

## ESTIMATION OF SUGARCANE YIELD USING MULTI-TEMPORAL SENTINEL 2 SATELLITE IMAGERY AND RANDOM FOREST REGRESSION

Alireza Taravat <sup>1</sup>, Gebeyehu Abebe <sup>2,3</sup>, Berhan Gessesse <sup>3</sup>, Tsegaye Tadesse <sup>4</sup>

<sup>1</sup>Deimos Space UK Ltd., Airspeed 1, 151 Eighth St., Harwell Oxford, Didcot OX11 0QR, UK

<sup>2</sup>Department of Natural Resources Management, Debre Berhan University, Debre Berhan, Ethiopia

<sup>3</sup>Department of Remote sensing, Space Science and Geospatial Institute, Addis Ababa, Ethiopia

<sup>4</sup>National Drought Mitigation Centre, University of Nebraska-Lincoln, Lincoln, USA

**KEY WORDS:** Random forest regression, Recursive feature elimination, Sugarcane yield estimation, Sentinel 2

### ABSTRACT:

Advancements in remote sensing techniques have greatly enhanced crop monitoring and yield estimation, with spectral vegetation indices (VIs) serving as a key component. Our study investigates the use of Sentinel-2 data, notable for its red-edge bands, in estimating sugarcane yield in Ethiopia's Awash Basin. Utilizing 22 VIs from S2 imagery, our approach combines Random Forest (RF) regression with the Recursive Feature Elimination (RFE) algorithm to improve the accuracy of sugarcane yield predictions. The results demonstrate the superior performance of the RF-RFE method over traditional RF with full datasets and Stepwise Multiple Regression (SMR). Particularly, VIs focusing on the red-edge spectral bands of S2 - such as NDVI<sub>re1n</sub>, NDVI<sub>re2n</sub>, NDVI<sub>re3n</sub>, NDRE1 and NDRE2 - were crucial in enhancing prediction precision. These indices from the red-edge and NIR narrow bands consistently influenced yield estimations in both the Wonji-Shoa and Metehara estates. The study underscores the critical role of the RFE algorithm in optimal variable selection, reinforcing earlier findings that precise variable choice can substantially boost model accuracy. The enhanced performance of the RF model when paired with the RFE algorithm was evident, emphasizing the importance of variable selection in accurate yield predictions. Employing the Out-of-Bag RMSE (OOB\_RMSE) error estimate for evaluation, we observed variations in OOB\_RMSE performance with different RF parameters, identifying the ntree value of 500 as optimal for the studied regions. The RF-RFE model's estimations showed lower errors and higher correlation coefficients, proving its efficacy over a full dataset approach, which faced challenges with traditional VIs' saturation. Our findings align with earlier studies, highlighting the efficiency of S2's red-edge bands in diverse estimation tasks and a shift towards using freely available broadband images like S2, over hyperspectral imagery, due to reduced data redundancy and processing costs. In conclusion, our findings reveal that RF regression, particularly when integrated with the RFE algorithm, is a powerful tool in remote sensing applications. The S2 imagery is optimal VIs, predominantly from the red-edge bands, exhibit significant potential for sugarcane yield estimations. The impressive results of the RF-RFE method, evident in metrics like MAE, MAPE, RMSE, Mean percentage, and R2, advocate its invaluable role in sugarcane yield prediction, highlighting its potential for optimizing irrigation management strategies and broad-spectrum agricultural planning.

### 1. INTRODUCTION

Early prediction of crop yield at local, national, and regional levels is crucial for effective planning and decision-making (Becker-Reshef et al., 2010). Traditionally, crop yield estimations have been made using field observation and survey techniques. However, these techniques are often time-consuming, costly, subjective, and prone to large errors (Reynolds et al., 2000). Furthermore, data processing and analysis may be completed several months after the harvest date, limiting the usefulness of these data for timely decision-making and planning (Dempewolf et al., 2014). Similarly, yield estimation using weather data is associated with several problems related to the spatial distribution of weather stations (Becker-Reshef et al., 2010), necessitating the need for alternative approaches that are effective in providing accurate crop yield estimates.

