2021

Combined Use of Landsat 8 and Sentinel 2A Imagery for Improved Sugarcane Yield Estimation in Wonji-Shoa, Ethiopia

G. Abebe

T. Tadesse
University of Nebraska-Lincoln

B. G. Awoke

Journal of the Indian Society of Remote Sensing (January 2022) 50(1):143–157
https://doi.org/10.1007/s12524-021-01466-8

RESEARCH ARTICLE

Combined Use of Landsat 8 and Sentinel 2A Imagery for Improved Sugarcane Yield Estimation in Wonji-Shoa, Ethiopia

Gebeyehu Abebe[1,2] · Tsegaye Tadesse[3] · Berhan Gessesse[2,4]

Received: 15 September 2020 / Accepted: 21 November 2021 / Published online: 2 December 2021

Abstract

In this study, a support vector regression (SVR) approach based on a radial basis function was used for estimating sugarcane yield in the Wonji-Shoa sugarcane plantation (Ethiopia) combining Landsat 8 (L8) and sentinel 2A (S2A) data. Vegetation Indices(VIs) involving visible, near-infrared, and shortwave infrared bands were calculated from the L8 and S2A sensor observations, and seasonal cumulative values were computed for the period June to October in the 9th month and June to November in the 10th month of the year for 2016/17 to 2018/19 cropping seasons. Sugarcane yield was predicted using the SVR, Multilayer perceptron neural network (MLPNN), and Multiple linear regression (MLR) methods. Then, a tenfold cross-validation approach was implemented for the performance evaluation. The results showed significant correlations between sugarcane yield and cumulative values of VIs computed during the 10th month in the growing season. The results also revealed that the estimation accuracy of sugarcane was better using the combined L8 and S2A (RMSE = 12.95 t/ha, and MAE = 10.14 t/ha) than using the S2A data alone (RMSE = 14.71 t/ha, and MAE = 12.18 t/ha). Comparing SVR results with MLPNN and MLR disclosed that SVR outperforms the other two models in terms of prediction accuracy. Overall, this study demonstrated the successful application of the SVR in developing a model for Sugarcane yield estimation and it may provide a guideline for improving the estimations of sugarcane in the study area.

Keyword Landsat 8 · Sentinel 2A · Sugarcane · Support vector regression · Wonji-Shoa · Yield estimation

Introduction

Sugarcane (*Saccharum officinarum L.*) is a perennial crop widely grown in the tropical and subtropical regions (Miphokasap & Wannasiri, 2018) and it has become one of the important crops which supports the economy in many developing countries (Abdel-Rahman & Ahmed, 2008). Sugarcane is one of the leading biofuel crops used for producing the highest renewable energy outputs and biofuel yield per unit area. It has the lowest production costs and smallest ecosystem 'carbon payback times' as compared to other crops used for biofuel energy production (Cuadra et al., 2012; Pagani et al., 2017). Predicting sugarcane yield before harvest and understanding its yield potential is important to manage production and maximize milling efficiency (Gunnula et al., 2012) as well as support marketing strategies and industry competitiveness(Pagani et al., 2017).

Due to their repetitive and synoptic coverage over a large area, remote sensing data have been recognized as an effective tool for estimating crop yield (Ban et al., 2017; Ngie & Ahmed, 2018). In this regard, several remote sensing-based approaches such as empirical regression (Franch et al., 2015, 2019; Johnson et al., 2016), integrated agro-meteorological spectral parameters (Huang et al., 2014; Saeed et al., 2017), semi-empirical radiation use efficiency model (Liu et al., 2010; Marshall et al., 2018; Sibley et al., 2014) and assimilating and spectral data assimilation with crop model (Huang et al., 2015; Li et al., 2015) have been developed and used to crop yield

✉ Gebeyehu Abebe
gebeyehuabebe2010@gmail.com

1 Department of Natural Resources Management, Debre Berhan University, Debre Berhan, Ethiopia

2 Department of Remote Sensing, Ethiopian Space Science and Technology Institute, P.O.Box: 33679, Addis Ababa, Ethiopia

3 National Drought Mitigation Center, University of Nebraska-Lincoln, Lincoln, USA

4 Department of Geography and Environmental Studies, Kotebe Metropolitan University, Addis Ababa, Ethiopia

estimation. However, empirical regression methods are the most widely used remote sensing approaches (Becker-reshef et al., 2010; Franch et al., 2019) which are based on Vegetation Indices (VIs) such as seasonal maximum value (Begue et al., 2010; Nuarsa et al., 2011), seasonal mean value (Svotwa et al., 2014), or cumulative value for the growing season (Begue et al. 2010; Lai et al., 2018).

Several studies that were carried out in crop yield estimation using VIs (Lai et al., 2018; Mutanga et al., 2013) have shown the importance of cumulative NDVI over the growing season to estimate crop yield. For instance, Begue et al. (2010) developed regression models using the maximum NDVI and integrated NDVI, to estimate sugarcane yield. They found that maximum NDVI and integrated NDVI extracted from SPOT4 and SPOT5 time series during the cropping season provided similar results. Besides, Mulianga et al, (2013) used weighted NDVI values extracted from MODIS products to estimate sugarcane yield. On the other hand, Rahman & Robson (2016) developed a regression model using maximum GNDVI from Landsat data to estimate sugarcane yield. Similarly, Mutanga et al., (2013) carried out a study at the local level using cumulative NDVI of the pre-harvest season and found that the best acquisition period of satellite images for estimating sugarcane yield is about two months preceding the beginning of harvest.

Sugarcane yield could be estimated using VIs calculated from individual sensor data (Robson et al., 2012; Mutanga et al., 2013; Mulianga et al., 2013). However, combing data from multi-source satellite data can provide improved information and overcome various limitations of data from individual sensors (He et al., 2018; Helder et al., 2018). It was reported that the accuracy of crop monitoring and yield estimation was better when using VIs derived from multi-source data than using VIs from one image data alone (Amorós-lópez et al., 2013; Skakun et al., 2017, 2019; He et al., 2018). Thus, using both Landsat 8(L8) and Sentinel 2A (S2A) could help to acquire relatively high spatial resolution (10–30 m) multispectral imageries with a temporal resolution of 3–5 days (Griffiths et al., 2019).

Sugarcane is the main cash crop in the upper and middle Awash irrigation basin of Ethiopia. The accurate estimation of yield for this crop is important for agricultural management and production in the area. Although, various studies have described the use of remotely sensed data for estimating sugarcane yield (Lofton et al., 2012; Mutanga et al., 2013; Rahman & Robson, 2016), yield estimation in Ethiopia is based on conventional techniques acquired through ground-based field visits and reports. However, such reports are often subjective, expensive, and prone to large errors (Reynolds et al., 2000) and data compilation and analysis are only completed several months after harvest (Dempewolf et al., 2014). As a result, the use of these data for decision-making and planning regarding production shortages or surplus is limited. Reliable crop yield estimation is critical for developing effective agricultural and food policies at a local, regional, and global level (Lai et al., 2018). In this respect, remotely sensed data becomes an important tool to provide better yield estimation. Using satellite images that can be constantly downloaded, crop yield forecasts can be produced earlier than conventional estimates and can be updated often throughout the growing season (Svotwa et al., 2013). In this regard, multi-source image data with high spatial and temporal resolutions can provide an opportunity to estimate sugarcane yield efficiently and effectively over large areas.

In this study, we used a support vector regression (SVR) approach using combined sensor observations from L8 and S2A vegetation indices (VIs) for predicting ratoon sugarcane yield before the harvest period. Series of VIs involving visible, NIR, and SWIR bands were calculated from the combined sensor observations and ratoon cane yield was predicted using yield data and the SVR based on a non-linear kernel Radial basis function (RBF). Studies have demonstrated that SVR is more robust than artificial neural networks owing to its efficient and good generalization capability performance (Miphokasap & Wannasiri, 2018; Chen et al., 2016). It has been shown quite recently that crop yield can be predicted using SVR (Li et al., 2009; Chen et al., 2016). SVR with RBF kernel function was used in their study and the authors stated that the model was able to predict crop yield better (Li et al., 2009; Chen et al., 2016).

To ascertain the robustness of the SVR model, a comparative analysis between SVR and other techniques (MLPNN and MLR) was used in this study. The models were expected to be useful for sugarcane estates to estimate spatial variability in crop yield using combined L8 and S2A images of at least two–three months before the beginning of harvest season. Therefore, the objectives of this study were to (i) investigate the relationships of sugarcane yield with spectral vegetation indices (VIs) extracted from Landsat 8(L8) and sentinel 2A (S2A), (ii) develop a sugarcane yield estimation model using SVR with the combined S2A-L8 data, and (iii) evaluate whether SVR model can effectively estimate sugarcane by comparing with the model estimates using MLR and MLPNN. The rest of the manuscript is organized as follows. Section 2 describes the proposed materials and methods. In Sects. 3 and 4, the results and discussion of the proposed model using a case study of the sugarcane yield dataset are presented. The conclusion is stated in Sect. 5.

