

## Original Article

# Smallholder maize yield estimation using satellite data and machine learning in Ethiopia



Zhe Guo<sup>a,\*</sup>, Jordan Chamberlin<sup>b</sup>, Liangzhi You<sup>a,c</sup>

<sup>a</sup> International Food Policy Research Institute (IFPRI), Washington, DC 20005, USA

<sup>b</sup> International Maize and Wheat Improvement Center (CIMMYT), Nairobi 1041-00621, Kenya

<sup>c</sup> Macro Agriculture Research Institute (MARI), College of Economics and Management, Huazhong Agricultural University, Wuhan, Hubei 430070, China

## ARTICLE INFO

## Keywords:

Africa  
Intensification  
Sentinel-2  
Smallholder agriculture  
Yield prediction

## ABSTRACT

The lack of timely, high-resolution data on agricultural production is a major challenge in developing countries where such information can guide the allocation of scarce resources for food security, agricultural investment, and other objectives. While much research has suggested that remote sensing can potentially help address these gaps, few studies have indicated the immediate potential for large-scale estimations over both time and space. In this study we described a machine learning approach to estimate smallholder maize yield in Ethiopia, using well-measured and broadly distributed ground truth data and freely available spatiotemporal covariates from remote sensing. A neural networks model outperformed other algorithms in our study. Importantly, our work indicates that a model developed and calibrated on a previous year's data could be used to reasonably estimate maize yield in the subsequent year. Our study suggests the feasibility of developing national programs for the routine generation of broad-scale and high-resolution estimates of smallholder maize yield, including seasonal forecasts, on the basis of machine learning algorithms, well-measured ground control data, and currently existing time series satellite data.

## 1. Introduction

Reliable and location-specific information about agricultural production is a critical source of guidance for efforts to improve food security, reduce poverty, ensure well functional markets, and make progress toward the Sustainable Development Goals (Davis et al., 2017). In many developing countries, where such agricultural production information is most urgently needed, data availability is particularly limited. This is certainly the case in most of sub-Saharan Africa, where crop statistics are often only available (if at all) for large aggregate areas such as states or regions, are of uncertain accuracy, and are typically released long after the relevant production season has ended (Fritz et al., 2019). As a result, market intelligence and food balance estimates are skewed, and scarce development resources may be significantly misallocated, all of which may result in significant human and economic costs (Carletto et al., 2015). Improving the ability to map crop yields at relatively high spatial resolution and with acceptable levels of accuracy would better inform policymaking on agricultural investments, subsidies, and initiatives at local, national, and regional levels, as well as improve crop insurance planning (Eze

et al., 2020), guide farm advisory services (Zinyengere et al., 2011), and forecast food price (Delincé, 2017).

In the absence of reliable large-scale statistical data, much research has focused on improving crop yield estimation methods (Fritz et al., 2019). Such methods may be categorized into three groups. The first approach is to use empirical statistical methods. The second one is to use processed-based (i.e. mechanistic) crop growth models (Lobell and Burke, 2010; Tao et al., 2009). Thirdly, farm surveys or national agricultural censuses are also commonly used to collect agricultural data within sampling frames designed to allow inference at some level of spatial aggregation. Conventional statistical models develop a regression relationship between yield and key climate and biophysical variables such as temperature, precipitation, and solar radiation. Regression-based methods can capture the spatial variation of the yield and derive the general relationship between yield and biophysical variable measurements. However, the explanatory power of the models is more acceptable when the predictions are made within the ranges of values observed in the training samples. When predictions are made outside the training area or for different periods, performance can be degraded considerably. More importantly, previous studies have developed and tested their

\* Corresponding author.

E-mail address: [z.guo@cgiar.org](mailto:z.guo@cgiar.org) (Z. Guo).

<https://doi.org/10.1016/j.crope.2023.07.002>

Received 8 December 2022; Received in revised form 3 July 2023; Accepted 4 July 2023

2773-126X/© 2023 Published by Elsevier Ltd on behalf of Huazhong Agricultural University. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

methods over relatively limited geographical areas (Cao et al., 2021; Lobell et al., 2020; Zhang et al., 2021).

