOVA) (Devore and Berk, 2012) is also used to examine the statistical significance of the results at the α =5% level. To this end, the average performance of the ML methods is compared across either original feature dataset or different feature subsets derived from the five feature selection (FS) methods (Azadbakht et al., 2019). In this way, the null hypothesis is that these average performances are similar, while at least one different performance refers to the alternative hypothesis (see Eq. 9). In the latter case, pairwise comparison is subsequently conducted using Tukey's honestly significant difference (HSD) among the scenarios to select models with significantly different performance.

$$\begin{cases} H_0 : \mu_i = \mu_j ;\ where\ i \neq j \\ \\ H_1 : \text{at least one is different} \end{cases} \tag{9}$$

Here, μ_i refers to the mean performance of the i-th ML method, based on the evaluation measures.

3 Results and Discussion

3.1 Alfalfa yield estimation using all features

Boxplots in Figure 6 show how well the predicted alfalfa yields fit the actual yields under 100 runs of the ML methods, in terms of the MAE and RMSE measures. As seen, inferior performance of lasso, ridge and ν-SVR is evident in terms of both measures, exhibting the median RMSE value of about 2848.6, 2826.0 and 2889.9 kg/ha, respectively. In total, of the four ML methods, ν-SVR and RFR performed with less deviations across 100 runs, showing smaller inter-quartile (IQR) values compared to the results of BRT and GPR. The boxplots in Figure 6, however, show better performance of GPR and BRT with more desirable values of upper quartile, lower quartile, and median values, followed by RFR.

The average predicted yield values using the ML methods over 100 runs were calculated and scattered against the actual yields of the corresponding alfalfa

Table 4:The lower, median and upper quartiles in terms of R^2 and RMSE (kg/ha).

Quartiles	BRT		GPR		RFR		SVR		Ridge		I
	R^2	$RMSE$	R^2	$RMSE$	R^2	$RMSE$	R^2	$RMSE$	R^2	$RMSE$	R^2
Lower quartile	0.48	2029.7	0.61	1594.4	0.48	2244.9	0.19	2855.7	0.24	2776.2	0.21
Median	0.55	2177.9	0.67	1860.2	0.51	2311.8	0.21	2889.9	0.25	2826.0	0.23
Upper quartile	0.62	2343.3	0.75	2020.6	0.54	2355.1	0.22	2928.2	0.27	2862.6	0.25
IQR	0.13	313.6	0.15	426.2	0.06	110.1	0.03	72.5	0.03	86.4	0.04

Figure 6:Boxplots of the machine learning methods over 100 runs in terms of MAE and RMSE.

farms (Figure 7). As seen, the average yield values in GPR, BRT, and RFR exhibit solid linear correlation with the observed values, while a less consistent correlation is evident in $\nu-$SVR, ridge and lasso. Larger deviations of the predicted alfalfa yields from the 1:1 line indicates that lasso, ridge and $\nu-$SVR experience higher levels of under/over-estimation of alfalfa yields at higher/lower actual yields. It must be remarked that more significant underestimation rates of predicted alfalfa yields are evident in lasso, ridge, $\nu-$SVR, and RFR, particularly for higher yield values.

For performance evaluation of the ML methods, deviations of the average predicted alfalfa yields were also calculated across farms in terms of the RMSE measure. The calculated RMSE values for BRT, GPR, RFR, and $\nu-$SVR were 1902.9, 1237.4, 2240.3, and 2829.8 kg/ha, respectively. This obviously shows that GPR predicted alfalfa yields, on average, with the maximum correlation (R^2=0.91) and the minimum RMSE value of 1237.4 kg/ha. As shown in Figure 3, unevenly distributed available images in each year are evident, in which influences the combined temporal SVIs and the calculated features, and thus the built ML models.

An ANOVA was carried out on performance of the ML methods across 100 runs in terms of the RMSE measure. The obtained p-value of smaller than 0.05 confirmed that the null hypothesis is rejected and there is at least one pair of ML methods with significantly different performances. The results of Tukey's HSD test showed adjusted p-values between all pairs of ML methods smaller than 0.05, which in turn indicates that all ML methods performed significantly different in terms of RMSE.

Figure 7:Scatter plots of the actual versus predicted alfalfa yields (kg/ha) using the machine learning methods and the complete feature dataset. (Red line and dashed green line are respectively the regression line and 1:1 line.)

Figure 8:Scatter plots of the actual and predicted alfalfa yields (in t/ha) using GPR under different temporal gaps. (Red line and dashed green line are respectively the regression line and 1:1 line.)

Table 5 Top ten selected features by the four feature selection methods.