Remotely sensed data, due to their synoptic and repetitive coverage over a large area, have been recognized as an imperative technique for crop yield estimation (Becker-Reshef et al., 2010; Ban et al., 2016). Since they are based on satellite images that can be constantly downloaded, yield estimates can be updated frequently during the growing season (Reynolds et al., 2000). Using Vegetation Indices (VIs), crop yield forecasts can thus be produced earlier than conventional estimates. Previous studies, such as (Franch et al., 2019), used empirical models to

estimate crop yield using VIs extracted from satellite imaging data. However, the accuracy of the predictions varied depending on how strong the correlations were between the crop yields estimated from the field data and the satellite-based VIs. In this regard, the saturation problem that affects the accuracy of the estimation is the major challenge of using VIs. Studies have demonstrated the use of VIs based on red-edge bands to improve the estimation of above-ground biomass (Franch et al., 2019), LAI (Herrmann et al., 2011), and plant chlorophyll and nitrogen content (Delegido et al., 2011) by reducing the saturation effect on traditional VIs extracted from multispectral bands. Information on such variables is very important for determining crop growth conditions and predicting crop yield.

In this respect, the recent S2 satellites, which have a high temporal frequency (i.e., 5-day revisit cycle) and spatial resolution (10m - 20m), are found to be promising in helping with crop monitoring and yield estimation. The S2 imagery has recently been used for estimating crop primary productivity (Wolanin et al., 2019), crop canopy and nitrogen content (Delloye et al., 2018), LAI and biomass (Punalekar et al., 2018), and for crop mapping and yield estimation (Jin et al., 2019). Therefore, the successful application of VIs extracted from S2 imagery for crop monitoring and yield estimation could be useful not only for yield estimation and monitoring at the local level but also for mapping crop productivity at a national and regional scale.

In recent years, several Machine Learning (ML) techniques have been employed to attain improved yield estimation for different crops. To improve crop yield estimation, a powerful method for selecting the optimal VIs is highly required because VIs are weakly correlated with crop yield or are highly correlated to each other. Recently, a data-driven ensemble learning technique called RF is increasingly being applied in the field of remote sensing for several applications including Land use/Land cover classification (Munyati, 2019), crop disease detection (Adam et al., 2017), nitrogen and chlorophyll content estimation (Abdel-Rahman et al., 2012), and LAI and biomass estimation (Pandit et al., 2018). Previous studies have also used the RF regression for predicting crop yield (Saeed et al., 2017).

The robustness and high accuracy for estimation, high computational speed, and capacity to rank the variables based on their importance makes the RF algorithm one of the best approaches for classification and regression problems in remote sensing. RF can overcome the limitations of other ML methods such as the black box constraint in ANNs, helps to select the optimal variables and can reduce the dimensionality of the dataset. The RF algorithm can be used for both developing models and variables selection. Besides, RF algorithm is less sensitive to over-fitting and runs efficiently on large datasets, and it has fewer parameters as compared with other ML approaches such as ANNs and Support Vector Machine (SVM) algorithms.

In this study, we evaluated the performance of VIs computed from S2 imaging data ( $n = 22$ ) to estimate sugarcane yield in Wonji-Shoa and Metehara estates, Ethiopia. Series of VIs involving visible, NIR, red-edge, and short-wave infrared (SWIR) bands were calculated from the S2 imagery, and sugarcane yield was predicted using yield data and the RF regression algorithm. The RF regression was used in this study because of its capability to select and rank important variables for sugarcane yield estimation. Therefore, the objectives of this study were to: (1) investigate the use of VIs derived from multi-temporal S2 data for estimating sugarcane yield (t/ha), and (2) evaluate the performance of the RF regression as a variable selection and prediction method. This approach is expected to provide reliable estimates of crop yield that can contribute to improving early estimates of sugarcane yield before harvest.

## 2. MATERIALS

### 2.1 Study Area

This research was carried out in the Wonji-Shoa and Metehara sugar estates during the 2016/17-2018/19 cropping seasons (Figure 1). The Wonji-Shoa sugarcane plantation, positioned about 108 km southeast of Addis Ababa, Ethiopia, spans the coordinates  $8^{\circ} 21' - 8^{\circ} 29' N$  and  $39^{\circ} 12' - 39^{\circ} 18' E$ . Its average altitude is 1540 meters above sea level, encompassing an area of 12,000 hectares. Located downstream of the Koka dam, within the upper Awash River Basin of Ethiopia's rift valley, the Wonji-Shoa plantation experiences a semi-arid climate. It records an average annual rainfall of 831.2 mm, with mean annual temperatures ranging from a maximum of  $27.6^{\circ}C$  to a minimum of  $15.2^{\circ}C$ .