Materials and Methods

Study Area

The area selected for this study was the Wonji-Shoa sugarcane plantation, which is the oldest and pioneer in the history of Ethiopia's sugar industry, producing sugar since 1954 (Girma & Awulachew, 2007). This plantation is located at 8°30′N, 39°20′E; about 108 km southeast of Addis Ababa, Ethiopia (Fig. 1). The average elevation of the site is 1540 m above the mean sea level (Firehun et al., 2013), and it covers a 12,000 hectares area (Degefa & Saito, 2017). The total size of out-growers and the estate grow sugarcane farm areas on 7000 ha and 5000 ha, respectively Furrow irrigation system is used to water a total land area of 7022.24 ha out of which 1118.67 ha, is owned by seven local cooperative farmer associations. The study area has a semi-arid climate and receives a mean annual rainfall of 831.2 mm, mean annual maximum and minimum temperatures of 27.6 °C and 15.2 °C, respectively (Girma & Awulachew, 2007). Furthermore, the area of the sample fields (ha) and other details are shown in supplementary Table B1.

Sugarcane Yield Statistics

Field level yield statistics [tons of stalks per hectare (t/ha)] for 2nd and 3rd ratoon cane crops in the study area were collected from the Wonj sugarcane research and development center for 2016/17 to 2018/19 cropping seasons. In this study, sugarcane fields ratoon in February and harvested 12 to 16 months, were selected with sizes ranging

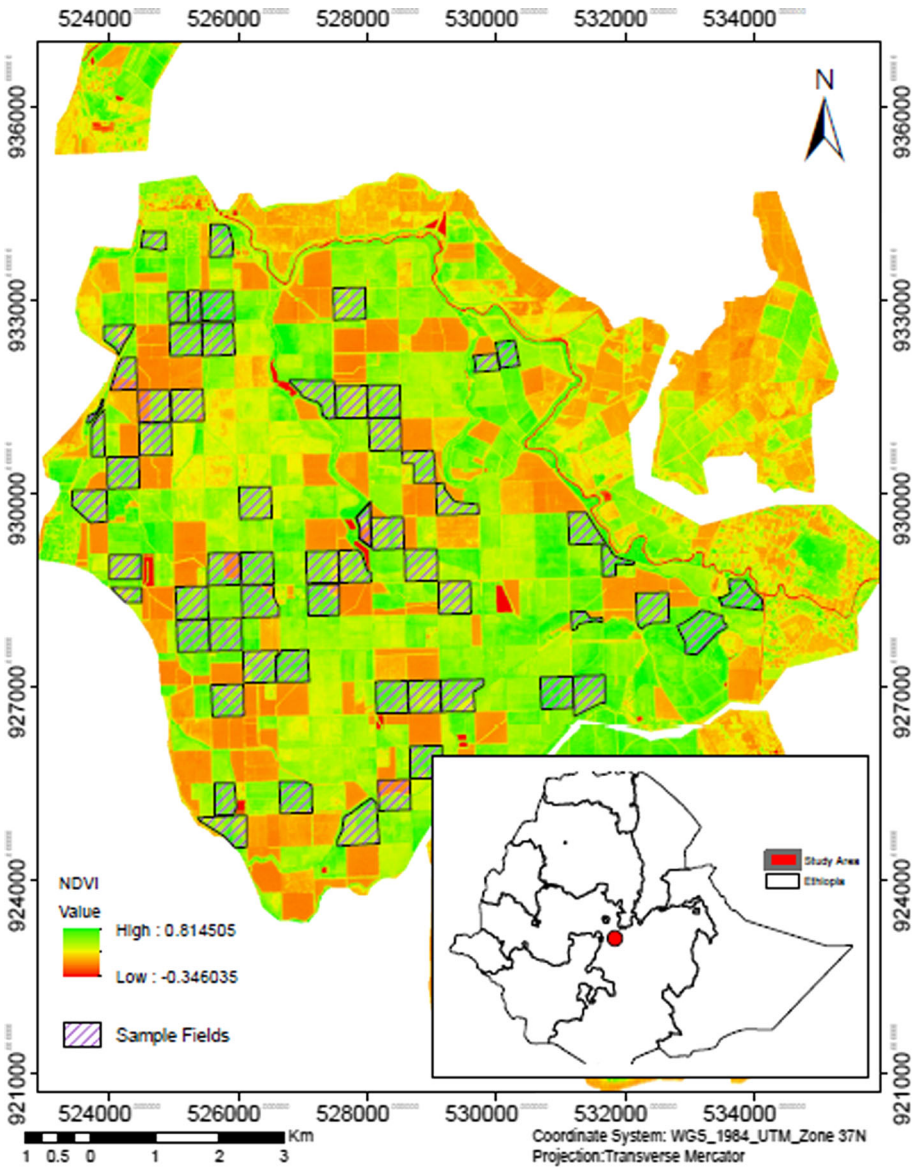

Fig. 1 Location map of the study area, superimposed on an NDVI image of Sentinel 2A (10/21/2016). The network of sample fields where field-level yield statistics were collected is highlighted in purple (hatched fields)

from 2 to 24.6 ha. In the study area, average yields were around 102 t/ha. The average age at harvest obtained for ratoon cane was 16 months. In the study area, harvest starts around November and continues until June, depending on the onset of first precipitations in summer. Planting takes place between December and June. Thus, field-level ratoon cane production statistics involved about 62, 68, and 57 sample plots in 2016/17, 2017/18, and 2018/19, respectively were used for developing an empirical yield estimation model. The cadastral map of the study area was used to extract zonal average statistics of the sample fields using their field number.

Satellite Image Acquisition and Pre-Processing

S2A and L8 Data

In this study, 30 multi-temporal L8 and 41 S2A images were acquired over the study area from the beginning of June to the end of November 2016–2018 (Table 1). The L8 images were downloaded from the USGS' Earth Explorer portal (https://earthexplorer.usgs.gov/), while the S2A images were obtained from the ESA's Sentinels data hub (https://scihub.copernicus.eu/dhus/ #/home). The L8 and S2A satellites are in circular sun-synchronous orbits with 185 km and 290 km swath widths, with 16-day and 10-day repeat cycles, respectively (Roy et al., 2019). The L8 has 9 bands with six of them (blue, green, red, NIR, SWIR-1, SWIR-2 designed for land applications at 30 m spatial resolution (Irons et al., 2012). S2A has 13 spectral bands with four bands at 10 m (visible and near-infrared), six bands at 20 m (red-edge and shortwave infrared), and three bands at 60 m spatial resolution (atmospheric correction) (Drusch et al., 2012). Approximately equivalent spectral bands that were used in this study are bands 2(blue), 3(-green), 4 (Red), 5 (NIR), 6(SWIR 1), and 7(SWIR 2) from Landsat-8, and bands 2(blue), 3(green), 4 (Red), 8 (NIR), 11(SWIR 1), and 12(SWIR 2) from Sentinel-2A.

Earlier studies suggested that the best image acquisition date for predicting sugarcane production is about two months prior to the beginning of harvesting time (Mutanga et al., 2013; Ueno et al., 2005) which supports an approach for sugarcane yield estimation using VIs from eight to ten months in the growing season(Almeida et al., 2006). For evaluating the capability of spectral VIs to estimate sugarcane yield, this study assumed that the most important relationship between VIs and sugarcane yield statistics occurred during the 9th and 10th months in the growing period. The seasonal cumulative values were then calculated from the multi-temporal VIs for five to six extended months surrounding the prediction dates. Figure 2 describes sugarcane phenological development in relation to the dates of image acquisition for the present study.

Image Pre-Processing

The pre-processing of images in this study included geometric co-registration, atmospheric correction, and adjustment for surface reflectance differences. Image registration is required to combine time series L8 and S2A data (Storey et al., 2017; Yan et al., 2016). Hence, to confirm spatial consistency between L8 and S2A data sets, L8 images were registered to their corresponding S2A images using an automated image-to-image registration approaches based on a set of ground control points (GCPs) selected from Google Earth described by (Forkuor et al., 2017). S2A images were then resampled to 30 m resolution using the bilinear resampling technique (Zhang et al., 2018).

Atmospheric correction is required to implement the yield estimation model (Griffiths et al., 2019) and both sentinel-2A and Landsat-8 sensors' top of atmosphere (TOA) reflectance data were corrected to surface reflectance using the same algorithm to minimize biases that might occur if different algorithms were used (Zhang et al., 2018). The L8 TOA reflectance L1T and S2A TOA reflectance (L1C) products were atmospherically corrected using image-based Dark Object Subtraction (DOS1) algorithm with the Semi-automatic Classification Plugin (SCP) V 6.2.9 (Congedo, 2016) in QGIS 3.6.3 software. To improve the consistency between the S2A and L8 image data, the L8 surface reflectance was then adjusted with the S2A surface reflectance using the band-specific surface reflectance ordinary least square regression parameters stated in Zhang et al. (2018).