Boruta	GS	RReliefF	RFE
$\boldsymbol{AUC}$(LSWI)	$\boldsymbol{AUC}$(EVI2)	$\boldsymbol{AUC}$(LSWI)	$\boldsymbol{AUC}$(EVI2)
$\boldsymbol{AUC}$(NDVI)	$\boldsymbol{Sum}$(EVI2)	$\boldsymbol{AUC}$(NDVI)	$\boldsymbol{AUC}$(LSWI)
$\boldsymbol{AUC}$(SR)	$\boldsymbol{Sum}$(LSWI)	$\boldsymbol{AUC}$(OSAVI)	$\boldsymbol{AUC}$(SR)
$\boldsymbol{Sum}$(LSWI)	$\boldsymbol{Sum}$(NDVI)	$\boldsymbol{Sum}$(EVI2)	$\boldsymbol{AUC}$(OSAVI)
$\boldsymbol{Sum}$(NDVI)	$\boldsymbol{Sum}$(OSAVI)	$\boldsymbol{Sum}$(LSWI)	$\boldsymbol{Sum}$(EVI2)
$\boldsymbol{Sum}_{Peaks}$(EVI2)	$\boldsymbol{\#Peaks}$(LSWI)	$\boldsymbol{Sum}$(NDVI)	$\boldsymbol{Sum}$(LSWI)
$\boldsymbol{Sum}_{Peaks}$(LSWI)	$\boldsymbol{Sum}_{\lvert Slopes\rvert}$(EVI2)	$\boldsymbol{Sum}$(OSAVI)	$\boldsymbol{Sum}$(NDVI)
$\boldsymbol{Sum}_{\lvert Slopes\rvert}$(EVI2)	$\boldsymbol{Sum}_{\lvert Slopes\rvert}$(OSAVI)	$\boldsymbol{\#Peaks}$(LSWI)	$\boldsymbol{Sum}_{Peaks}$(EVI2)
$\boldsymbol{Sum}_{+Slopes}$(EVI2)	$\boldsymbol{Sum}_{+Slopes}$(EVI2)	$\boldsymbol{Sum}_{Peaks}$(LSWI)	$\boldsymbol{Sum}_{\lvert Slopes\rvert}$(EVI2)
$\boldsymbol{Sum}_{+Slopes}$(SR)	$\boldsymbol{Sum}_{+Slopes}$(OSAVI)	$\boldsymbol{Sum}_{+Slopes}$(EVI2)	$\boldsymbol{Sum}_{+Slopes}$(EVI2)

3.2 Alfalfa yield estimation under undesirable temporal gaps

In order to consider undesirable temporal gaps in time series, a monthly spectral reflectance data in the blue, red, NIR, and SWIR bands was created. We then applied different random temporal gaps (0%, 10%, 20%, 30%, 40%, 50%) in the dataset and calculated the spectral-temporal features in Eq. 1-6. Random temporal gaps were applied to the dataset, since temporal gaps in satellite data do not follow a specific pattern. Of the ML methods, GPR was applied, as the best ML method, on the datasets to evaluate performance of the developed features under different levels of temporal gaps. As can be seen in Figure 8, temporal gaps of up to 50% do not significantly affect the performance of GPR in alfalfa yield estimation. Indeed, temporal gaps of 30% and 40% result in higher RMSE and lower $\boldsymbol{R}^2$ values as compared to the original dataset. Therefore, we can conclude that the introduced spectral-temporal features can compensate for undesirable temporal gaps in time series of satellite images.

3.3 Alfalfa yield estimation using the selected features

To exhibit the importance of the calculated features and examine whether a smaller number of features can provide similar alfalfa yield predictions, we employed Boruta, GS, RReliefF, and RFE to select the most relevant features. Table 5 summarizes the ten most important features based on the four FS methods.

As can be seen in Table 5, of the whole feature dataset, $\boldsymbol{AUC}(.)$ and $\boldsymbol{Sum}(.)$ of different SVIs were selected more frequently using all of the FS methods. The FS methods also selected features related to the slopes (either

Figure 9:Boxplots of the machine learning methods and feature selection techniques over 100 runs in terms of RMSE.

Figure 10:Scatter plots of the actual and predicted alfalfa yields (in t/ha) using the machine learning methods and selected features by different feature selection methods. (Red line and dashed green line are respectively the regression line and 1:1 line.)

positive slopes or absolute values of the entire slopes) of temporal curves, in addition to $\boldsymbol{Sum}_{Peaks}(.)$. Among the SVIs, LSWI and EVI2 more frequently emerged in the selected features. The selected features can play critical roles in describing temporal characteristics of alfalfa curves during the growing season. For example, $\boldsymbol{Sum}_{Peaks}(.)$ refers to the summation of given SVI values placed on the peaks extracted across the time series dataset, $\boldsymbol{AUC}(.)$ represents the area under temporal SVI curve throughout the time series curve, and slope-based features explain growth stages of the crop.