The Metehara sugarcane plantation lies at  $8^{\circ} 35' - 8^{\circ} 54' N$  and  $39^{\circ} 40' - 39^{\circ} 55' E$ , approximately 200 km southeast of Addis Ababa in the rift valley region of Ethiopia. It stands at an elevation of around 950 meters above sea level. The Chercher highlands and Mount Fentale, as well as adjacent undulating plateaus flank the Metehara area, largely comprising level floodplains along the Awash River. With a similar semi-arid

climate, the Metehara region receives an average annual rainfall of 532 mm. The temperature in this area varies between an average maximum of  $32.7^{\circ}C$  and a minimum of  $17.4^{\circ}C$ . Currently, the estate extends over more than 13,000 hectares and includes eleven principal sugarcane fields.

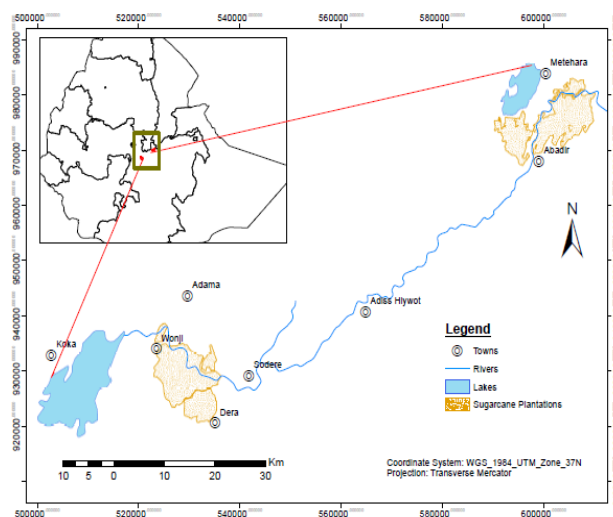


Figure 1. Map of the study area

### 2.2 Sugarcane Yield Data

Sugarcane yield data, measured in tons of stalks per hectare (t/ha), were collected for the ratoon cane crops in the aforementioned estates from the Wonji and Metehara sugarcane research and development centres for the 2016/17 to 2018/19 cropping seasons. This study focused on ratoon cane fields harvested 11 to 14 months after ratooning during: 1) the March 2016/17 to 2018/19 cropping seasons, and 2) the May 2016/17 to 2017/18 cropping seasons. The study area's shape file facilitated the extraction of zonal statistics for the sampled fields, identified by their respective field numbers in the management units of the sugarcane schemes. Table 1 presents a summary of the plot area, age at harvest, and sugarcane yield for the ratoon cane in both the Wonji-Shoa and Metehara estates.

Table 1. Descriptive statistics of sugarcane yield (t/ha)

	Mean	St.Dv.	Min	Max	CV (%)
<b>Wonji-Shoa</b>					
Plot area (ha)	11.03	6.73	0.1	70.00	61.00
Age at harvest (months)	16.00	3.00	10.00	34.00	18.75
Yield (tons/ha)	120.75	37.22	12.59	317.44	30.82
<b>Metehara</b>					
Plot area (ha)	6.17	5.30	0.1	29.70	85.89
Age at harvest (months)	16.00	3.80	0.00	47.00	23.75
Yield (tons/ha)	145.59	39.90	9.90	394.00	27.41

### 2.3 Satellite Data Acquisition and Pre-processing

For this study, Level 1C S2 satellite images, which represent the grand growth stage of sugarcane, were acquired from the USGS Earth Explorer portal. The specific tile encompassing the study area was identified as T37PEK, with a relative orbit number of

92. The research focused on sugarcane growing seasons spanning from 2016/17 to 2018/19. Five cropping seasons were selected for analysis, constrained by the availability of historical S2 data prior to 2016.

To enable reliable yield estimation using these images, atmospheric correction was essential. Level 1C S2 data were processed to Level 2A Bottom of Atmosphere (BOA) reflectance images using the Image-based Dark Object Subtraction (DOS1) algorithm, employed via the Semi-automatic Classification Plugin (SCP) V 6.2.9 (Congedo, 2021) in QGIS 3.6.3. In this study, we utilized S2 images with 10m spatial resolution bands (Blue, Green, Red, NIR,) and 20m resolution bands (Red-edge 1, Red-edge 2, Red-edge 3, NIR narrow 1, SWIR1, SWIR2). These surface reflectance images were then cropped to the study area, followed by the computation of various VIs from the selected spectral bands.

#### 2.4 Vegetation Indices and Seasonal Composites

In our research, 22 VIs, as reported in existing literature, were chosen for the development of empirical sugarcane yield estimation models. These VIs were computed from seasonal S2A images using QGIS 3.6.3 software. We then calculated the seasonal cumulative values of each VI during the grand growth stage, specifically from July to November and September to January for the March and May ratoon dates, respectively. This stage is critical as it encompasses key processes like stem elongation and yield formation. The zonal mean values of these seasonal composites within each farm field (i.e., management unit) were subsequently calculated using the Processing Toolbox in QGIS 3.6.3. These values were later utilized for modelling sugarcane yield estimation.