Vegetation Indices and the Seasonal Composites

In this study, the potential of satellite images for estimating sugarcane yield was evaluated using VIs extracted from multi-temporal L8 and S2A data. The normalized

Table 1 Sugarcane ratoon date and associated image acquisition dates

Crop year	Ratoon date	Image acquisition dates	Number of images	
			L8 OLI	S2A MSI
2016/17	February 2016	June to the end of November 2016	10	13
2017/18	February 2017	June to the end of November 2017	10	14
2018/19	February 2018	June to the end of November 2018	10	14

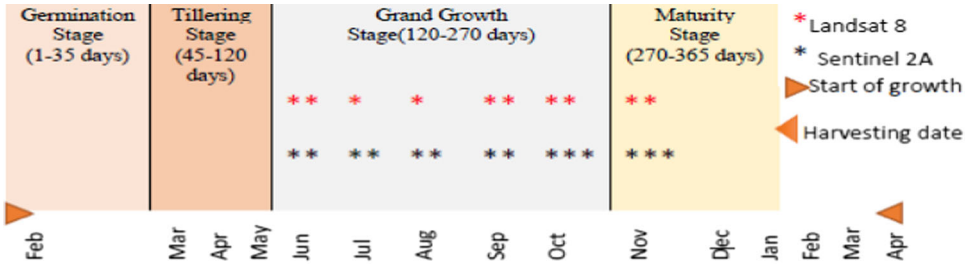

Fig. 2 Sugarcane phenological development and image acquisition scheme

difference vegetation index (NDVI) (Rouse et al., 1973), enhanced vegetation index (EVI) (Liu & Huete, 1995), soil adjusted vegetation index (SAVI) (Huete, 1988), Modified soil adjusted vegetation index (MSAVI) (Qi et al., 1994), Simple ratio (SR) (Jordan, 1969), Green Normalized difference vegetation index (GNDVI) (Gitelson et al., 1996), and Short wave infrared ratio (SIRI) (Henrich et al., 2009) were selected to analyze the relationships between VIs and sugarcane yield. The QGIS software was used to compute the VIs from the S2A and L8 images. Then, seasonal cumulative values were calculated for the period June to October in the 9th month and June to November in the 10th month for 2016/17 to 2018/19 cropping seasons. The spatial mean values of each seasonal composite within each farm field were then calculated using zonal statistics tools in ArcGIS 10.5 software and used for sugarcane yield estimation modeling. The data were also normalized and scaled to the range of 0 to 1 using min–max scaling Eq. (1), to securely apply the transfer function in the hidden (sigmoid) and output layer (linear) of MLPNN.

$$\hat{X} = \frac{X_i - X_{\min}}{X_{\max} - X_{\min}}, \tag{1}$$

where $\hat{X}$ is the normalized value, X_i is the input variable x with ith' training case, and $X_{\max}$ and $X_{\min}$ are maximum and minimum values of input variable, respectively.

Model Development and Performance Metrics

Support Vector Regression (SVR)

In this study, a machine-learning (ML) algorithm known as Support vector machines (SVMs) was applied to estimate sugarcane yield. SVMs are a supervised ML technique, which is nonlinear and is used for both classification and regression problems (Chen & Wang., 2007). The theoretical basis of the SVMs is the principle of Structural Risk Minimization (Vapnik, 1998) and support vector regression (SVR) is the extension of SVMs when they are applied to deal with regression problems (Chen & Wang., 2007).

Given a set of the training dataset, $D = \{(xi, yi|xi \in R^H, y \in R, i = 1, 2, \ldots n\}$, where, xi is a multivariate input consisting of all the independent variables,$\boldsymbol{y_i}$ is the corresponding scalar output, and $\boldsymbol{n}$ is the number of the training samples.

The SVR can be expressed by the following formula:

$$y_i = f(x, \omega) = \sum \omega.\varphi i(x) + b \tag{2}$$

where ω is the weight vector corresponding to $\varphi i(x), \varphi i(x)$ is nonlinear transformations mapping function, and b is a constant threshold. The parameters ω and b need to be estimated. The variables yi—a vector of values for the field level sugarcane yield statistics [tons of stalks per hectare (t/ha)] for the period 2016/17–2018/19, x–matrix of variables consist of NDVI, EVI, SAVI, MSAVI, GNDVI, SR and SIRI corresponding yi.

Flattens in the regression model means that one is seeking a small ω. The value of ω and b can be estimated using the structural risk minimization principle (Lagat et al., 2018) and can be expressed as;

$$(R(C) = 1/2\left(\|\omega\| + C\frac{1}{n}\sum_{i=1}^{n} L\varepsilon(f(x, \omega)) \right) \tag{3}$$

The different type of loss function termed an ε-insensitive loss is defined as:

$$L\varepsilon(f(xi, \omega)) = \begin{cases} 0, & \text{if } |y_i - f(x_i, \omega) < \varepsilon \\ |y_i - f(x_i, \omega)|, & \text{otherwise} \end{cases} \tag{4}$$

where ε is a predetermined value, $L\varepsilon(x, \omega)$ is the empirical error measured by ε-insensitive loss function.

SVR estimates weights w by minimizing the following regularized functional (Verrelst et al., 2012):

$$1/2\|\omega\|^2 + C\sum_{i=1}^{n}(\xi i + \xi_I^*) \tag{5}$$

With respect to w and $\{\xi_I^*\}_i^n = 1$, constrained to:

$$\begin{cases} y_i - f(x_i, \omega) - b \le \varepsilon + \xi i \\ f(x_i, \omega) + b - y_i \le \varepsilon + \xi_I^* \\ \xi i, \xi_I^* \ge 0 \end{cases} \tag{6}$$

where ξi and ξ_I^* are positive slack variables to deal with training samples with a prediction error larger than ε ($\varepsilon > 0$), and C is the penalization parameter applied to these. Note that C controls the trade-off between the minimization of errors and the regularization term, thus controlling the generalization capabilities.

The SVRs problem can be easily solved in its dual formulation using standard quadratic programming procedures (Verrelst et al., 2012), which yields the final solution:

Maximize W (α)

$$\widehat{y_i} = \sum_{i=1}^{n} (ai - a_i^*) k(x_i, x_j) + b, \tag{7}$$

where K is a kernel function that is expressed as the dot product of mapped examples $k(x_i, x_j) = \langle \varphi(x_i), \varphi(x_j) \rangle$. Some of the kernel functions are the linear, the polynomial, and the Gaussian function (RBF) kernel. A Gaussian radial basis kernel function (RBF) (Nanda et al., 2018; Zhang & Huihua, 2013) was used in this study, and can be expressed by the following equation:

$$k(x,\ x') = \exp\left(-\frac{\|x - x\|^2}{2\sigma^2}\right) \tag{8}$$

where $\|x - x'\|^2$ recognized as the squared Euclidean distance between the two feature vectors and sigma (σ) is the spread of the distribution used in the kernel function. Hence, a grid search algorithm with a cross-validation strategy was implemented to find the optimal parameters for SVRs (See Appendix A). For statistical modeling, we used the open-source R 3.6.2(R. Core Team, 2019) statistical computing environment with the caret package (Kuhn, 2008).

Performance Metrics

The performance of the SVR model was evaluated using the k-fold cross-validation (in this case tenfold cross-validation) approach. This procedure splits the original data randomly into k partitions, folds of equal size (Jing et al., 2017). The performance of the SVR in estimating sugarcane yield was then compared against the multiple linear regression (MLR) model (Noi et al., 2017; Oguntunde et al., 2018) and the multilayer perceptron neural networks (MLPNN) with backpropagation algorithm (Anitha & Chakravarthy, 2019; Panda et al., 2010). To analyze the performance of the regression models, three accuracy metrics of regression models were used. These are: root mean squared error (RMSE), mean absolute error (MAE), and R-square (R^2).

Results

Characteristics of Vegetation Indices Profiles

An example used to illustrate smooth VIs profiles shows the temporal pattern characteristics of sugarcane plantations for the 2016/17 growing season (Fig. 3). The cropping calendar of sugarcane in the study region varied from 12 to 16 months and the temporal intensity of VIs responses from sugarcane fields generally characterized seasonal changes of sugarcane crop phenology. This seasonal information confirmed the effectiveness of using spectral indices for tracking phenological events of sugarcane crops and was useful for assessing the crop growth conditions and yield estimates. The temporal profiles of spectral data of the sugarcane crop showed peak values surrounding the day of the year (DOY) 232 and 274 (Fig. 3), indicating the grand growth dates of the sugarcane crop. Based on the sugarcane phenology analysis, we found that the maximum intensity of the NDVI (upper bound) of L8 during the grand growth stage of the sugarcane crop was 0.63, while that of the S2A was 0.57. The maximum EVI value of the L8 was 0.61, and that of the S2A was 0.48.

Relationship Between Spectral Vegetation Indices and Sugarcane Yield

The correlation coefficients for sugarcane were calculated for cumulative values of selected VIs in the L8 and S2A images. Among L8 9th month cumulative values, EVI, NDVI, SAVI, GNDVI, and SIRI exhibited a better correlation with sugarcane yield ($0.36 \leq R^2 \leq 0.57$) for the 2016/17 cropping season. GNDVI and SIRI showed a good agreement in 2018/19 with R^2 of 0.67 and 0.68, respectively. Among the 10th month cumulative values of VIs, the best performance was obtained for NDVI, MSAVI, GNDVI, and SIRI with an R^2 between 0.52 and 0.68 in the 2016/17 cropping season. Good correlations were found between L8 VIs and sugarcane yield for the 2017/18 and 2018/19 cropping seasons ($R^2 \geq 0.42$). SIRI and MSAVI showed a lower performance ($R^2 \leq 0.26$) for one cropping season. Compared to the 9th-month cumulative values, 10th-month values showed a relatively better relationship with ratoon cane yield in the 2017/18 and 2018/19 cropping seasons (See Supplementary Table B2).