In total, features extracted from EVI2 and LSWI were more selected by most of the FS methods, indicating $\boldsymbol{Sum}_{+Slopes}$(EVI2) and $\boldsymbol{Sum}_{|Slopes|}$(EVI2) as the most frequently selected features of the former, while $\boldsymbol{Sum}$(LSWI) and $\boldsymbol{AUC}$(LSWI) were the most common features derived from the temporal profile of the latter. This clearly demonstrates the importance of considering the area under the temporal SVI curves and summation of slope values of SVI curves across time.

To evaluate the selected features, and for the sake of simplicity, only GPR, RFR and BRT were applied on the selected features. Figure 9 shows the RMSE measure's boxplots for the ML methods based on the selected features by the four FS methods. Although both RFR and BRT exhibit shorter boxplots and thus higher performance stability, superior performance of GPR using the selected features is evident, showing smaller values of the first, second (median), and third quartiles. In particular, features selected by GS and RReliefF provide smaller interquartile values using GPR, respectively 332.4 and 442.5 kg/ha.

Figure 10 shows scatter plots of the predicted versus actual alfalfa yields through implementation of the ML methods on the feature datasets extracted by the four FS methods. Analogous to Figures 7 and 9, GPR outperformed both BRT and RFR using the features selected by the FS methods, with the $\boldsymbol{R}^2$ values of the linear regression equations of two circumstances with the selected features by GS ($\boldsymbol{R}^2$=92%) and RReliefF ($\boldsymbol{R}^2$=92%) being higher than the case of implementing this ML method on the complete feature dataset ($\boldsymbol{R}^2$=91%).

Table 6 shows the lower and upper quartiles, median and IQR metrics in terms of $\boldsymbol{R}^2$ and RMSE values of the predicted alfalfa yields, over 100 runs, against the actual alfalfa yields for combinations of ML-FS methods. As seen in this table, the RMSE values of GPR using the features selected by GS, RReliefF, and Boruta are smaller than those of implementing this ML method on the entire feature dataset (Table 4). RFR and BRT, however, showed inferior

Table 6:The lower, median, upper quartiles, and IQR values in terms of $\boldsymbol{R}^2$ and RMSE values (in kg/ha) of the ML methods using the selected features by the FS methods.

ML Method	Quartiles	FS Method							
		Boruta		GS		RFE		ReliefF	
		$\boldsymbol{R}^2$	RMSE	$\boldsymbol{R}^2$	RMSE	$\boldsymbol{R}^2$	RMSE	$\boldsymbol{R}^2$	RMSE
BRT	Lower quartile	0.46	2239.8	0.41	2304.9	0.42	2366.5	0.42	2344.5
	Median	0.50	2318.4	0.46	2394.1	0.44	2437.7	0.46	2393.7
	Upper quartile	0.53	2395.5	0.50	2500.4	0.48	2479.3	0.48	2457.2
	IQR	0.07	155.6	0.09	195.5	0.06	112.9	0.06	112.6
GPR	Lower quartile	0.60	1520.0	0.67	1516.3	0.56	1646.9	0.65	1461.9
	Median	0.69	1810.7	0.72	1696.9	0.65	1896.2	0.71	1726.7
	Upper quartile	0.78	2038.3	0.78	1848.7	0.74	2141.9	0.80	1904.4
	IQR	0.18	518.3	0.11	332.4	0.18	495.0	0.15	442.5
RFR	Lower quartile	0.44	2354.5	0.47	2317.8	0.45	2335.9	0.43	2409.4
	Median	0.46	2398.1	0.48	2359.4	0.47	2373.5	0.44	2430.8
	Upper quartile	0.48	2432.5	0.50	2386.0	0.49	2411.7	0.45	2461.0
	IQR	0.04	78.0	0.03	68.2	0.04	75.7	0.03	51.6

performances with higher RMSE values when coupled with the FS methods compared to when implemented on the total feature dataset. Noticeably, RFR, coupled with the FS methods, provided the least IQR values for both $\boldsymbol{R}^2$ and RMSE measures.

Table 6 and Figure 10 also show that both GPR and RFR provide the least average RMSE and highest $\boldsymbol{R}^2$ values using the features provided by GS, while BRT exhibits the smallest RMSE value using the features selected by Boruta. Among the full feature dataset, GS selected the area under the temporal curve of EVI2 (AUC($\boldsymbol{EVI2}$)), summation of temporal values of SVIs (Sum(.)), number of peaks of temporal LSWI ($\boldsymbol{\#Peaks}$(LSWI)), summation of either poisitive slopes ($\boldsymbol{Sum}_{+Slopes}$(.)) or absolute values of the slopes ($\boldsymbol{Sum}_{|Slopes|}$(.)) of EVI2 and OSAVI.