**Table 2.** List of Sentinel 2A VIs and their formulas used in this study

Spectral Index	Abb.	Equation
Normalized Difference Vegetation	NDVI	$(B8 - B4) / (B8 + B4)$
Enhanced Vegetation	EVI	$2.5 * \frac{B8-B4}{1+B8+6B4-7.5B2}$
Soil adjusted Vegetation	SAVI	$(1 + L) * \frac{B8-B4}{B8+B4+L}, L = 0.5$
Green NDVI	GNDVI	$(B8 - B3) / (B8 + B3)$
Nonlinear Vegetation	NLI	$(B8^2 - B4) / (B8^2 + B4)$
Re-normalized Difference Vegetation	RDVI	$(B8 - B4 / B8 + B4)^{1/2}$
Ratio Vegetation	RVI	$B8 / B4$
Green red difference	GRDI	$(B3 - B4) / (B3 + B4)$
Simple ratio(Red/green)	Red/Gr	$B4 / B3$
Red-edge Ratio1	RR1	$B8 / B5$
Red-edge chlorophyll	CIred-edge	$(B7 / B5) - 1$
Normalized difference red-edge 1	NDRE1	$(B6 - B5) / (B6 + B5)$
Normalized difference red-edge 2	NDRE2	$(B7 - B5) / (B7 + B5)$

Biochemical compounds	BC	$B8/B12$
Moisture Stress	MSI	$B11/B8$
Normalized difference Moisture	NDMI	$(B8 - B11) / (B8 + B11)$
Red-edge NDVI	NDVIre	$(B8 - B5) / (B8 + B5)$
NDVI red-edge 2	NDVIre 2	$(B8 - B6) / (B8 + B6)$
NDVI red-edge 3	NDVIre 3	$(B8 - B7) / (B8 + B7)$
NDVI red-edge 1 narrow	NDVIre 1n	$(B8a - B5) / (B8a + B5)$
NDVI red-edge 2 narrow	NDVIre 2n	$(B8a - B6) / (B8a + B6)$
NDVI red-edge 3 narrow	NDVIre 3n	$(B8a - B7) / (B8a + B7)$

### 3. METHOD

In our study, we utilized a comprehensive methodology for estimating sugarcane yield, which involved the integration of various statistical techniques and data processing methods.

We employed the RF regression method, a non-parametric and ensemble-learning approach based on decision trees. The optimization of the RF algorithm included adjusting the number of regression trees (ntree), the number of variables for growing each tree (mtry), and the number of terminal nodes (node size). The ntree values were varied from 500 to 2000 in increments of 500, while mtry was tested across all 22 predictor variables. Optimal values for ntree and mtry were determined using the Out-Of-Bag (OOB) Root Mean Square Error (OOB\_RMSE) from the training data set. Default values were used for node size for computational efficiency, and the percent increment of mean squared error (% IncMSE) was used to measure variable importance.

Additionally, we included the Recursive Feature Elimination (RFE) algorithm to select the most relevant Vegetation Indices (VIs) for yield estimation (Pullanagari et al., 2018). RFE, a wrapper-based variable selection method, is particularly effective for analysing high-dimensional datasets (Degenhardt et al., 2019). To optimize variable selection, a 10-fold cross-validation was employed, using Root Mean Square Error (RMSE) to identify the optimal predictor variables. This method was preferred over the OOB error, which can be biased in evaluating the overall error of the selection procedure. RFE was implemented using the caret R package and the RF algorithm (Kuhn, 2008).  $(1+L) *$

The predictive performance of our regression models was evaluated using the Leave-One-Out Cross-Validation (LOOCV) method. This involved sequentially removing one cropping season at a time to generate new models for predicting the yield of the omitted season. Regression analyses were performed on the training dataset, with subsequent cross-validation using an independent validation dataset. Various metrics, including Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), RMSE, Mean percent (Mean %), and the coefficient of determination ( $R^2$ ) were used to assess the models. These assessments established a direct comparison between observed

and predicted yield values and allowed for a comparison of the RF regression's performance against the SMR.

#### 4. RESULTS AND DISCUSSION

The RF regression parameters were optimized using training datasets and Out-Of-Bag Root Mean Square Error (OOB\_RMSE), considering ntree values of 500, 1000, 1500, and 2000, and exploring all 22 possible mtry variables. This optimization revealed that the ideal ntree and mtry values varied based on the prediction date and the specific growing estate. Interestingly, models generally favored mtry values above the default setting, except for the 2016/17-I season at Wonji-Shoa. Across various mtry values, the number of trees ranging from 500 to 2000 consistently produced lower prediction errors, with 500 trees often outperforming the others, indicating a less sensitive OOB\_RMSE to ntree values and suggesting stability in model performance with varying ntree values.