Considering, the relationship between the S2A spectral VIs and sugarcane yields, strong correlations were found between EVI, NDVI, SAVI, and GNDVI ($R^2 \geq 0.53$, $p < 0.05$) during the 9th month in the growing period for 2016/17 cropping season. Except for SR and GNDVI, there was no significant correlation between S2A VIs and sugarcane yield for the 2017/18 and 2018/19 cropping periods. Considering the 10th month period, better correlations were found between S2A VIs and sugarcane yield from the 2016/17—2018/19 cropping period ($0.44 \leq R^2 \leq 0.76$, $p < 0.05$). Compared to the 9th-month cumulative values, 10th-month values showed a relatively better relationship with ratoon cane yield. Compared to L8 VIs, the

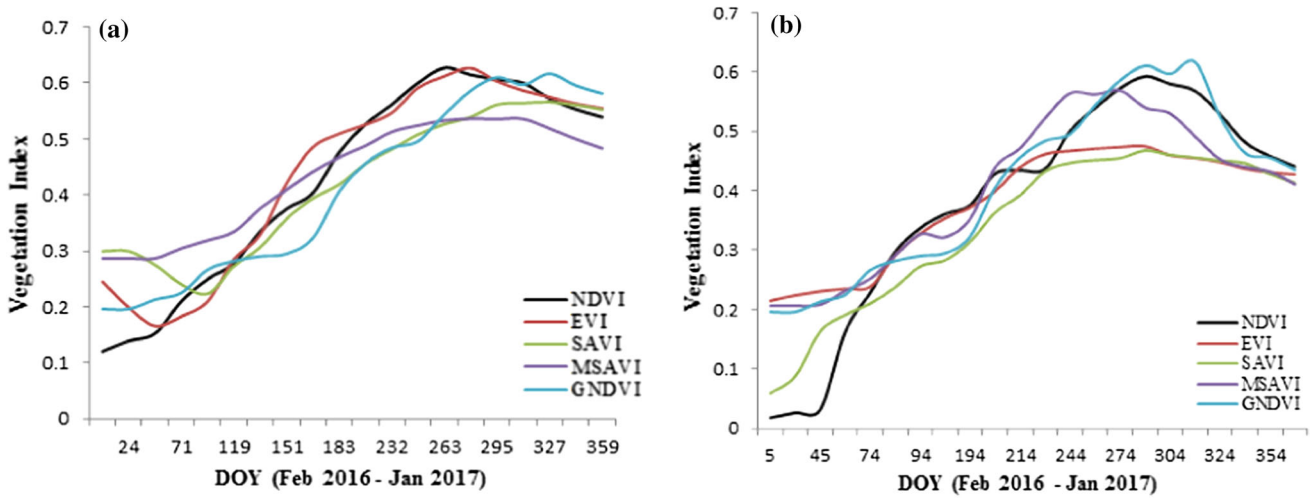

Fig. 3 Smooth VIs profiles of sugarcane plantation in the study area: **a** Landsat 8, **b** Sentinel 2A

correlation coefficients of S2A VIs with ratoon cane yield were higher (Supplementary Table B3).

The results showed that VIs based on different combinations of visible, NIR and SWIR part of the spectrum was significantly related to sugarcane yield. However, the majority of VIs used in this study tend to explore the NIR and red part of the spectral bands, while some others used either green or SWIR bands (e.g., GNDVI and SIRI). In this regard, the coefficient of determination values varies from 0.1 to 0.76, where GNDVI, SR, NDVI, and SIRI were strongly correlated with sugarcane yield. This is mainly because, each VI was initially developed in order to reduce atmospheric noise, soil background effect, or to improve a specific canopy reflectance parameter, etc. Hence, several studies combine spectral data from NIR and red bands in different ways according to their specific objectives (Xue & Su, 2017).

Estimation of Sugarcane Yield Using Multi-Temporal S2A Vegetation Indices Data

In this section, the cumulative value of VIs derived from the S2A data was used as a variable to develop the sugarcane yield estimation model using the MLR, MLPNN, and SVR algorithms. The results are presented in Table 2. There was a significant relationship between predicted and observed yield across all three years. Predictive performance varied between the three growing years and across yield estimation algorithms. The comparisons of model performance indices revealed that the SVR algorithm gave better predictive accuracy than MLR and MLPNN. MLR and MLPNN models gave poor results for both the 9th and 10th month periods. During the first period (9th month), the SVR provides the best performance with a coefficient of determination (R^2) and RMSE values of 0.78 and 18.9 t/ha in 2016/17, 0.63, and 21.19 t/ha in 2017/18, 0.79 and 15.12 t/ha in 2018/19, respectively. The value of MAE ranged from 12.37 to 16.65 t/ha. There was a wider range of predicted yield values in 2017/18 than in 2016/17 and 2018/19, probably because of the quality of the image data. During the later yield estimation (10th month), the best results were also obtained with the SVR model. The R^2 and RMSE values were 0.62 and 14.74 t/ha for 2016/2017, 0.74 and 16.74 t/ha for 2017/18, 0.77 and 12.65 t/ha for 2018/2019, respectively. The MAE ranged between 9.65 and 13.73 t/ha.

The Combined Use of S2A and L8 Data for Sugarcane Yield Estimation

The predicted yield of sugarcane was obtained after the L8 and S2A data were combined. The relationship between the measured and predicted yields is shown in Fig. 4 and Table 3. There was a significant relationship between estimated and observed sugarcane yield across all three years (Table 3). The estimation accuracies of our results varied between growing years, prediction dates, and among the tested regression models. The comparisons of model performance indicators confirmed that the results were not good when using MLR and MLPNN models (Table 3). In this respect, using models based on MLR R^2 values ranged from 0.47 to 0.78 during the 9th month in the growing period for the 2016/17–2018/19 cropping season. The ranges of RMSE and MAE were 12.31 to 25.5 t/ha and 10.78 to 20.42 t/ha, respectively, for the same period. Satisfactory results were obtained during the 10th month with the RMSE and MAE of yield models, mean values equaled 17t/ha and 14.06 t/ha, respectively, for the same

Table 2 Performance of the MLR, MLPNN and SVR models based on S2A data for predicting sugarcane yield

Year	Prediction date	MLR			MLPNN			SVR(radial)		
		R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE
2016/17	9th month	0.53	23.44	18.99	0.69	20.12	17.3	0.78	18.9	16.59
2017/18	9th month	0.37	32.55	27.67	0.56	26.02	23.86	0.63	21.19	16.65
2018/19	9th month	0.45	33.64	28.62	0.75	19.73	16.45	0.79	15.12	12.37
Average		0.45	29.87	25.09	0.66	21.95	19.2	0.73	18.4	15.2
2016/17	10th month	0.54	22.63	17.73	0.61	14.73	13.16	0.62	14.74	13.17
2017/18	10th month	0.57	25.36	22.08	0.78	19.59	17.6	0.74	16.74	13.73
2018/19	10th month	0.59	25.63	19.37	0.76	17.86	14.77	0.77	12.65	9.65
Average		0.57	24.54	19.73	0.72	17.39	15.18	0.71	14.71	12.18

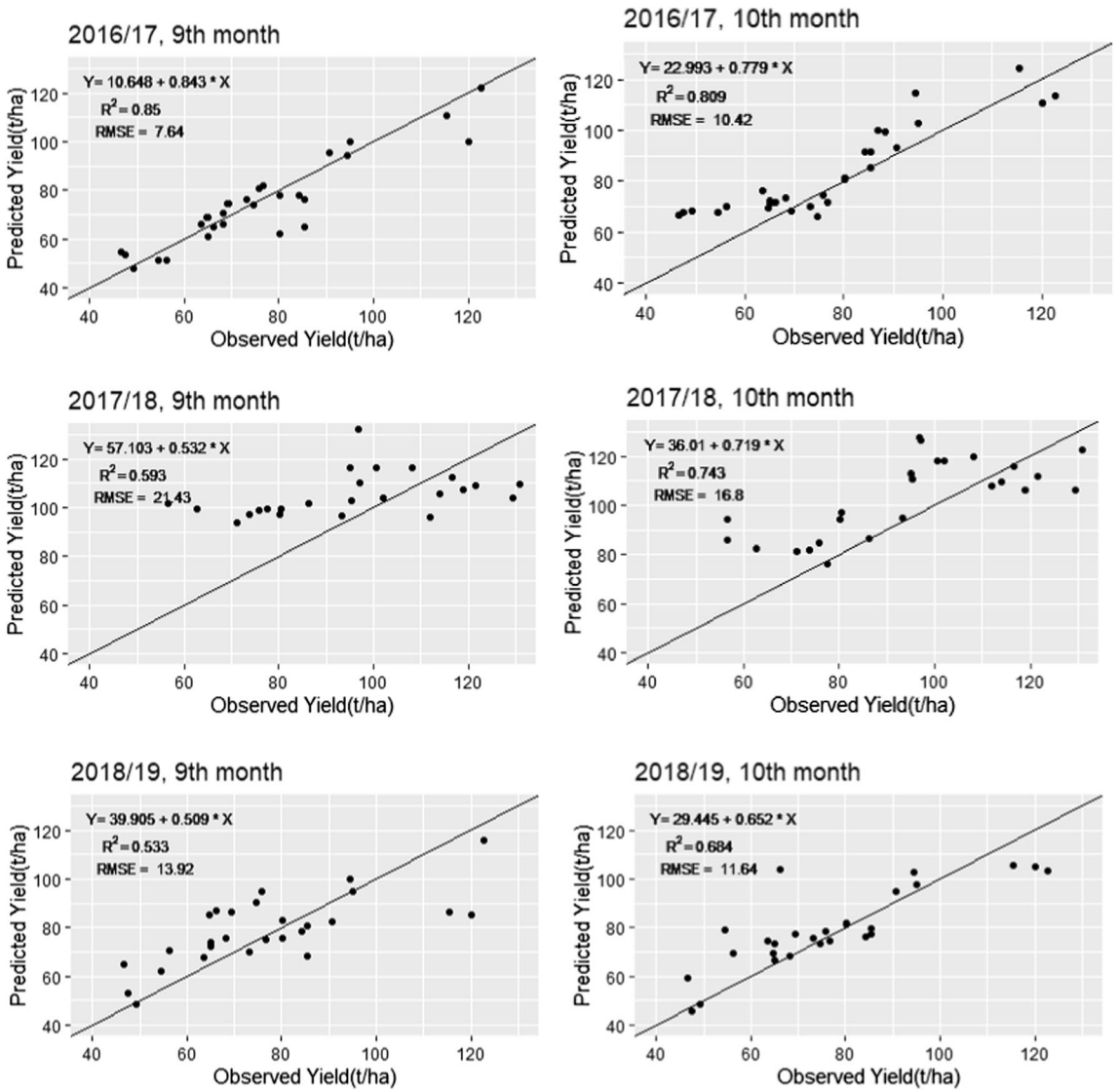

Fig. 4 Relationship between observed and predicted sugarcane yield for validation analysis using the SVR model based on the combined S2A-L8 data. *Note*: The solid line is a one-to-one line