Noticeably, there was no common trend among the FS techniques in terms of average RMSE values using the three selected ML methods. For example, GS provided the smallest average RMSE values in both GPR and RFR, while the minimum RMSE value for BRT occured using features selected by Boruta. GPR, particularly coupled with GS (RMSE=1114.0 kg/ha), outperformed other combinations of the FS techniques and ML methods. It was followed by GPR integrated with RReliefF. Among the top 10 features provided by GS (Table 5), $\boldsymbol{Sum}_{+Slopes}$(OSAVI) and $\boldsymbol{Sum}_{|Slopes|}$(OSAVI) were not selected by the other FS methods. This fact may indicate the better performance of GPR-GS compared to the cases that other FS methods were combined with. These two features mainly characterize the importance of slopes of temporal SVI profiles. Moreover, $\boldsymbol{\#Peaks}$(LSWI) and $\boldsymbol{Sum}$(OSAVI) were common between GS and RReliefF that provided the least average RMSE values in combination with GPR.

Table 7:Tukey's HSD test between the pairs of ML methods implemented on the selected features by the FS methods in terms of the RMSE values.

	BRT-Boruta	BRT-GS	BRT-RFE	BRT-RReliefF	GPR-Boruta	GPR-GS	GPR-R
BRT-Boruta	-	0.86	0.19	0.70	**0.00**	**0.00**	**0.00**
BRT-GS		-	1.00	1.00	**0.00**	**0.00**	**0.00**
BRT-RFE			-	1.00	**0.00**	**0.00**	**0.00**
BRT-RReliefF				-	**0.00**	**0.00**	**0.00**
GPR-Boruta					-	0.16	0.92
GPR-GS						-	**0.00**
GPR-RFE							-
GPR-RReliefF							
RFR-Boruta							
RFR-GS							
RFR-RFE							
RFR-RReliefF							

An ANOVA analysis was performed on the RMSE values obtained from the 100 times replications of the joint ML-FS methods to compare the performance of the regression models. The p-values/F-statistics of all combinations of ML and FS methods were smaller/larger than the significance level/critical F-values. Therefore, the null hypothesis of similar performance of the joint ML-FS methods, on average, was rejected and at least one of the pairs performed significantly different at the 95% confidence level. Table 7 shows Tukey's HSD results for pairwise comparisons between the ML methods coupled with the FS techniques. In this table, the adjusted p-values of smaller than 0.05 indicate significant differences between the pairs. Noticeably, of the three ML methods, GPR performed significantly different from both RFR and BRT, regardless of the FS methods. On the other hand, no significant differences were found between RFR and BRT's performance in terms of the average RMSE values. Moreover, according to Tables 6 and 7, GPR-RFE and GPR-GS performed significantly different, while no significant differences were found between these and the other combinations of GPR and FS methods.

As shown in Table 5, GS shared 40% and 60% similarity in the selected features with Boruta and RReliefF, respectively. The higher similarity of the latter with GS emerged in the smaller mean RMSE measure of GPR-RReliefF than GPR-Boruta. Despite the high similarity (60%) between the selected features using RFE and GS, Tukey's HSD showed that their performances, on average, were significantly different. Of the six common features between GS and RFE, three were similar to those selected by RReliefF and Boruta. The only features in GS that were not in common with the other FS methods were $Sum_{+Slopes}$(OSAVI) and $Sum_{|Slopes|}$(OSAVI), which can represent the importance of slopes of time series of OSAVI.

An ANOVA analysis was also performed between the joint GPR-FS methods and implementations of GPR on the entire feature dataset, with no significant differences being emerged. Although the average performance of GPR coupled with FS methods did not show significant improvement compared to the complete feature dataset, yet the computational complexity was reduced.

4 Conclusions

In this study, inversion models using the ML methods were built to estimate alfalfa yields. The proposed features mainly describe the geometrical changes of the temporal SVIs across time, representing the phenological characteristics of the alfalfa throughout the cultivation year. GPR outperformed the other ML methods in alfalfa yield estimation across several farms using features extracted from temporal profiles of different SVIs, resulting in higher $\boldsymbol{R}^2$ and lower RMSE values across the three years.

Furthermore, various feature selection techniques were utilized to select the most relevant features for alfalfa yield estimation. The selected features by different feature selection methods could acceptably participate in alfalfa yield estimation while reducing the training time of the ML methods. Of the selected features, the area under the entire temporal SVIs and features related to the slope of the temporal SVIs were among the most common features selected by the four FS techniques.