The importance of predictor variables, specifically Vegetation Indices (VIs), was evaluated using the percentage IncMSE, which measured the deterioration in the predictive performance of the sugarcane yield estimation models when each predictor was permuted. VIs such as NDVIre1, NDVIre3n, NDVIre2n, NDRE1, and NDRE2 emerged as significant contributors to yield estimation. The RF-RFE algorithm further refined the selection, identifying the optimal VIs that improved the RF algorithm's predictive performance. This process indicated that the accuracy of yield estimation enhanced for both Wonji-Shoa and Metehara study areas, with a significant proportion of the VIs being selected as important. Notably, a combination of red-edge bands and NIR narrow bands, as well as NDMI and MSI, yielded the lowest RMSE.

The overall performance of the RF methods was evaluated by comparing the RF-RFE with the full datasets and the SMR. The RF-RFE algorithm demonstrated superior predictive accuracy, as evidenced by lower Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE) values in comparison to the RF with full datasets. This improvement was consistent across both Wonji-Shoa and Metehara estates. The RF-RFE algorithm also outperformed the SMR in all prediction dates. The observed versus estimated sugarcane yield, represented in graphical form, further confirmed the enhanced predictive capability of the RF-RFE algorithm using selected VIs.

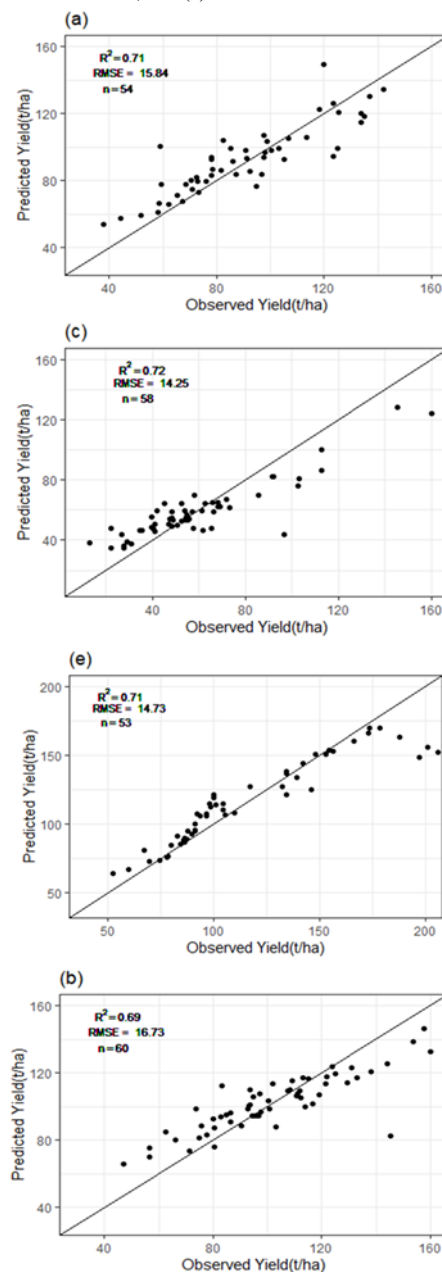
**Table 2.** Comparative analysis of Sugarcane yield estimation models utilizing S2 VIs at Wonji-Shoa