Crop growth models simulate daily crop growth using structural/process modeling approaches that have been validated elsewhere. Such approaches require large amounts of high-quality input data on soil, water, fertilizer, and land management which are limited in availability and/or quality for much of Africa and are thus not suitable for reliable yield estimation. Farm survey and agricultural census data are expensive to collect and often rely on self-reported data, which are known to have systematic bias (Lobell et al., 2020; Pradhan, 2001).

In the last two decades, remotely sensed data and spatial analytical tools have expanded substantially, making it easier and cheaper to incorporate the spatial perspective in estimating and mapping crop types and crop yields with higher temporal frequency and spatial resolution (Pradhan, 2001; Sagan et al., 2021; Tuvdendorj et al., 2019). Repeated observations over large spatial extents enable researchers to produce and update land cover products with marginally additional efforts and shorter processing times. Remote sensing technologies have been widely applied in studies of land cover and land use changes including agricultural land use at different spatial and temporal scales. Remote sensing together with Geographical Information Systems (GIS) and Global Position Systems (GPS) can be used to detect and identify characteristics of land cover types and growing conditions. Quantitative assessments of the spectrum of time series observation from space are used to monitor and map vegetation activity, density, area, and yield. U.S. Department of Agriculture (USDA) has drawn on Landsat to monitor dozens of crops including corn, wheat, soy, and cotton (Gao et al., 2017; Nguyen et al., 2020). The University of Maryland developed GEOCIF (Global Earth Observation Crop Yield and Condition Forecasting) model, a machine learning algorithm for crop yield prediction at a large scale. GEOCIF uses an ensemble-based machine-learning model to estimate and monitor agricultural yields and conditions at a large scale. The vegetation index (VI) from MODIS, together with other earth observation datasets (e.g. temperature, precipitation, and soil moisture), have been used to generate yield estimates for highly-developed agricultural regions (Becker-Reshef et al., 2018, 2019; Dempewolf et al., 2014; Franch et al., 2019). These studies demonstrate that remotely sensed data from space offer a practical means for crop yield estimation in U.S., Europe, and countries in Southern America where cropland are relatively large and homogeneous with relatively simple farming practices. Similar applications of yield estimation in sub-Saharan Africa are somewhat less successful due to the highly fragmented and complex agricultural landscapes, heterogeneous management practices, and lack of reliable training datasets (Lobell et al., 2020). Consequently, the very high-resolution data from commercial satellites that have been used to monitor crop status in Africa have generally been deployed at the landscape level, i.e. over relatively limited spatial extents, due to the high cost of image acquisition (Sagan et al., 2021).

In recent years, the rapid development of remote sensing technologies, particularly the improvements in spatial and temporal resolutions, makes it more suitable to monitor agricultural activities promptly. To be more specific, the improvements can be organized into three categories. First, there is an increased amount of satellite data with high spatial and temporal resolution and more importantly, the increasing public availability of such data is at no direct cost to users. For example, the Sentinel-1 and Sentinel-2 satellites from Europe Space Agency (ESA) provide earth observations at about 10-m resolution with 5-d temporal frequency. These publicly available datasets are now frequently used in agricultural analyses (Battude et al., 2016; Hunt et al., 2019; Jin et al., 2017). Second, the low cost of cloud computing makes it possible to process and analyze huge volumes of spatial data. Google Earth Engine (GEE) combines a multi-petabyte catalog of satellite imagery and geospatial datasets with planetary-scale analysis capabilities, which makes it available for scientists, researchers, and developers to detect changes, map trends, and

quantify differences on the earth's surface. Third, powerful machine learning algorithms are increasingly available from commercial and open-source software and computing environments. Many new libraries and packages related to machine learning and deep learning have been made publicly available in recent years. The Sentinel satellite series and advanced machine learning algorithms have opened new opportunities for monitoring agriculture outcomes like crop area and yield, even in hitherto data-sparse environments.