Table 3 Performance of the MLR, MLPNN and SVR models based on combined L8-S2A data for predicting sugarcane yield

Year	Prediction date	MLR			MLPNN			SVR(radial)		
		R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE
2016/17	9th month	0.78	12.31	10.78	0.85	10.63	8.73	0.84	7.64	5.51
2017/18	9th month	0.47	25.5	20.42	0.43	26.82	22.86	0.59	21.43	17.81
2018/19	9th month	0.54	18.01	14.99	0.67	14.25	11.56	0.53	13.92	10.57
Average		0.59	18.61	15.39	0.65	17.23	14.38	0.65	14.33	11.29
2016/17	10th month	0.68	12.27	9.17	0.66	11.46	9.86	0.81	10.42	8.5
2017/18	10th month	0.56	19.36	16.53	0.72	19.93	16.08	0.74	16.8	13.78
2018/19	10th month	0.42	19.36	16.48	0.71	15.4	12.62	0.68	11.64	8.13
Average		0.55	17.00	14.06	0.69	15.59	12.85	0.74	12.95	10.14

period. The same ranges of performance indicators were obtained in the case of models based on MLPNN, with R^2, RMSE, and MAE ranging from 0.43 to 0.85, 10.63 to 26.82 t/ha and 8.73 to 22.86 t/ha, respectively, during the 9th month in the growing period. Satisfactory values of RMSE and MAE (Mean value; RMSE = 15.59 t/ha and MAE = 12.85 t/ha) were obtained during the 10th month.

The comparisons of model performance indices, demonstrated that the results were quite high when using S2A-L8 data on the SVR algorithm than MLR and MLPNN (Table 3). During the first period (9th month), the SVR provides the best performance, with a coefficient of determination values, ranged from 0.53 to 0.84. The RMSE and MAE values were 7.64 and 5.51 t/ha for 2016/2017, 21.43, and 17.81 t/ha for 2017/18, 13.92, and 10.57 t/ha for 2018/2019. However, the deviation between the predicted and observed sugarcane yield in 2017/2018 was larger than that for the other years (Table 3) and this suggests that the models did not capture well the extreme input values.

During the second period (10th month), the SVR provides the best estimates. During this period, the mean RMSE and MAE are slightly improved, respectively, from 14.33 t/ha to a minimum of 12.95 t/ha and from 11.29 t/ha to a minimum of 10.14 t/ha (Table 3). Relatively good sugarcane yield estimation performance was obtained in the 10th month, compared with the 9th-month forecasts. This suggests that the sugarcane crop status in the 10th month is determining the final sugarcane yield in the study area.

The results of sugarcane yield estimation using the cumulative value of selected L8-S2A indices as input for MLR, MLPNN, and SVR models are shown in Table 4. In overall comparison, higher statistical results were achieved by using the cumulative value of selected VIs as input for SVR than using the same indices as input for MLR and MLPNN for the estimation of sugarcane yield (Table 4). Compared with the full data sets, relatively higher accuracies were observed using 9th month cumulative value of selected VIs as input for the SVR model (R^2 = 0.85, RMSE = 7.16 t/ha, MAE = 5.22 t/ha for 2016/17, R^2 = 0.63, RMSE = 18.7 t/ha, MAE = 13.65 for 2017/18; R^2 = 0.67, RMSE = 13.21 t/ha, MAE = 8.74 t/ha for 2018/19). The difference in R^2, RMSE and MAE between full data sets and selected variables used as input for SVR model were 0.01, 0.48 t/ha, and 0.29 t/ha for 2016/17, 0.04, 2.73 t/ha, 4.16 t/ha for 2017/18, 0.14, 0.71 t/ha and 1.83 t/ha for 2018/19, respectively. The results also revealed that SVR using 10th month cumulative value of selected VI as input resulted in higher statistical results (R^2 = 0.83, RMSE = 8.39 t/ha, MAE = 6.38 t/ha for 2016/17; R^2 = 0.75, RMSE = 15.36 t/ha, MAE = 12.57 for 2017/18; R^2 = 0.77, RMSE = 10.47 t/ha, MAE = 7.76 t/ha for 2018/19). Compared with the full data sets, the result for selected VIs shows a slight improvement. In this respect, the difference in R^2, RMSE and MAE between the full data sets and selected variables used as input for the SVR model were 0.02, 2.03 t/ha and 2.12 t/ha for 2016/17, 0.01, 1.44 t/ha and 1.21 t/ha for 2017/18, 0.1, 1.17 t/ha, and 0.37 t/ha for 2018/19, respectively.

Discussion

In this study, seven VIs were used to analyze the relationships of VIs with sugarcane yield for estimating sugarcane production in the study area. The results showed that cumulative values of multi-temporal L8 VIs were significantly correlated with cane yield and the current results in our study confirmed previous findings (Begue et al. 2010; Mulianga et al., 2013 and Morel et al., 2014). Previous studies in sugarcane yield estimation have been published using Landsat TM and ETM + and have yielded mixed results. For instance, Ueno et al. (2005) reported poor correlations between sugarcane yield and VIs extracted from Landsat imageries. In contrast, the finding of Almeida et al. (2006) confirmed the usefulness of Landsat data in estimating sugarcane cane yield. Similar to L8, seven VIs were used to analyze the relationships

Table 4 Performance of the MLR, MLPNN and SVR models based on selected L8-S2A indices for predicting sugarcane yield

Year	Prediction date	Variables	MLR			MLPNN			SVR(radial)		
			R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE
2016/17	9th month	EVI, NDVI, SAVI, GNDVI	0.76	12.85	10.29	0.81	9.84	7.13	0.85	7.16	5.22
2017/18	9th month	NDVI, SR, GNDVI	0.58	23.04	18.63	0.57	23.68	17.34	0.63	18.70	13.65
2018/19	9th month	SR, GNDVI, SIRI	0.65	17.26	13.18	0.69	14.46	11.13	0.67	13.21	8.74
Average			0.66	17.72	14.00	0.69	16.00	11.87	0.72	13.02	9.20
2016/17	10th month	MSAVI, GNDVI, SIRI	0.72	13.44	10.28	0.72	15.08	12.26	0.83	8.39	6.38
2017/18	10th month	NDVI, MSAVI, SR, GNDVI	0.69	16.32	11.43	0.62	20.80	16.13	0.75	15.36	12.57
2018/19	10th month	SAVI, SR, GNDVI, SIRI	0.65	19.60	15.18	0.73	14.59	10.53	0.77	10.47	7.76
Average			0.69	16.45	12.29	0.69	16.82	12.97	0.78	11.41	8.9

between cane yield and S2A VIs. The results from this study demonstrated strong correlations between S2A VIs and cane yield during the 10th month (Table 3). Since S2A is a newly launched satellite, no empirical studies are demonstrating the usefulness of this imagery in estimating sugarcane yield at the field level. However, earlier studies have demonstrated that it could be superior to other Landsat-like multispectral imageries (Davis et al., 2019; Korhonen, et al., 2017). In their study on the comparison of S2A and L8, Davis et al. (2019) reported that S2A VIs has a slightly higher correlation than L8.

The results demonstrated that the VIs based on different combinations of visible, NIR and SWIR spectral bands were significantly related to sugarcane yield, where GNDVI, SR, NDVI, and SIRI were strongly correlated with sugarcane yield. The results demonstrated that GNDVI more accurately estimates sugarcane yield than other VIs which confirmed findings of previous studies (Robson et al., 2012). In their study on the sugarcane, Rahman & Robson (2016) found that the GNDVI performed better than the other VIs, which is in agreement with our results. Moreover, the GNDVI is an optimized index designed to reduce the effects of saturation (Gitelson et al., 1996). This increases the sensitivity of the GNDVI and explains the relatively good results obtained in this study. The NDVI and SR include the NIR and Red bands. The decrease or increase in these spectral bands influences the value of the VIs, which was highly related to sugarcane yield. Moreover, the VI based on the SWIR part of the spectral bands has also revealed the potential use for estimating sugarcane yield. This is due to, VIs extracted using SWIR spectral bands carrying water and nutrient content such as nitrogen and carbon information (Laurin et al., 2016). Hence, SIRI extracted from SWIR bands improves the estimation accuracy of sugarcane yield.

The findings of this study demonstrated that the multivariate approach using the seasonal composites of S2A VIs and SVR algorithm provides a reasonable predictive accuracy. In this regard, Li et al. (2009) and Chen et al. (2016) have used SVR approaches to crop yield and found that this method can be used to improve the estimation accuracy of crop yield. Therefore, we have used cumulative values of VIs derived from S2A and SVR algorithms to estimate ratoon cane yield at the field level. Our study is consistent with the previous findings on crop yield estimation reported by Gaffar & Sitanggang, 2019). In this study, the 10th-month cumulative values computed using the S2A VIs in combination with the SVR approach have produced higher cane yield accurate results compared to MLPNN and MLR.