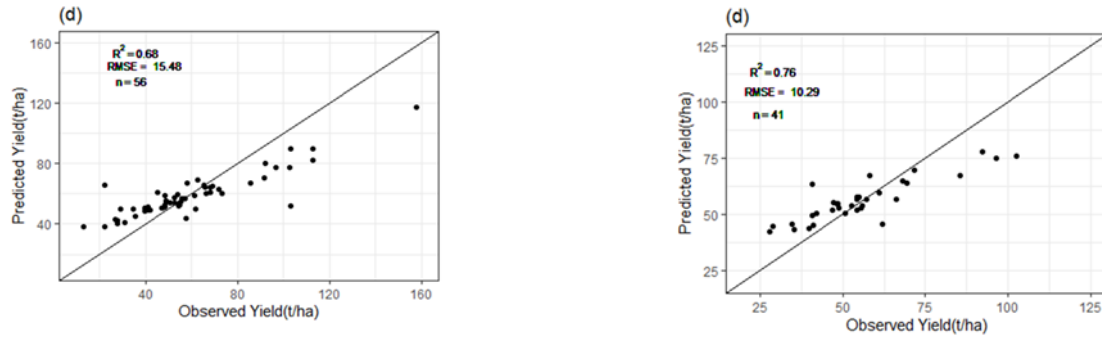
Methods		Year					Average
		2016/17-I	2016/17-II	2017/18-I	2017/18-II	2018/19-I	
SMR	MAE	17.27	17.55	18.15	18.91	18.76	18.13
	MAPE	23.76	22.01	23.75	18.94	23.46	22.38
	RMSE	23.40	22.88	24.70	25.07	25.23	24.26
	Mean %	19.38	18.95	20.46	20.76	20.89	20.09
	R <sup>2</sup>	0.57	0.68	0.65	0.47	0.59	0.59
RF	MAE	20.55	20.19	20.91	22.29	21.18	21.02
	MAPE	28.43	26.52	27.87	21.92	26.6	26.27
	RMSE	25.76	25.76	27.58	28.52	27.5	27.02
	Mean %	21.33	21.33	22.84	23.62	22.77	22.38
	R <sup>2</sup>	0.47	0.62	0.6	0.32	0.52	0.51
RF-RFE	MAE	11.56	11.77	10.44	11.24	9.51	10.9
	MAPE	12.53	11.55	22.21	25.02	15.59	17.38
	RMSE	15.84	16.73	14.25	15.48	14.73	15.41
	Mean %	13.12	13.85	11.8	12.82	12.2	12.76
	R <sup>2</sup>	0.71	0.69	0.72	0.68	0.71	0.7

**Table 3.** Comparative analysis of Sugarcane yield estimation models utilizing S2 VIs at Metehara

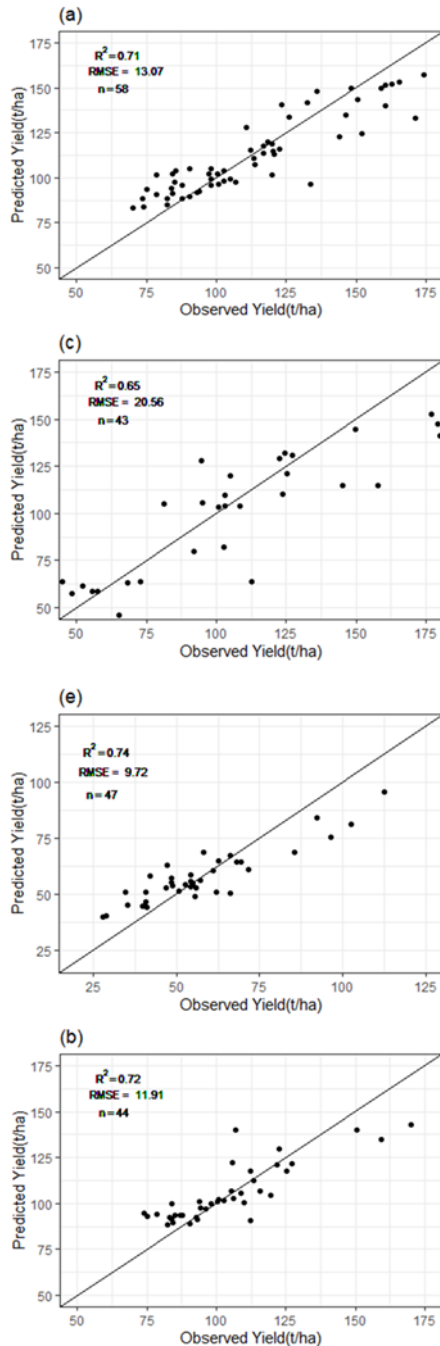
Method		Year					Average
		2016/17-I	2016/17-II	2017/18-I	2017/18-II	2018/19-I	
SMR	MAE	17.87	23.54	22.42	19.72	26.05	21.92
	MAPE	24.14	27.51	25.75	19.84	29.00	25.25
	RMSE	22.97	30.59	28.98	26.42	32.76	28.34
	Mean %	15.78	21.01	19.91	18.15	22.50	19.47
	R <sup>2</sup>	0.64	0.64	0.67	0.69	0.46	0.62
RF	MAE	18.75	24.47	25.52	22.39	26.42	23.51
	MAPE	26.26	28.83	29.37	22.51	27.93	26.98
	RMSE	23.55	31.81	31.82	29.47	32.83	29.89
	Mean %	16.18	21.85	21.86	20.24	22.55	20.53
	R <sup>2</sup>	0.59	0.57	0.53	0.61	0.43	0.55
RF-RFE	MAE	8.63	8.26	15.46	7.75	7.58	9.53
	MAPE	15.58	16.81	31.25	15.07	14.2	18.58
	RMSE	13.07	11.91	20.56	10.29	9.72	13.11
	Mean %	8.97	8.18	14.12	7.07	6.68	9.00
	R <sup>2</sup>	0.71	0.72	0.65	0.76	0.74	0.72