Burke and Lobell (2017) began using Sentinel-1 and Sentinel-2 seasonal composite images to predictively identify croplands and maize fields with a random forest classifier in GEE. A scalable satellite-based crop yield mapper was developed using crop model simulations to train statistical models (Lobell et al., 2015). Jin et al. (2017, 2019) built a multi-linear regression model for maize yield estimation using harmonic derived peak VI values of the growing season together with mean temperature and total seasonal rainfall. Additional work has shown that images from Sentinel-1, Sentinel-2, and other fine-resolution sensors can be used to generate yield estimates for smallholdings of staple cereals such as maize and wheat (Jain et al., 2016; Jin et al., 2017; Lambert et al., 2018). Most of the studies are at small scales with limited spatial coverage except Lobell et al. (2020) who expanded the study area over a larger spatial extent using a huge amount of ground truth data from various sources with different quality, resulting in promising results of  $R^2$  close to 0.40. However, challenges remain for estimating smallholder yields over larger extents in Africa. Most researchers use mean, maximum, or seasonal composite of different derived VI instead of using the entire time series of the VI profiles due to the cloud contamination of the satellite observation, even though the time series can represent biomass and crop conditions in a better way (Guerini Filho et al., 2020). Many studies focus more on spatial variation of yield distribution and the effect of different periods of time series datasets on yield estimation is not clear. Traditional regression approaches are good to assess the significance between the crop yield and selected independent variables. Machine learning based approaches are generally superior in prediction performance. Meanwhile, African countries usually have limited resources to provide large amounts of annual ground truth data for both crop types and crop yield. The inability to systematically ground-truth crop production estimates and predictions driven by data has given rise to a situation where competing models and evidence generate markedly different results.

To fill these knowledge gaps, our main research objective is to develop a machine-learning based approach to estimate and predict maize yield at a fine scale (e.g. plots) but over a large area of interest (e.g. a region or country). To do this, we deployed supervised learning algorithms using Sentinel datasets and use high-quality ground truth yield data to calibrate and validate model performance. Specifically, this paper addresses the following questions.

- How does model performance vary across different regression and machine-learning approaches?
- Which period of the time series earth observation data are most important to estimate maize yield?
- How useful are ancillary spatial datasets, such as soil properties, rainfall, and temperature variables besides satellite datasets?
- Can the model in one year be applied to predict the yield after? What is the prediction accuracy?

We investigated numerous indicators in the models including remote sensing, weather, soil, and soil moisture data from GEE to develop machine-learning models for maize yield predictions in Ethiopia. We tested a variety of prediction approaches, including regression-based predictions and several alternative machine learning algorithms, i.e. linear region model, support vector machine (SVM), regression trees, random forests, Gaussian process regression, neural networks, and deep learning networks.

## 2. Datasets and data processing

### 2.1. Study area

Ethiopia is the second most populated nation in Africa, with a current population estimated at 112 million. The economy is predominantly agrarian, with agriculture accounting for 46.3% of total GDP and more than 80% of the population residing in rural areas. Annual per capita income is around \$783 and poverty rates, while improving, remain high, particularly in rural areas. Maize is one of the major staple crops and widely planted throughout the country (Abate et al., 2015). There are two growing seasons in the country: Belg and Meher, which receive rainfall from February to June and from August to November, respectively. The country is comprised of eight agroecological zones (AEZs; Fig. 1): tropical cool arid, tropical cool semiarid, tropical cool sub-humid, tropical cool humid, tropical warm arid, tropical warm semiarid, tropical warm sub-humid, and tropical warm humid. Agriculture is relatively limited in the tropical warm arid zone but found widely elsewhere.

### 2.2. Data

Datasets used in this study include remotely sensed data, rainfall data, temperature data, soil properties, soil moisture, and yield data from 2017 to 2018. All data processing was conducted on the GEE platform and ArcGIS.

#### 2.2.1. Remote sensing datasets

Our research leveraged GEE to write generalizable modules that can be easily applied to different geographic locations, satellite datasets, and temporal specifications. Ultimately, we used 10-d intervals of Sentinel-2 images in Ethiopia to run crop yield estimations. These 10-d intervals of 2017 and 2018 datasets were generated by first mosaicking all images taken on a given day within the study region. If there were multiple day-images within a 10-d interval, a “qualityMosaic” function was