Crop monitoring and yield estimation need a sufficient number of time series images during critical phases in the crop growing season. With the growing number of earth observation satellites at the moderate spatial resolution, imageries from multisource sensors can be combined to provide improved temporal coverage (Li & Roy, 2017). Several high temporal resolution satellites are available for crop monitoring and yield estimation applications (Justice et al., 2013), however, at a coarser spatial resolution. In this respect, combined use of S2A and L8 sensors data together have the potential to support regional and local coverage with a moderate spatial resolution (10–30 m) and high temporal frequency for crop monitoring and yield estimation efforts. Besides, cloud as well as other atmospheric contamination during the rainy season reduces the number of observations significantly (Skakun et al. 2019). Hence, crop monitoring and yield estimation based on the combined use of L8 and S2A might be a better alternative to improve model performance and data availability. In this study, a series of pre-processing steps are performed to obtain a combined L8-S2A surface reflectance data.

In the present study, the L8 and S2A complementary sensors data have been combined to transform data from one earth observation satellite to another satellite data. The results of the SVR, MLPNN, and MLR methods indicated that the prediction accuracy of sugarcane yield was slightly improved when combining L8 and S2A data than when using only S2A data (Table 3). Previous studies have combined multisource optical image data to estimate crop yield (Amorós-lópez et al., 2013; He et al., 2018, Skakun et al. 2019). Their findings confirmed that the integration of multi-source image data could be used to improve crop yield estimations accuracy. The results of this study are consistent with their findings. These results confirmed that the combined L8-S2A data based on the SVR algorithm provide better yield estimates for sugarcane grown in the study area. It demonstrated that the combined use of image data is an applicable method for estimating sugarcane yield in the study area.

The results of the MLR, MLPNN, and SVR approaches demonstrated that the combined L8-S2A data were highly related to sugarcane (Table 3 and Fig. 2). The estimation accuracy of sugarcane yield was higher with the SVR than the MLR and MLPNN methods. Previous studies have revealed that SVR outperforms MLR (Li et al., 2009; Oguntunde et al., 2018) and MLPNN (Chen & Wang, 2007; Ramedani et al., 2014) in estimating crop yield. This might be since the SVR technique implements the structural risk minimization principle to minimize an upper bound of the generalization error rather than reduce the training error (Chen et al. 2007). This inherent feature of SVR leads to a better estimation error than that of MLR and MLPNN. Moreover, another possible reason for the improved performance of the SVR model might have been the characteristic of the data due to the facts that statistical models require a set of data and that their robustness depends on the properties of the datasets such as the quality, number and representativeness of the available samples. So, the result of model performances for sugarcane yield indicated that the SVR model is more likely to catch the nonlinear relationship for the given data and hence, the SVR method obtained the best estimation accuracy of sugarcane yield.

The results for predicted yield were more accurate when combining the S2A imaging data with the L8 using 10th-month cumulative values than when using the 9th-month values during the growing period (Table 3 and Fig. 2). In general, higher correlation levels were clearly observed for the relationship between VIs and yield data of sugarcane during the 10th month for sugarcane cropping seasons in all years. The main reason explaining the better performance of VIs datasets for 10th month cumulative values was the use of additional spectral information in the optimal time (i.e. the transition period from grand growth to maturity stage) that is critical to improving the potential of VIs for sugarcane yield estimation. Moreover, error sources such as cloud cover in the month of July and August might have a stronger effect on earlier date images and could lower the model accuracy using 9th month cumulative values.

These results are in agreement with previous studies that have reported the optimal time for obtaining image data to be related to the final yield of sugarcane (Almeida et al., 2006). These results in the present study indicated that the combined L8 and S2A sensors data provide good estimates of sugarcane yield and the high interest in using image data acquired earlier, before the onset of maturity. The estimates of yield (at least two months before harvest) are promising and consistent with previous studies (Mutanga et al., 2013). The results from this study also demonstrated that the predicted yield was more accurate when combining the S2A imaging data with the L8 (Table 3) than when using only S2A imaging data (Table 2) with the SVR algorithm. Furthermore, the relatively low estimation error and high correlation coefficient (Table 4) demonstrated that the selected VIs improved the predictive performance of the SVR model compared to the use of the full data set. Overall, the integration of S2A with the L8 imaging data based on SVR provides reasonable estimates of sugarcane yield during the growing period.

Conclusions

The primary objective of this study was to test if the integration of L8 and S2A data with its temporal resolution of 3–5 days compared to using only S2A imaging data. Hence, the sentinel 2A vegetation indices (VIs), combined L8 and S2A (L8-S2A), and Support vector regression (SVR), Multilayer perceptron (MLPNN), and Multiple linear regression (MLR) methods were explored to determine the most accurate empirical regression equations for cane yield estimation. Based on the results of this study, we concluded that the cumulative values of combined S2A-L8 data could be effectively used to optimally estimate ratoon cane yield compared to only S2A data. Integrating multispectral optical imageries are increasingly being used for crop monitoring and yield estimation at regional to local scales. In this study, the combined L8-S2A data was used with MLR, MLPNN, and SVR, at the field level to improve the estimation accuracy of ratoon cane yield in different growth periods (9th and 10th month) and irrigation management schemes. The conclusions of this study are: (i) cumulative value of VIs computed from L8 and S2A data at 10th month in the growing period was highly correlated with ratoon cane yield; ii) The multivariate SVR algorithm with multi-temporal S2A VIs (cumulative

values) produced better results(R^2 = 0.71, RMSE = 14.71 t/ha, and MAE = 12.18 t/ha) compared to the MLR(R^2 = 0.57, RMSE = 24.54 t/ha, and MAE = 19.73 t/ha) and MLPNN(R^2 = 0.72, RMSE = 17.39 t/ha, and MAE = 15.18 t/ha) models, (iii) The combined L8-S2A method achieved more accurate sugarcane yield estimations(RMSE = 12.95 t/ha, and MAE = 10.14 t/ha) than only the S2A method(RMSE = 14.71 t/ha, and MAE = 12.18 t/ha); and (iv) Yield estimated with the VIs in the 10th month growth period was consistent with observed yield across all three years. Our findings indicated that the empirical prediction error could be significantly reduced by making use of SVR with the combined L8-S2A data in the 10th month in the sugarcane-growing season. This confirmed that the proposed SVR algorithm based on S2A-L8 data performed relatively better than the other two algorithms. Overall, the results demonstrated that a combined L8-S2A method is an effective approach for estimating ratoon cane yield and it provides guidance for optimizing irrigation management strategies for sugarcane production in the study area. Future studies will verify these results in different crops and agro-ecological regions.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12524-021-01466-8.

Acknowledgments This study was conducted with funding provided by the Ethiopian Space Science and Technology Institute (ESSTI). The field-level sugarcane yield data were provided by the Wonji Sugarcane Research and Development Centre. The Copernicus Program of the European Space Agency (ESA) and the United States Geological Survey (USGS) Landsat Program are thanked for their open-source Sentinel-2A MSI and L8 OLI datasets, respectively. Our gratitude is also extended to the R Development Team for the open-source packages for the statistical analysis.

Authors' Contribution All authors contributed to the study conception and design. Material preparation, data collection, and analysis were performed by Gebeyehu Abebe. The first draft of the manuscript was written by Gebeyehu Abebe and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding This study was funded by the Ethiopian Space Science and Technology Institute (ESSTI).

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

Abdel-Rahman, E. M., & Ahmed, F. B. (2008). The application of remote sensing techniques to sugarcane (Saccharum spp. hybrid) production a review of the literature. *International Journal of Remote Sensing, 29*(13), 753–3767. https://doi.org/10.1080/01431160701874603

Almeida, T. I. R., De Souza Filho, C. R., & Rossetto, R. (2006). ASTER and Landsat ETM+ images applied to sugarcane yield forecast. *International Journal of Remote Sensing, 27*, 4057–4069. https://doi.org/10.1080/01431160600857451

Amorós-lópez, J., Gómez-chova, L., Alonso, L., Guanter, L., Zurita-milla, R., Moreno, J., & Camps-valls, G. (2013). Multi-temporal fusion of Landsat/TM and ENVISAT/MERIS for crop monitoring. *International Journal of Applied Earth Observations and Geoinformation, 23*, 132–141. https://doi.org/10.1016/j.jag.2012.12.004

Anitha, P., & Chakravarthy, T. (2019). Agricultural crop yield prediction using artificial neural network with feed forward algorithm. *International Journal of Computer Sciences and Engineering, 6*(11), 178–181. https://doi.org/10.26438/ijcse/v6i11.178181

Ban, H. Y., Kim, K. S., Park, N. W., & Lee, B. W. (2017). Using MODIS data to predict regional corn yields. *Remote Sensing, 9*(1), 1–18. https://doi.org/10.3390/rs9010016

Becker-Reshef, I., Vermote, E., Lindeman, M., & Justice, C. (2010). A generalized regression-based model for forecasting winter wheat yields in Kansas and Ukraine using MODIS data. *Remote Sensing of Environment, 114*(6), 1312–1323. https://doi.org/10.1016/j.rse.2010.01.010

Bégué, A., Lebourgeois, V., Bappel, E., Todoroff, P., Pellegrino, A., Baillarin, F., & Siegmund, B. (2010). Spatio-temporal variability of sugarcane fields and recommendations for yield forecast using NDVI. *International Journal of Remote Sensing, 31*(20), 5391–5407. https://doi.org/10.1080/01431160903349057

Chen, H., Wu, W., & Liu, H. B. (2016). Assessing the relative importance of climate variables to rice yield variation using support vector machines. *Theoretical and Applied Climatology, 126*(1–2), 105–111. https://doi.org/10.1007/s00704-015-1559-y

Chen, K. Y., & Wang, C. H. (2007). Support vector regression with genetic algorithms in forecasting tourism demand. *Tourism Management, 28*(1), 215–226.