Compared to the complete feature datasets, the selected features could marginally improve alfalfa yield estimation using GPR in terms of the RMSE measure. GPR implemented on the complete feature datasets provided the average RMSE value of 1237.4 kg/ha, while the average RMSE values of 1114.0, 1157.7, and 1210.2 kg/ha emerged using GPR and the selected features by GS, RReliefF, and Boruta. The introduced features in the current study can estimate alfalfa yield for other areas and estimate the yield of other crops. This approach can be further developed to predict alfalfa yield during the growth period and before harvest.

References

Aghighi, H., Azadbakht, M., Ashourloo, D., Shahrabi, H.S., Radiom, S., 2018. Machine learning regression techniques for the silage maize yield prediction using time-series images of landsat 8 oli. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 11, 4563–4577.

Ashourloo, D., Aghighi, H., Matkan, A.A., Mobasheri, M.R., Rad, A.M., 2016. An investigation into machine learning regression techniques for the leaf rust disease detection using hyperspectral measurement. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 9, 4344–4351. doi:10.1109/JSTARS.2016.2575360.

Ashourloo, D., Shahrabi, H.S., Azadbakht, M., Aghighi, H., Matkan, A.A., Radiom, S., 2018. A novel automatic method for alfalfa mapping using time series of landsat-8 oli data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 11, 4478–4487.

Atzberger, C., 2013. Advances in remote sensing of agriculture: Context description, existing operational monitoring systems and major information needs. Remote sensing 5, 949–981.

Azadbakht, M., Ashourloo, D., Aghighi, H., Radiom, S., Alimohammadi, A., 2019. Wheat leaf rust detection at canopy scale under different lai levels using machine learning techniques. Computers and electronics in agriculture 156, 119–128.

Azadbakht, M., Fraser, C.S., Khoshelham, K., 2018. Synergy of sampling techniques and ensemble classifiers for classification of urban environments using full-waveform lidar data. International journal of applied earth observation and geoinformation 73, 277–291.

Azzari, G., Jain, M., Lobell, D.B., 2017. Towards fine resolution global maps of crop yields: Testing multiple methods and satellites in three countries. Remote Sensing of Environment 202, 129–141.

Battude, M., Al Bitar, A., Morin, D., Cros, J., Huc, M., Sicre, C.M., Le Dantec, V., Demarez, V., 2016. Estimating maize biomass and yield over large areas using high spatial and temporal resolution sentinel-2 like remote sensing data. Remote Sensing of Environment 184, 668–681.

Bolton, D.K., Friedl, M.A., 2013. Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics. Agricultural and Forest Meteorology 173, 74–84.

Breiman, L., 2001. Random forests. Machine learning 45, 5–32. doi:10.1023/A:1010933404324.

Breiman, L., Friedman, J., Stone, C., Olshen, R., 1984. Classification and Regression Trees. The Wadsworth and Brooks-Cole statistics-probability series, Taylor & Francis. URL: https://books.google.com/books?id=JwQx-WOmSyQC.

Cawley, G.C., Talbot, N.L., 2010. On over-fitting in model selection and subsequent selection bias in performance evaluation. Journal of Machine Learning Research 11, 2079–2107.

Chahbi, A., Zribi, M., Lili-Chabaane, Z., Duchemin, B., Shabou, M., Mougenot, B., Boulet, G., 2014. Estimation of the dynamics and yields of cereals in a semi-arid area using remote sensing and the safy growth model. International Journal of Remote Sensing 35, 1004–1028.

Chang, C., Lin, C., 2002. Training nu-support vector regression: theory and algorithms. Neural Computation 14, 1959–1978. doi:10.1162/089976602760128081.

Cherkassky, V., Mulier, F., 1998. Learning from data. adaptive and learning systems for signal processing, communications and control.

Claverie, M., Demarez, V., Duchemin, B., Hagolle, O., Ducrot, D., Marais-Sicre, C., Dejoux, J.F., Huc, M., Keravec, P., Béziat, P., et al., 2012. Maize and sunflower biomass estimation in southwest france using high spatial and temporal resolution remote sensing data. Remote Sensing of Environment 124, 844–857.

Devore, J.L., Berk, K.N., 2012. Regression and correlation, in: Modern Mathematical Statistics with Applications. Springer, pp. 613–722.

Duchemin, B., Maisongrande, P., Boulet, G., Benhadj, I., 2008. A simple algorithm for yield estimates: Evaluation for semi-arid irrigated winter wheat monitored with green leaf area index. Environmental Modelling & Software 23, 876–892.

Durbha, S.S., King, R.L., Younan, N.H., 2007. Support vector machines regression for retrieval of leaf area index from multiangle imaging spectroradiometer. Remote sensing of environment 107, 348–361.

Ebrahimy, H., Azadbakht, M., 2019. Downscaling modis land surface temperature over a heterogeneous area: An investigation of machine learning techniques, feature selection, and impacts of mixed pixels. Computers & Geosciences 124, 93–102.