performed which chose from multiple available pixels based on cloud quality ratings. For each 10-d interval of 2017 and 2018, the individual band values related to land surface are extracted, including Blue, Green, Red, Red Edge1, Red Edge2, Red Edge3, and Near-Infra-Red (total 7 individual bands). The normalized difference vegetation index (NDVI) and green chlorophyll index (GCI) were computed along with a quality pixel indicating the presence of clouds. We aimed to evaluate the efficacy of NDVI and GCI by comparing two different processing methods. Specifically, we examined the 10-d NDVI and GCI values obtained from the original observations and the 10-d NDVI and GCI values derived after applying Savitsky-Golay (SG) curve fitting to the time series of each pixel. By analyzing these different approaches, we sought to assess the impact of the SG curve fitting technique on the accuracy and reliability of NDVI and GCI measurements (Li et al., 2011). The SG curve fitting algorithms are designed to remove the cloud contamination during the growing season since the growing seasons are largely overlapped with raining season in Ethiopia. The SG filter could remove the cloud effects by smoothing the time series of VI. SG fits successive subsets of adjacent data points with a low-degree polynomial using linear least squares.

In the end, the 10-d time series of Blue, Green, Red, Red Edge1, Red Edge2, Red Edge3, Near-Infra-Red bands, NDVI and GCI without filter, NDVI and GCI with filter in both 2017 and 2018 were produced.

#### 2.2.2. Climate data

Spatiotemporal estimates of weather parameters proved to be important predictors of maize yield. The variables include monthly precipitation, monthly minimum temperature (Tmin), and monthly maximum temperature (Tmax). The Climate Hazards Group Infrared Precipitation with Station data (CHIRPS) provides precipitation measure of the region of interests at 0.05° (ca. 5 km) of spatial resolution (Funk et al., 2015). The data incorporated satellite imagery with in-situ station data to create gridded rainfall time series for trend analysis and seasonal drought monitoring. The original CHIRPS data were aggregated into monthly precipitation data. Tmin and Tmax were acquired from the

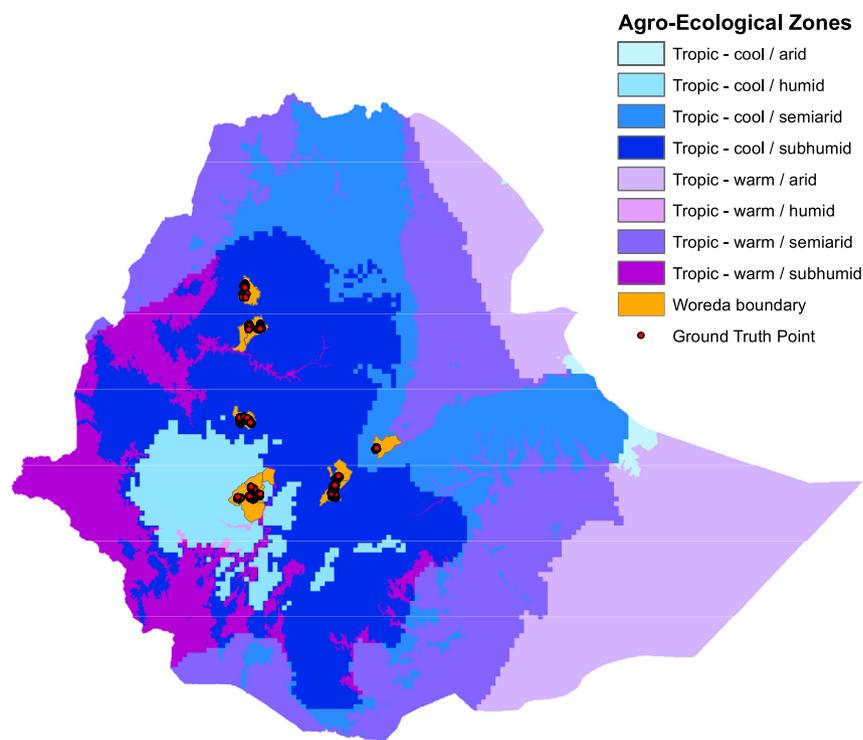


Fig. 1. Agro-ecological zones and ground truth data locations in Ethiopia.

MODIS sensor (MYD11a2 and MOD11a2 products; <https://adsweb.modaps.eosdis.nasa.gov/>). The MOD and MYD provided 8-d composites of land surface temperature at daytime and nighttime at a 1 km resolution. The data were aggregated monthly in 2017 and 2018 before applying to the analysis. The GEE platform was used to process all the climate data in Ethiopia.