Congedo, L. (2016). Semi-automatic classification plugin documentation. *Release, 4*(1), 29.

Cuadra, S. V., Costa, M. H., Kucharik, C. J., Da Rocha, H. R., Tatsch, J. D., Inman-Bamber, G., Da Rocha, R. P., Leite, C. C., & Cabral, O. M. R. (2012). A biophysical model of sugarcane growth. *Gcb Bioenergy, 4*(1), 36–48. https://doi.org/10.1111/j.1757-1707.2011.01105.x

Davis, E., Wang, C., & Dow, K. (2019). Comparing S2A and L8 OLI in soil salinity detection: A case study of agricultural lands in coastal North Carolina. *International Journal of Remote Sensing, 40*(16), 6134–6153. https://doi.org/10.1080/01431161.2019.1587205

Degefa, S., & Saito, O. (2017). Assessing the impacts of large-scale agro-industrial sugarcane production on biodiversity: A case study of Wonji Shoa Sugar Estate, Ethiopia. *Agriculture, 7*(12), 99. https://doi.org/10.3390/agriculture7120099

Dempewolf, J., Adusei, B., Becker-Reshef, I., Hansen, M., Potapov, P., Khan, A., & Barker, B. (2014). Wheat yield forecasting for Punjab Province from vegetation index time series and historic crop statistics. *Remote Sensing, 6*(10), 9653–9675. https://doi.org/10.3390/rs6109653

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., et al. (2012). Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sensing of Environment, 120*, 25–36.

Firehun, Y., Tamado, T., Abera, T., & Yohannes, Z. (2013). Weed Interference in the sugarcane (*Saccharum officinarum* L.) plantations of Ethiopia. *Agriculture, Forestry and Fisheries, 2*(6), 239–247.

Forkuor, G., Dimobe, K., Serme, I., Ebagnerin, J., & Forkuor, G. (2017). Landsat-8 vs. Sentinel-2: Examining the added value of sentinel-2's red-edge bands to land-use and land-cover mapping in Burkina Faso. *Giscience and Remote Sensing*. https://doi.org/10.1080/15481603.2017.1370169

Franch, B., Vermote, E. F., Becker-reshef, I., Claverie, M., Huang, J., Zhang, J., Justice, C., & Sobrino, J. A. (2015). Improving the timeliness of winter wheat production forecast in the United States of America, Ukraine and China using MODIS data and ncar growing degree day information. *Remote Sensing of Environment, 161*, 131–148. https://doi.org/10.1016/j.rse.2015.02.014

Franch, B., Vermote, E. F., Skakun, S., Roger, J. C., Becker-Reshef, I., Murphy, E., & Justice, C. (2019). Remote sensing based yield monitoring: Application to winter wheat in United States and Ukraine. *International Journal of Applied Earth Observation and Geoinformation, 76*, 112–127. https://doi.org/10.1016/j.jag.2018.11.012

Gaffar, A. W. M., & Sitanggang, I. S. (2019). Spatial model for predicting sugarcane crop productivity using support vector regression. In *IOP Conference Series: Earth and Environmental Science*, 335(1). https://doi.org/10.1088/1755-1315/335/1/012009

Girma, M. M., & Awulachew, S. B. (2007). Irrigation Practices in Ethiopia: Characteristics of Selected Irrigation Schemes. Vol. 124. IWMI.

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

Griffiths, P., Nendel, C., & Hostert, P. (2019). Intra-annual reflectance composites from sentinel-2 and landsat for national-scale crop and land cover mapping. *Remote Sensing of Environment, 220*, 135–151. https://doi.org/10.1016/j.rse.2018.10.031

Gunnula, W., Kosittrakun, M., Righetti, T. L., Weerathaworn, P., Prabpan, M., Caldwell, J. S., & Sukchan, S. (2012). Evaluating sugarcane growth and maturity using ground-based measurements and remote sensing data. *Thai Journal of Agricultural Science, 45*(1), 17–28.

He, M., Kimball, J. S., Maneta, M. P., Maxwell, B. D., Moreno, A., Beguería, S., & Wu, X. (2018). Regional crop gross primary productivity and yield estimation using fused landsat-MODIS data. *Remote Sensing, 10*(3), 372. https://doi.org/10.3390/rs10030372

Helder, D., Markham, B., Morfitt, R., Storey, J., Barsi, J., Gascon, F., & Lewis, A. (2018). Observations and recommendations for the calibration of landsat 8 OLI and sentinel 2 MSI for improved data interoperability. *Remote Sensing, 10*(9), 1340. https://doi.org/10.3390/rs10091340

Henrich V., Jung A., Götze C., Sandow C., Thürkow D., & Gläßer C. (2009). Development of an online indices database: Motivation, concept and implementation. 6th EARSeL Imaging Spectroscopy.

Huang, J., Dai, Q., Wang, H., & Han, D. (2014). Empirical regression model using NDVI, meteorological factors for estimation of wheat yield in Yunnan, China. *CUNY Academic Works*. http://academicworks.cuny.edu/cc_conf_hic/5.

Huang, J., Tian, L., Liang, S., Ma, H., Becker-Reshef, I., Huang, Y., Su, W., Zhang, X., Zhu, D., & Wu, W. (2015). Improving winter wheat yield estimation by assimilation of the leaf area index from Landsat TM and MODIS data into the WOFOST model. *Agricultural and Forest Meteorology, 204*, 106–121. https://doi.org/10.1016/j.agrformet.2015.02.001

Huete, A. R. (1988). A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment, 25*(3), 295–309. https://doi.org/10.1016/0034-4257(88)90106-X

Irons, J. R., Dwyer, J. L., & Barsi, J. A. (2012). The next Landsat satellite: The Landsat data continuity mission. *Remote Sensing of Environment, 122*, 11–21.

Jing, W., Pengyan, Z., Hao, J., & Xiaodan, Z. (2017). Reconstructing satellite-based monthly precipitation over Northeast China using machine learning algorithms. *Remote Sensing*. https://doi.org/10.3390/rs9080781

Johnson, M. D., Hsieh, W. W., Cannon, A. J., Davidson, A., & Bédard, F. (2016). Crop yield forecasting on the Canadian Prairies by remotely sensed VIs and machine learning methods. *Agricultural and Forest Meteorology, 218*, 74–84. https://doi.org/10.1016/j.agrformet.2015.11.003

Jordan, C. F. (1969). Derivation of leaf-area index from quality of light on the forest floor. *Ecology, 50*(4), 663–666.

Justice, C. O., Román, M. O., Csiszar, I., Vermote, E. F., Wolfe, R. E., Hook, S. J., Friedl, M., Wang, Z., Schaaf, C. B., Miura, T., et al. (2013). Land and cryosphere products from Suomi NPP VIIRS: Overview and status. *Journal of Geophysical Research: Atmospheres, 118*, 9753–9765.

Korhonen, L., Packalen, P., & Rautiainen, M. (2017). Comparison of Sentinel-2 and Landsat 8 in the estimation of boreal forest canopy cover and leaf area index. *Remote Sensing of Environment, 195*, 259–274. https://doi.org/10.1016/j.rse.2017.03.021

Kuhn, M. (2008). Caret package. *Journal of Statistical Software, 28*, 1–26.

Lagat, A. K., Waititu, A. G., & Wanjoya, A. K. (2018). Support vector regression and artificial neural network approaches: case of economic growth in east africa community. *American Journal of Theoretical and Applied Statistics, 7*(2), 67–79. https://doi.org/10.11648/j.ajtas.20180702.13

Lai, Y. R., Pringle, M. J., Kopittke, P. M., Menzies, N. W., Orton, T. G., & Dang, Y. P. (2018). An empirical model for prediction of wheat yield, using time-integrated Landsat NDVI. *International Journal of Applied Earth Observation and Geoinformation, 7*, 99–108. https://doi.org/10.1016/j.jag.2018.07.013

Laurin, G. V., Puletti, N., Hawthorne, W., Liesenberg, V., Corona, P., Papale, D., Chen, Q., & Valentini, R. (2016). Discrimination of tropical forest types, dominant species, and mapping of functional guilds by hyperspectral and simulated multispectral Sentinel-2 data. *Remote Sensing of Environment, 176*, 163–176.

Li, J., & Roy, D. P. (2017). A global analysis of sentinel-2A, sentinel-2B and Landsat-8 data revisit intervals and implications for terrestrial monitoring. *Remote Sensing, 9*(9), 902.

Li, R., Li, C., Xu, X., Wang, J., Yang, X., Huang, W., & Pan, Y. (2009). Winter wheat yield estimation based on support vector machine regression and multi-temporal remote sensing data. *Transactions of the Chinese Society of Agricultural Engineering, 25*(7), 114–117.