Elith, J., Leathwick, J.R., Hastie, T., 2008. A working guide to boosted regression trees. Journal of Animal Ecology 77, 802–813. doi:10.1111/j.1365-2656.2008.01390.x.

Feng, L., Zhang, Z., Ma, Y., Du, Q., Williams, P., Drewry, J., Luck, B., 2020. Alfalfa yield prediction using uav-based hyperspectral imagery and ensemble learning. Remote Sensing 12, 2028.

Friedman, J., Hastie, T., Tibshirani, R., et al., 2000. Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). The annals of statistics 28, 337–407. doi:10.1214/aos/1016120463.

Gerstmann, H., Möller, M., Gläßer, C., 2016. Optimization of spectral indices and long-term separability analysis for classification of cereal crops using multi-spectral rapideye imagery. International Journal of Applied Earth Observation and Geoinformation 52, 115–125.

Ghaseminik, F., Aghamohammadi, H., Azadbakht, M., 2021. Land cover mapping of urban environments using multispectral lidar data under data imbalance. Remote Sensing Applications: Society and Environment 21, 100449.

Gitelson, A.A., Gritz, Y., Merzlyak, M.N., 2003. Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. Journal of plant physiology 160, 271–282.

Guan, K., Wu, J., Kimball, J.S., Anderson, M.C., Frolking, S., Li, B., Hain, C.R., Lobell, D.B., 2017. The shared and unique values of optical, fluorescence, thermal and microwave satellite data for estimating large-scale crop yields. Remote sensing of environment 199, 333–349.

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

Guyon, I., Weston, J., Barnhill, S., Vapnik, V., 2002. Gene selection for cancer classification using support vector machines. Machine learning 46, 389–422.

Hamdi Ahmadabad, Y., Liaghat, A., Sohrabi, T., Rasoulzadeh, A., Ebrahimian, H., 2021. Improving performance of furrow irrigation systems using simulation modelling in the moghan plain of iran. Irrigation and Drainage 70, 131–149.

Hastie, T., Tibshirani, R., Friedman, J., 2013. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer Series in Statistics, Springer New York. URL: https://books.google.com/books?id=yPfZBwAAQBAJ.

Hoerl, A.E., Kennard, R.W., 1970. Ridge regression: applications to nonorthogonal problems. Technometrics 12, 69–82.

Huang, J., Sedano, F., Huang, Y., Ma, H., Li, X., Liang, S., Tian, L., Zhang, X., Fan, J., Wu, W., 2016. Assimilating a synthetic kalman filter leaf area index series into the wofost model to improve regional winter wheat yield estimation. Agricultural and Forest Meteorology 216, 188–202.

Huete, A., Justice, C., Liu, H., 1994. Development of vegetation and soil indices for modis-eos. Remote Sensing of Environment 49, 224–234.

James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An Introduction to Statistical Learning: with Applications in R. Springer Texts in Statistics, Springer New York. URL: https://books.google.com/books?id=qcI_AAAAQBAJ.

Johnson, M.D., Hsieh, W.W., Cannon, A.J., Davidson, A., Bédard, F., 2016. Crop yield forecasting on the canadian prairies by remotely sensed vegetation indices and machine learning methods. Agricultural and forest meteorology 218, 74–84.

Kursa, M.B., Jankowski, A., Rudnicki, W.R., 2010. Boruta–a system for feature selection. Fundamenta Informaticae 101, 271–285.

Lambert, M.J., Traoré, P.C.S., Blaes, X., Baret, P., Defourny, P., 2018. Estimating smallholder crops production at village level from sentinel-2 time series in mali's cotton belt. Remote Sensing of Environment 216, 647–657.

Lázaro-Gredilla, M., Titsias, M.K., Verrelst, J., Camps-Valls, G., 2014. Retrieval of biophysical parameters with heteroscedastic gaussian processes. IEEE Geoscience and Remote Sensing Letters 11, 838–842. doi:10.1109/LGRS.2013.2279695.

Liu, R., Wang, H., Wang, S., 2018. Functional variable selection via gram–schmidt orthogonalization for multiple functional linear regression. Journal of Statistical Computation and Simulation 88, 3664–3680.

Lobell, D.B., Thau, D., Seifert, C., Engle, E., Little, B., 2015. A scalable satellite-based crop yield mapper. Remote Sensing of Environment 164, 324–333.

Marcano-Cedeno, A., Quintanilla-Domínguez, J., Cortina-Januchs, M., Andina, D., 2010. Feature selection using sequential forward selection and classification applying artificial metaplasticity neural network, in: IECON 2010-36th annual conference on IEEE industrial electronics society, IEEE. pp. 2845–2850.