### 2.2.3. Soil properties

Crop yield may be evidently affected by soil properties such as fertility level and water holding capacity. Six properties were selected in the analysis (Fischer et al., 2008). They were soil organic carbon content, soil texture, soil bulk density, soil pH, soil clay contents, and soil sand contents. The data are available at different depths from 0 to 200 cm. Based on previous studies, the soil depth of 30 cm which has the biggest impact on the crop root zones was used in our analysis. The data are available at 250 m spatial resolution.

### 2.2.4. Soil moisture

Soil moisture can affect maize production, particularly in arid areas. The NASA-USDA Global soil moisture and the NASA-USDA SMAP Global soil moisture dataset provide soil moisture information across the globe at  $0.25 \times 0.25^\circ$  spatial resolution (O'Neill et al., 2016). These datasets include surface and subsurface soil moisture (mm) and surface soil moisture were used at monthly time steps for 2017 and 2018.

### 2.2.5. Ground truth data of maize yield

The maize yield of the main (Meher) growing season was collected in 2017 and 2018 in major maize-producing areas of the Ethiopian highlands in Amhara and Oromia regions. The plots were randomly selected using a spatial sampling frame covering the major maize-producing agroecological zones. There are 469 samples in 2018 and 536 samples in 2017. In each sample plot, maize grain yield was calculated based on grain harvested at maturity from three  $5 \times 5$  m quadrats. The quadrats were placed diagonally to capture in-field variation with the first quadrat at the center of the field and remaining two offsets equidistantly from the edge along the main diagonal, as shown in Fig. 2. The moisture content of fresh grain samples was recorded and used to compute the dry weight equivalent at 12.5% moisture content. The average maize yield in 2017 was  $6,070 \text{ kg ha}^{-1}$ , while in 2018 it increased to  $6,480 \text{ kg ha}^{-1}$ .

## 3. Methodology

### 3.1. Combination of the predictor variables

There are two main categories of spatial data utilized in our study: time series data and static spatial data. The time series data consists of

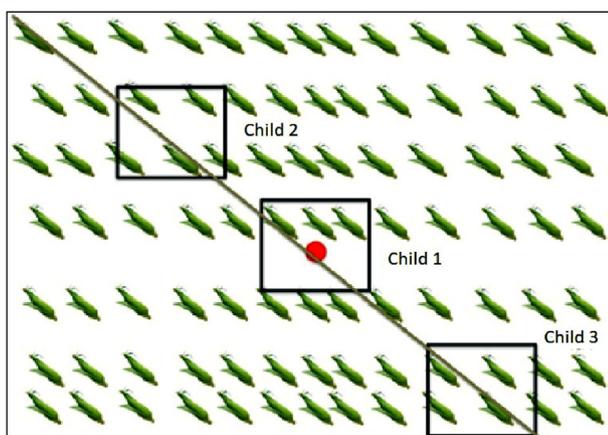


Fig. 2. Positions of quadrats in a local plot.

various datasets capturing temporal information, such as 10-d composites of original spectral bands, monthly climate variables (e.g. rainfall and temperature), monthly soil moisture, and 10-d composite of VI like NDVI and GCI. Each VI is represented by two products: SG filtered and unfiltered. The second category, static spatial data, includes soil properties and biophysical variables such as elevation. These datasets provide information that does not vary over time and remain constant throughout the year. To construct our machine learning model, we experimented with different combinations of predictors from both categories of data, aiming to identify the most informative variables for predicting crop yield. We overlaid the maize ground truth data with both time series and static remote sensing datasets. The values from the static data layers and each interval of the time series data (e.g. each month for monthly rainfall data, 10 d composite for NDVI data) were summarized and attached to the ground truth data of the maize yield based on the locations of the ground truth data. It is important to note that all predictors derived from the time series data were treated as independent variables without explicitly considering their temporal sequencing. In other words, we did not incorporate the temporal sequence information when utilizing the time series data in our models. In the end, we constructed a matrix of dimensions  $m \times n$ , where  $m$  represents the number of ground truth data points of maize yield and  $n$  represents the number of predictors, including both the time series data and static data. This matrix formed the basis for our machine learning approach to estimating crop yield. Two separate matrices were prepared for the years 2017 and 2018. Each matrix consisted of the relevant data for the respective year, including both the time series and static spatial data. These matrices were constructed to facilitate the analysis and modeling of crop yield for each specific year. For the 2017 matrix, the data from that year's time series and static spatial variables were combined into a matrix where row represented the number of ground truth data points for 2017 and column represented the total number of predictors included in the analysis. Similarly, for the 2018 matrix, the data from the corresponding time series and static spatial variables for that year were consolidated into a matrix. Below is a list of the thematic combinations:

- 10-d composite of spectral bands
- 10-d composite of spectral bands + NDVI (unfiltered)
- 10-d composite of spectral bands + NDVI (filtered)
- 10-d composite of spectral bands + GCI (unfiltered)
- 10-d composite of spectral bands + GCI (filtered)
- 10-d composite of spectral bands + climate + soil + NDVI (unfiltered)
- 10-d composite of spectral bands + climate + soil + NDVI (filtered)
- 10-d composite of spectral bands + climate + soil + GCI (unfiltered)
- 10-d composite of spectral bands + climate + soil + GCI (filtered)
- NDVI (unfiltered) + climate + soil
- NDVI (filtered) + climate + soil
- GCI (unfiltered) + climate + soil
- GCI (filtered) + climate + soil

### 3.2. Machine learning algorithms for maize yield estimation

We evaluated seven alternative advanced machine learning algorithms. All the combinations of the variables in 2017 were used as training datasets and validation datasets. For all models, 70% of the ground truth data were used for model development and calibration and the remaining 30% of the data were used as validation datasets. We applied a holdout procedure to estimate the  $R^2$  values and the root mean square errors (RMSE) to compare model performance.

#### 3.2.1. Linear regression

We evaluated several alternative regression approaches: linear, iteration linear, robust linear, and stepwise linear regression. The results

from these linear regression models were compared with the machine-learning algorithms.

### 3.2.2. Support vector machine

Support vector machine (SVM) is a popular machine learning tool for classification and regression. The objective of the SVM algorithm is to find a hyperplane that, to the best degree possible, separates data points of one class from those of another class (Hastie et al., 2009). Since the algorithm depends on kernel functions, it is considered as a non-parametric technique. Here we tested linear, polynomial, cubic, and Gaussian kernel functions at different scales to fine-tune prediction model hyper-parameters.

### 3.2.3. Regression trees

The regression tree is considered a simple and useful machine learning algorithm. In order to predict a response, the model constructs a tree-based decision-making procedure from the root, node down to a leaf node (Breiman et al., 1984). The regression tree is designed to handle large datasets. In this study, the minimum leaf sizes of 4, 8, and 16 (fine, medium, and coarse trees, respectively) were evaluated.

### 3.2.4. Gaussian process regression

The Gaussian process regression is a nonparametric probabilistic model (Hastie et al., 2009). Instead of calculating the probability distribution of parameters of a specific function, it calculate the probability distribution of all candidate functions that fit the data. The method can provide uncertainty measurements of the predictions. All the parameters are optimized using Bayesian approach. Since the method is computationally expensive, it is working well on high-dimensional but small datasets.

### 3.2.5. Random forest regression

The random forest regression approach can operationalize different learning algorithms, including boosted trees and bagged trees (Rasmussen, 2004). The random forest contains three user-defined parameters including a number of trees, number of variables at each tree, and minimum size of each node. The random forest model is designed to handle high dimensional large datasets efficiently and has been widely applied in remote sensing studies such as land classification and vegetation density estimation. In this study, both boosted and bagged trees were used.

### 3.2.6. Neural networks

Neural networks use interconnected nodes or neurons in a layer structure that mimic a human brain (Wang et al., 2016). Neural networks consist of several layers including input layers, output layers, and multiple hidden layers. Neural networks that operate on one hidden layer are known as shallow neural networks. It includes one input layer, one output layer, and one hidden layer. In each layer, there are several neurons that interconnect the different layers. Each neuron has weights that are adjusted during the learning process. The neural networks are especially large exploited to detect complex non-linear relationships between independent and dependent variables.