Li, Z., Jin, X., Zhao, C., Wang, J., & Xu, X. (2015). Estimating wheat yield and quality by coupling the DSSAT-CERES model and proximal remote sensing. *European Journal of Agronomy, 71*, 53–62. https://doi.org/10.1016/j.eja.2015.08.006

Liu, H. Q., & Huete, A. (1995). A feedback based modification of the NDVI to minimize canopy background and atmospheric noise. *IEEE Transaction on Geoscience and Remote Sensing, 33*(2), 457–465. https://doi.org/10.1109/36.377946

Liu, J., Pattey, E., Miller, J. R., Mcnairn, H., Smith, A., & Hu, B. (2010). Estimating crop stresses, aboveground dry biomass and yield of corn using multi-temporal optical data combined with a radiation use efficiency model. *Remote Sensing of Environment, 114*(6), 1167–1177. https://doi.org/10.1016/j.rse.2010.01.004

Lofton, J., Tubana, B. S., Kanke, Y., Teboh, J., Viator, H., & Dalen, M. (2012). Estimating sugarcane yield potential using an in-season determination of normalized difference vegetative index. *Sensors, 12*(6), 7529–7547. https://doi.org/10.3390/s120607529

Marshall, M., Tu, K., & Brown, J. (2018). Optimizing a remote sensing production efficiency model for macro-scale GPP and yield estimation in agroecosystems. *Remote Sensing of Environment, 217*, 258–271. https://doi.org/10.1016/j.rse.2018.08.001

Miphokasap, P., & Wannasiri, W. (2018). Estimations of nitrogen concentration in sugarcane using hyperspectral imagery. *Sustainability, 10*(4), 1266. https://doi.org/10.3390/su10041266

Morel, J., Todoroff, P., Bégué, A., Bury, A., Martiné, J. F., & Petit, M. (2014). Toward a satellite-based system of sugarcane yield estimation and forecasting in smallholder farming conditions: A case study on Reunion Island. *Remote Sensing, 6*(7), 6620–6635. https://doi.org/10.3390/rs6076620

Mulianga, B., Bégué, A., Simoes, M., & Todoroff, P. (2013). Forecasting regional sugarcane yield based on time integral and spatial aggregation of MODIS NDVI. *Remote Sensing, 5*(5), 2184–2199. https://doi.org/10.3390/rs5052184

Mutanga, S., Van Schoor, C., Olorunju, P. L., Gonah, T., & Ramoelo, A. (2013). Determining the best optimum time for predicting sugarcane yield using hyper-temporal satellite imagery. *Advances in Remote Sensing, 2*, 269–275. https://doi.org/10.4236/ars.2013.23029

Nanda, M. A., Seminar, K. B., Nandika, D., & Maddu, A. (2018). A comparison study of kernel functions in the support vector machine and its application for termite detection. *Information, 9*, 5.

Ngie, A., & Ahmed, F. (2018). Estimation of Maize grain yield using multispectral satellite data sets (SPOT 5) and the random forest algorithm. *South African Journal of Geomatics, 7*(1), 11–30.

Noi, P. T., Degener, J., & Kappas, M. (2017). Comparison of multiple linear regression, cubist regression, and random forest algorithms to estimate daily air surface temperature from dynamic combinations of MODIS LST data. *Remote Sensing*. https://doi.org/10.3390/rs9050398

Nuarsa, I. W., Nishio, F., & Hongo, C. (2011). Relationship between rice spectral and rice yield using MODIS data. *Journal of Agricultural Science, 3*(2), 80–88. https://doi.org/10.5539/jas.v3n2p80

Oguntunde, P. G., Lischeid, G., & Dietrich, O. (2018). Relationship between rice yield and climate variables in southwest Nigeria using multiple linear regression and support vector machine analysis. *International Journal of Biometeorology, 62*(3), 459–469.

Pagani, V., Stella, T., Guarneri, T., Finotto, G., van den Berg, M., Marin, F. R., Acutis, M., & Confalonieri, R. (2017). Forecasting sugarcane yields using agro-climatic indicators and Canegro model: A case study in the main production region in Brazil. *Agricultural Systems, 154*, 45–52. https://doi.org/10.1016/j.agsy.2017.03.002

Panda, S. S., Ames, D. P., & Panigrahi, S. (2010). Application of VIs for agricultural crop yield prediction using neural network techniques. *Remote Sensing, 2*(3), 673–696. https://doi.org/10.3390/rs2030673

Qi, J., Chehbouni, A., Huete, A. R., Kerr, Y. H., & Sorooshian, S. (1994). A modified soil adjusted vegetation index. *Remote Sensing of Environment, 48*(2), 119–126.

R. Core Team. (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing.

Rahman, M. M., & Robson, A. J. (2016). A novel approach for sugarcane yield prediction using Landsat time series imagery: A case study on Bundaberg region. *Advances in Remote Sensing, 5*, 93–102. https://doi.org/10.4236/ars.2016.52008

Ramedani, Z., Omid, M., Keyhani, A., Shamshirband, S., & Khoshnevisan, B. (2014). Potential of radial basis function based support vector regression for global solar radiation prediction. *Renewable and Sustainable Energy Reviews, 39*, 1005–1011. https://doi.org/10.1016/j.rser.2014.07.108

Reynolds, C. A., Yitayew, M., Slack, D. C., Hutchinson, C. F., Huete, A., & Petersen, M. S. (2000). Estimating crop yields and production by integrating the FAO Crop Specific Water Balance model with real-time satellite data and ground-based ancillary data. *International Journal of Remote Sensing, 21*(18), 3487–3508.

Robson, A., Abbott, C., Lamb, D., and Bramley, R. (2012). Developing sugarcane yield prediction algorithms from satellite imagery. In *Proceedings of the Australian Society of Sugar Cane Technologists*, 34(11).

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

Roy, D. P., Huang, H., Boschetti, L., Giglio, L., Yan, L., & Zhang, H. H. (2019). Landsat-8 and Sentinel-2 burned area mapping - A combined sensor multi- temporal change detection approach. *Remote Sensing of Environment, 231*, 111254. https://doi.org/10.1016/j.rse.2019.111254

Saeed, U., Dempewolf, J., Becker-Reshef, I., Khan, A., & Wajid, S. A. (2017). Forecasting wheat yield from weather data and MODIS NDVI using random forests for Punjab province, Pakistan. *International Journal of Remote Sensing, 38*(17), 4831–4854. https://doi.org/10.1080/01431161.2017.1323282

Sibley, A. M., Grassini, P., Thomas, N. E., & Cassman, K. G. (2014). Testing remote sensing approaches for assessing yield variability among maize fields. *Agronomy Journal, 106*(1), 24–32. https://doi.org/10.2134/agronj2013.0314

Skakun, S., Vermote, E., Franch, B., Roger, J. C., Kussul, N., Ju, J., & Masek, J. (2019). Winter wheat yield assessment from landsat 8 and sentinel-2 data: Incorporating surface reflectance, through phenological fitting, into regression yield models. *Remote Sensing*. https://doi.org/10.3390/rs11151768

Skakun, S., Vermote, E., Roger, J. C., & Franch, B. (2017). Combined use of Landsat-8 and Sentinel-2A images for winter crop mapping and winter wheat yield assessment at regional scale. *AIMS Geosciences, 3*(2), 163–186. https://doi.org/10.3934/geosci.2017.2.163

Storey, J., Roy, D. P., Masek, J., Gascon, F., Dwyer, J., & Choate, M. (2017). A note on the temporary misregistration of Landsat-8 operational land imager (OLI) and sentinel-2 multi spectral instrument (MSI) imagery. *Remote Sensing of Environment, 186*(2016), 121–122. https://doi.org/10.1016/j.rse.2016.08.025

Svotwa, E., Masuka, A. J., Maasdorp, B., & Murwira, A. (2014). Estimating tobacco crop area and yield in Zimbabwe using operational remote sensing and statistical techniques. *International Journal of Agricultural Research and Review, 2*(5), 84–91.

Svotwa, E., Masuka, A. J., Maasdorp, B., Murwira, A., & Shamudzarira, M. (2013). Remote sensing applications in tobacco yield estimation and the recommended research in Zimbabwe. *International Scholarly Research Network (ISRN) Agronomy*. https://doi.org/10.1155/2013/941873

Ueno, M., Kawamitsu, Y., Sun, L., Taira, E., & Maeda, K. (2005). Combined applications of NIR, RS, and GIS for sustainable sugarcane production. *Sugarcane International,* 23.

Vapnik, V. N. (1998). *Statistical Learning Theory*. Wiley.

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. https://doi.org/10.1016/j.rse.2011.11.002

Xue, J., & Su, B. (2017). Significant remote sensing vegetation indices: A review of developments and applications. *Journal of Sensors, 2017*, 17p. https://doi.org/10.1155/2017/1353691

Yan, L., Roy, D. P., Zhang, H., Li, J., & Huang, H. (2016). An automated approach for sub-pixel registration of landsat-8

operational land imager (OLI) and sentinel-2 multi spectral instrument (MSI) imagery. *Remote Sensing, 8*(6), 520.

Zhang, G., & Huihua, G. (2013). Support vector machine with a Pearson VII function kernel for discriminating halophilic and non-halophilic proteins. *Computational Biology and Chemistry, 46*, 16–22.

Zhang, H. K., Roy, D. P., Yan, L., Li, Z., Huang, H., Vermote, E., Skakun, S., & Roger, J. (2018). Characterization of Sentinel-2A and Landsat-8 top of atmosphere, surface, and nadir BRDF adjusted reflectance and NDVI differences. *Remote Sensing of Environment, 215*, 482–494. https://doi.org/10.1016/j.rse.2018.04.031