Matton, N., Canto, G.S., Waldner, F., Valero, S., Morin, D., Inglada, J., Arias, M., Bontemps, S., Koetz, B., Defourny, P., 2015. An automated method for annual cropland mapping along the season for various globally-distributed agrosystems using high spatial and temporal resolution time series. Remote Sensing 7, 13208–13232.

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F., 2017. e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien. URL: https://CRAN.R-project.org/package=e1071. r package version 1.6-8.

Panda, S.S., Ames, D.P., Panigrahi, S., 2010. Application of vegetation indices for agricultural crop yield prediction using neural network techniques. Remote Sensing 2, 673–696.

Pantazi, X.E., Moshou, D., Alexandridis, T., Whetton, R.L., Mouazen, A.M., 2016. Wheat yield prediction using machine learning and advanced sensing techniques. Computers and Electronics in Agriculture 121, 57–65.

Pittman, J.J., Arnall, D.B., Interrante, S.M., Moffet, C.A., Butler, T.J., 2015. Estimation of biomass and canopy height in bermudagrass, alfalfa, and wheat using ultrasonic, laser, and spectral sensors. Sensors 15, 2920–2943.

R Core Team, 2017. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. URL: https://www.R-project.org/.

Rasmussen, C., Williams, C., 2006. Gaussian Processes for Machine Learning. Adaptative computation and machine learning series, University Press Group Limited. URL: shttps://books.google.com/books?id=vWtwQgAACAAJ.

Robnik-Šikonja, M., Kononenko, I., 1997. An adaptation of relief for attribute estimation in regression, in: Machine Learning: Proceedings of the Fourteenth International Conference (ICML'97), pp. 296–304.

Roffo, G., 2016. Feature selection library (matlab toolbox). arXiv preprint arXiv:1607.01327 .

Sakamoto, T., Gitelson, A.A., Arkebauer, T.J., 2013. Modis-based corn grain yield estimation model incorporating crop phenology information. Remote Sensing of Environment 131, 215–231.

Schapire, R.E., 2003. The boosting approach to machine learning: An overview, in: Nonlinear estimation and classification. Springer, pp. 149–171. doi:10.1007/978-0-387-21579-2_9.

Silvestro, P.C., Pignatti, S., Yang, H., Yang, G., Pascucci, S., Castaldi, F., Casa, R., 2017. Sensitivity analysis of the aquacrop and safye crop models for the assessment of water limited winter wheat yield in regional scale applications. PloS one 12.

Snoek, J., Larochelle, H., Adams, R.P., 2012. Practical bayesian optimization of machine learning algorithms, in: Advances in neural information processing systems, pp. 2951–2959.

Stańczyk, U., 2015. Feature evaluation by filter, wrapper, and embedded approaches, in: Feature Selection for Data and Pattern Recognition. Springer, pp. 29–44.

Steduto, P., Hsiao, T.C., Raes, D., Fereres, E., 2009. Aquacrop—the fao crop model to simulate yield response to water: I. concepts and underlying principles. Agronomy Journal 101, 426–437.

Stoppiglia, H., Dreyfus, G., Dubois, R., Oussar, Y., 2003. Ranking a random feature for variable and feature selection. Journal of machine learning research 3, 1399–1414.

Tang, K., Zhu, W., Zhan, P., Ding, S., 2018. An identification method for spring maize in northeast china based on spectral and phenological features. Remote Sensing 10, 193.

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

Tucker, C., Sellers, P., 1986. Satellite remote sensing of primary production. International journal of remote sensing 7, 1395–1416.

Vapnik, V., 2013. The Nature of Statistical Learning Theory. Springer New York. URL: `https://books.google.com/books?id=EoDSBwAAQBAJ`.

Verrelst, J., Muñoz, J., Alonso, L., Delegido, J., Rivera, J.P., Camps-Valls, G., Moreno, J., 2012. Machine learning regression algorithms for biophysical parameter retrieval: Opportunities for sentinel-2 and-3. Remote Sensing of Environment 118, 127–139.

Van der Walt, C.M., Barnard, E., 2006. Data characteristics that determine classifier performance .

Williams, C.K., 1998. Prediction with gaussian processes: From linear regression to linear prediction and beyond. Nato asi series d behavioural and social sciences 89, 599–621. doi:10.1007/978-94-011-5014-9_23.

Xiao, X., Boles, S., Liu, J., Zhuang, D., Liu, M., 2002. Characterization of forest types in northeastern china, using multi-temporal spot-4 vegetation sensor data. Remote Sensing of Environment 82, 335–348.

Xing, N., Huang, W., Xie, Q., Shi, Y., Ye, H., Dong, Y., Wu, M., Sun, G., Jiao, Q., 2020. A transformed triangular vegetation index for estimating winter wheat leaf area index. Remote Sensing 12, 16.