## 3.3. Model validation

In order to validate the model performance, we used a testing dataset randomly selected from 30% of the total ground truth points. A 30% sample is widely used as providing a good tradeoff between efficiency in testing the model accuracy and evaluating overfitting problems. The  $R^2$  values between the predicted values and the observed values were calculated and used to validate the model performance for both training dataset and validation dataset. The RMSE were also calculated and used to validate all the models.

$$R^2 = \frac{(\sum_{i=1}^n (y_i - \bar{y})(f_i - \bar{f}))^2}{\sum_{i=1}^n (y_i - \bar{y})^2 \sum_{i=1}^n (f_i - \bar{f})^2} \quad (1)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - f_i)^2} \quad (2)$$

Where  $n$  ( $i = 1, 2, \dots, n$ ) is the number of samples used for machine learning model,  $y_i$  is the observed maize yield,  $\bar{y}$  is the corresponding mean value,  $f_i$  is the predicted maize yield, and  $\bar{f}$  is the corresponding mean value. The  $R^2$  value of 1 indicates perfect prediction. A small RMSE indicates less discrepancy between the observed yield and the predicted yield.

## 4. Results

### 4.1. Comparison of the yield estimation performance by using different combinations of thematic predictors

In this study, we evaluated a total of 13 combinations of thematic predictors individually, along with the corresponding yield data from 2017. The main objective was to examine how the selection of different variable combinations or groups influenced the performance of the models. To ensure consistency, the same partitioning of developing testing data and training data was applied to each combination. Although all models were tested with different predictor combinations, they generally yielded similar results across the various combinations. To illustrate this, we presented the results from the random forest models as an example. Table 1 displayed the  $R^2$  and RMSE values for the random forest models derived from each combination using the testing datasets, providing an assessment of the performance. Furthermore, for the 10-d composites of individual spectral bands, we specifically utilized bands 2 (blue band) through band 8 (near-infrared), which are commonly employed for terrestrial land monitoring. By systematically examining the performance of different combinations of thematic predictors, we aimed to gain insights into their impact on the model's effectiveness in predicting crop yield in the context of Ethiopia. The training data were used to develop the model and the testing datasets were used to evaluate the model performance.

The first set of combinations included the 10-d composite of spectral bands alone. These combinations resulted in relatively low  $R^2$  value of 0.20, indicating that the spectral bands alone had limited explanatory power in predicting crop yield. The corresponding RMSE value was 2,069 kg ha<sup>-1</sup>, suggesting a moderate to high level of prediction error.

Incorporating the NDVI (unfiltered) or NDVI (filtered) along with the spectral bands led to higher  $R^2$  values of 0.30–0.42. This suggested that the inclusion of NDVI, which represents vegetation health, contributed to a better fit of the model. However, the improvement in model performance was relatively modest, as the RMSE values remained around 1,700–1,900 kg ha<sup>-1</sup>. Similar trends could be observed when adding the GCI (unfiltered) or GCI (filtered) to the spectral bands. The  $R^2$  values were in a similar range, indicating a limited increase in explanatory power. The corresponding RMSE values also remained in similar ranges.

Moving on to the combination of climate, soil, and vegetation indices (NDVI or GCI) resulted in higher  $R^2$  values of 0.46–0.54. These combinations showed a better improvement in model performance, suggesting that the inclusion of additional variables related to climate and soil conditions enhances the model's predictive capabilities. The corresponding RMSE values ranged from 1,522 to 1,692 kg ha<sup>-1</sup>, indicating a reduction in prediction error compared to the previous combinations.

Among these comprehensive combinations, those including NDVI (filtered) or GCI (filtered) tended to be slightly higher  $R^2$  values, indicating a stronger relationship between these variables and crop yield.

**Table 1**

The list of R<sup>2</sup> and RMSE of different combinations of thematic layers for predicting crop yield.

Combination of thematic layers	R <sup>2</sup>	RMSE (kg ha <sup>-1</sup> )
10-d composite of spectral bands	0.20	2,069
10-d composite of spectral bands + NDVI (unfiltered)	0.35	1,801
10-d composite of spectral bands + NDVI (filtered)	0.30	1,930
10-d composite of spectral bands + GCI (unfiltered)	0.26	1,942
10-d composite of spectral bands + GCI (filtered)	0.31	1,928
10-d composite of spectral bands + climate + soil + NDVI (unfiltered)	0.41	1,720
10-d composite of spectral bands + climate + soil + NDVI (filtered)	0.42	1,728
10-d composite of spectral bands + climate + soil + GCI (unfiltered)	0.38	1,762
10-d composite of spectral bands + climate + soil + GCI (filtered)	0.39	1,744
NDVI (unfiltered) + climate + soil	0.53	1,527
NDVI (filtered) + climate + soil	0.54	1,522
GCI (unfiltered) + climate + soil	0.46	1,692
GCI (filtered) + climate + soil	0.51	1,561

Abbreviations: NDVI, the normalized difference vegetation index; GCI, the green chlorophyll index.