**Figure 2.** Illustrates the correlation between actual and forecasted sugarcane yields in the validation study conducted with RF model. The data covers five periods at Wonji-Shoa: (a) 2016/17 first season, (b) 2016/17 second season, (c) 2017/18 first season, (d) 2017/18 second season, and (e) 2018/19 first season





**Figure 3.** The correlation between measured and projected sugarcane yields in Metehara, as part of a validation study employing the RF model. The periods represented include (a) the first season of 2016/17, (b) the second season of 2016/17, (c) the first season of 2017/18, (d) the second season of 2017/18, and (e) the first season of 2018/19. Note: A solid line indicates an exact one-to-one correspondence.



## 5. CONCLUSION

RF regression is a powerful statistical model that has found many applications in remote sensing. It has been applied to such a wide variety of applications including crop monitoring and yield prediction problems such as the estimation of crop biomass, LAI and yield, and crop disease detection and prediction. Features that have added to the popularity of RF algorithms especially in the remote sensing field of studies include the feature selection methods it provides to select the optimal variables that provide the best predictive power rather than prediction only. In this study, 22 VIs derived from multi-temporal S2A imagery were considered to estimate sugarcane yield in Wonji-Shoa and Metehara estates, Ethiopia. The RF regression was explored to select optimal VIs that could be used to accurately estimate sugarcane yield. In this study, the predictive performance of RF-RFE method were evaluated and compared with the RF and SMR. The results of the study demonstrated that the optimal VIs involving the additional red edge spectral bands of S2 that include NDVI<sub>red1n</sub>, NDVI<sub>red2n</sub>, NDVI<sub>red3n</sub>, NDRE1, and NDRE2 could improve the predictions of the RF model compared to using the RF with the full datasets.

Overall, RF-RFE method was able to provide a distinct subset of VIs with a reasonable prediction performance. The results obtained showed that the RF-RFE produce better sugarcane yield estimation (MAE = 10.9 t/ha, MAPE = 17.38%, RMSE = 15.41 t/ha, Mean % = 12.76 t/ha, R<sup>2</sup> = 0.7 for Wonji-Shoa, MAE = 9.53 t/ha, MAPE = 18.58%, RMSE = 13.11 t/ha, Mean % = 9 t/ha, R<sup>2</sup> = 0.72 for Metehara). Thus, we recommend employing the implementation of the RF-RFE algorithm that provides reasonable variable selection in crop yield estimation. When RF-RFE is applied using a cross-validation approach, the resulting variable importance measures can be used reliably for optimal variable selection. In this study, the use of S2 VIs and the RF regression based on RFE is illustrated for sugarcane yield estimation. In conclusion, the results demonstrated that the RF-RFE method is an effective method for variable selection and prediction in sugarcane yield estimation purposes and can be applied directly in crop yield monitoring and prediction studies for irrigation management strategies.

## ACKNOWLEDGEMENTS

Special thanks to Wonji and Metehara sugarcane research centers for their yield data.