Yu, B., Shang, S., 2018. Multi-year mapping of major crop yields in an irrigation district from high spatial and temporal resolution vegetation index. Sensors 18, 3787.

Zhao, Y., Potgieter, A.B., Zhang, M., Wu, B., Hammer, G.L., 2020. Predicting wheat yield at the field scale by combining high-resolution sentinel-2 satellite imagery and crop modelling. Remote Sensing 12, 1024.

Journal Pre-proof

Highlights:

- A set of spectral-temporal features is introduced to describe phenological characteristics of alfalfa during the cultivation year.
- Feature selection methods were implemented to identify the most important variables.
- Feature selection could acceptably increase cost-effectiveness of the alfalfa yield estimation procedure.
- Long gaps between available cloud-free satellite images consistently affect the performance of the inversion methods.
- Area under the whole temporal curve of spectral vegetation indices as well as features related to the slope of the temporal curves were among the most common features.

Journal Pre-proof

Journal Pre-proof

Ethical Statement

Hereby, I /**Mohsen Azadbakht**/ consciously assure that for the manuscript /**Alfalfa yield estimation based on time series of Landsat 8 and PROBA-V images: an investigation of Machine Learning techniques and spectral-temporal features**/ the following is fulfilled:

1) This material is the authors' own original work, which has not been previously published elsewhere.

2) The paper is not currently being considered for publication elsewhere.

3) The paper reflects the authors' own research and analysis in a truthful and complete manner.

4) The paper properly credits the meaningful contributions of co-authors and co-researchers.

5) The results are appropriately placed in the context of prior and existing research.

6) All sources used are properly disclosed (correct citation). Literally copying of text must be indicated as such by using quotation marks and giving proper reference.

7) All authors have been personally and actively involved in substantial work leading to the paper, and will take public responsibility for its content.

The violation of the Ethical Statement rules may result in severe consequences.

To verify originality, your article may be checked by the originality detection software iThenticate. See also http://www.elsevier.com/editors/plagdetect.

I agree with the above statements and declare that this submission follows the policies of Remote Sensing Applications: Society and Environment as outlined in the Guide for Authors and in the Ethical Statement.

Date: 23 October 2021
Mohsen Azadbakht
Corresponding author

Journal Pre-proof

AUTHORSHIP STATEMENT

Manuscript title: **Alfalfa yield estimation based on time series of Landsat 8 and PROBA-V images: an investigation of Machine Learning techniques and spectral-temporal features**.

All persons who meet authorship criteria are listed as authors, and all authors certify that they have participated sufficiently in the work to take public responsibility for the content, including participation in the concept, design, analysis, writing, or revision of the manuscript. Furthermore, each author certifies that this material or similar material has not been and will not be submitted to or published in any other publication before its appearance in the **Remote Sensing Applications: Society and Environment**.

Authorship contributions
Category 1
Conception and design of study: M.Azadbakht; D.Ashourloo; H.Aghighi
acquisition of data: M.Azadbakht; H.S.Shahrabi; S.Radiom
analysis and/or interpretation of data: M.Azadbakht; D.Ashourloo; H.Aghighi; A.Matkan

Category 2
Drafting the manuscript: M.Azadbakht; D.Ashourloo; H.Aghighi; S.Homayouni
revising the manuscript critically for important intellectual content: M.Azadbakht; D.Ashourloo; H.Aghighi; S.Homayouni; A.Matkan; H.S.Shahrabi; S.Radiom

Category 3
Approval of the version of the manuscript to be published (the names of all authors must be listed):
M.Azadbakht; D.Ashourloo; H.Aghighi; S.Homayouni; A.Matkan; H.S.Shahrabi; S.Radiom

Acknowledgements
All persons who have made substantial contributions to the work reported in the manuscript (e.g., technical help, writing and editing assistance, general support), but who do not meet the criteria for authorship, are named in the Acknowledgements and have given us their written permission to be named. If we have not included an Acknowledgements, then that indicates that we have not received substantial contributions from non-authors.

Journal Pre-proof

Dear Editor-in-Chief,

I, on behalf of the authors, would like to submit an original and unpublished research article entitled ‘**Alfalfa yield estimation based on time series of Landsat 8 and PROBA-V images: an investigation of Machine Learning techniques and spectral-temporal features’** for publication in Computers and Electronics in Agriculture.

We hereby declare that we have no conflict of interest by submitting our manuscript to this journal.

Thank you for your consideration of this manuscript.

Best regards,

Mohsen Azadbakht

Assistant Professor

Centre for Remote Sensing and GIS research

Shahid Beheshti University, Tehran, Iran.

E-mail address: m_azadbakht@sbu.ac.ir; m.azadbakht@gmail.com

Journal Pre-proof