## REFERENCES

- Abdel-Rahman, E. M., Ahmed, F. B., and Ismail, R.: Random forest regression and spectral band selection for estimating sugarcane leaf nitrogen concentration using EO-1 Hyperion hyperspectral data, *International Journal of Remote Sensing*, 34, 712-728, 10.1080/01431161.2012.713142, 2012.
- Adam, E., Deng, H., Odindi, J., Abdel-Rahman, E. M., and Mutanga, O.: Detecting the Early Stage of *Phaeosphaeria* Leaf Spot Infestations in Maize Crop Using In Situ Hyperspectral Data and Guided Regularized Random Forest Algorithm, *Journal of Spectroscopy*, 2017, 1-8, 10.1155/2017/6961387, 2017.
- Ban, H.-Y., Kim, K., Park, N.-W., and Lee, B.-W.: Using MODIS Data to Predict Regional Corn Yields, *Remote Sensing*, 9, 16, 10.3390/rs9010016, 2016.
- Becker-Reshef, I., Vermote, E., Lindeman, M., and Justice, C.: A generalized regression-based model for forecasting winter wheat yields in Kansas and Ukraine using MODIS data, *Remote Sensing of Environment*, 114, 1312-1323, 10.1016/j.rse.2010.01.010, 2010.
- Congedo, L.: Semi-Automatic Classification Plugin: A Python tool for the download and processing of remote sensing images in QGIS, *Journal of Open Source Software*, 6, 3172, 10.21105/joss.03172, 2021.
- Degenhardt, F., Seifert, S., and Szymczak, S.: Evaluation of variable selection methods for random forests and omics data sets, *Brief Bioinform*, 20, 492-503, 10.1093/bib/bbx124, 2019.
- Delegido, J., Verrelst, J., Alonso, L., and Moreno, J.: Evaluation of Sentinel-2 red-edge bands for empirical estimation of green LAI and chlorophyll content, *Sensors (Basel)*, 11, 7063-7081, 10.3390/s110707063, 2011.
- Delloye, C., Weiss, M., and Defourny, P.: Retrieval of the canopy chlorophyll content from Sentinel-2 spectral bands to estimate nitrogen uptake in intensive winter wheat cropping systems, *Remote Sensing of Environment*, 216, 245-261, 10.1016/j.rse.2018.06.037, 2018.
- Dempewolf, J., Adusei, B., Becker-Reshef, I., Hansen, M., Potapov, P., Khan, A., and Barker, B.: Wheat Yield Forecasting for Punjab Province from Vegetation Index Time Series and Historic Crop Statistics, *Remote Sensing*, 6, 9653-9675, 10.3390/rs6109653, 2014.
- Franch, B., Vermote, E. F., Skakun, S., Roger, J. C., Becker-Reshef, I., Murphy, E., and Justice, C.: 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, 10.1016/j.jag.2018.11.012, 2019.
- Herrmann, I., Pimstein, A., Karnieli, A., Cohen, Y., Alchanatis, V., and Bonfil, D. J.: LAI assessment of wheat and potato crops by VEN $\mu$ S and Sentinel-2 bands, *Remote Sensing of Environment*, 115, 2141-2151, 10.1016/j.rse.2011.04.018, 2011.
- Jin, Z., Azzari, G., You, C., Di Tommaso, S., Aston, S., Burke, M., and Lobell, D. B.: Smallholder maize area and yield mapping at national scales with Google Earth Engine, *Remote Sensing of Environment*, 228, 115-128, 10.1016/j.rse.2019.04.016, 2019.
- Kuhn, M.: Building Predictive Models in R Using the caret Package, *Journal of Statistical Software*, 28, 10.18637/jss.v028.i05, 2008.
- Munyati, C.: Comparative performance of regression tree and parametric classification of savannah woody cover on SPOT 6 NAOMI imagery, *Remote Sensing Applications: Society and Environment*, 13, 171-182, 10.1016/j.rsase.2018.10.015, 2019.
- Pandit, S., Tsuyuki, S., and Dube, T.: Estimating Above-Ground Biomass in Sub-Tropical Buffer Zone Community Forests, Nepal, Using Sentinel 2 Data, *Remote Sensing*, 10, 601, 10.3390/rs10040601, 2018.
- Pullanagari, R., Kereszturi, G., and Yule, I.: Integrating Airborne Hyperspectral, Topographic, and Soil Data for Estimating Pasture Quality Using Recursive Feature Elimination with Random Forest Regression, *Remote Sensing*, 10, 1117, 10.3390/rs10071117, 2018.
- Punalekar, S. M., Verhoef, A., Quaife, T. L., Humphries, D., Birmingham, L., and Reynolds, C. K.: Application of Sentinel-2A data for pasture biomass monitoring using a physically based radiative transfer model, *Remote Sensing of Environment*, 218, 207-220, 10.1016/j.rse.2018.09.028, 2018.
- Reynolds, C. A., Yitayew, M., Slack, D. C., Hutchinson, C. F., Huete, A., and Petersen, M. S.: 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, 3487-3508, 10.1080/014311600750037516, 2000.
- Saeed, U., Dempewolf, J., Becker-Reshef, I., Khan, A., Ahmad, A., and Wajid, S. A.: Forecasting wheat yield from weather data and MODIS NDVI using Random Forests for Punjab province, Pakistan, *International Journal of Remote Sensing*, 38, 4831-4854, 10.1080/01431161.2017.1323282, 2017.
- Wolanin, A., Camps-Valls, G., Gómez-Chova, L., Mateo-García, G., van der Tol, C., Zhang, Y., and Guanter, L.: Estimating crop primary productivity with Sentinel-2 and Landsat 8 using machine learning methods trained with radiative transfer simulations, *Remote Sensing of Environment*, 225, 441-457, 10.1016/j.rse.2019.03.002, 2019.