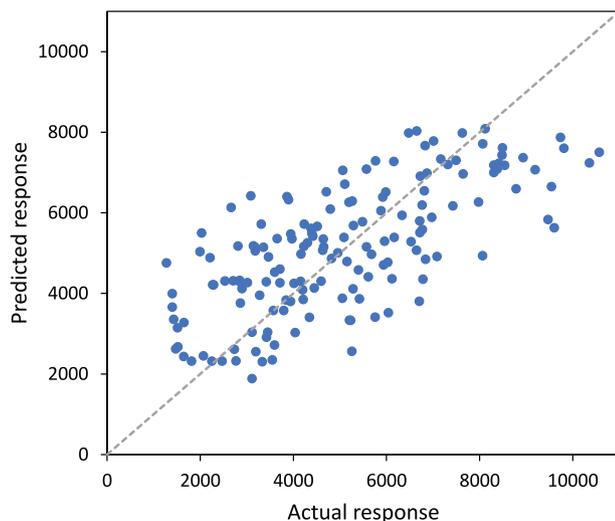
The RMSE values for these combinations also tended to be relatively lower, suggesting a more accurate prediction of crop yield.

The study further provided a scatter plot of observed and predicted yields of the testing datasets when using VI + climate + NDVI, as shown in Fig. 3. The plot illustrated that the predicted yield values using this combination of variables were relatively close to the observed yield values.

4.2. Comparison of the yield estimation using filtered VI and original VI

The research highlights that maize cultivation, occurring primarily during the rainy season, could lead to issues with cloud interference when utilizing multi-spectral Sentinel images for ground-level cropland observation. To address this challenge, the study employed the SG curve fitting method, which corrects and recalculates values by leveraging temporal information from the same geographical area. The purpose of the SG curve fitting method is to mitigate cloud effects by smoothing the VI's time series.

Interestingly, the study discovered that models employing both filtered and unfiltered (original) NDVI and GCI demonstrated comparable overall performance. However, during the validation process, the



**Fig. 3.** Scatter plot of the observed and predicted yields (Random Forest model, NDVI + climate + soil, RMSE = 1,522 kg ha<sup>-1</sup>, R<sup>2</sup> = 0.54).

filtered time series of NDVI exhibited slightly superior results. This implies that the SG curve fitting method reduced cloud contamination and enhanced the precision of yield prediction, albeit with only a marginal improvement.

4.3. Comparison of yield estimation using different machine learning algorithms

In this study, our research aimed to comprehensively assess and compare the performance of a diverse set of machine learning algorithms across six major categories: linear regression, SVM, regression tree, random forest regression, Gaussian process regression, and neural networks. Within each major category, we further explored various sub-categories, considering their unique characteristics and functionalities. While all the machine learning algorithms examined in this study were of great interest, we recognized that neural networks, in particular, warrant special attention due to their capacity for complex pattern recognition and non-linear modeling. Therefore, we have dedicated a specific paragraph within the section to discuss the neural network algorithm in more detail. By doing so, we could delve deeper into the intricacies of neural networks and provide a more comprehensive analysis of their performance within the context of our study.

To evaluate the performance of the models, we utilized a rigorous evaluation process. The training datasets were used to develop the model and the testing datasets were used to measure the model's ability to accurately predict crop yield. The key evaluation metrics employed were the RMSE from the testing datasets. It provides insights into the goodness of fit and the magnitude of errors in the predictions.

Since the combination of time series NDVI + climate + soil variables consistently demonstrated better performance, we chose to focus our subsequent analyses on evaluating the performance of the models using this specific combination of predictors. This approach allowed us to delve deeper into the predictive capabilities and understand the underlying relationships between these variables and crop yield.

A summary of the performance of various machine learning methods and their sub-groups in terms of RMSE was presented in Table 2. These algorithms were carefully selected to cover a wide range of techniques and approaches, allowing us to thoroughly assess their performance in predicting crop yield. In the linear regression category, different sub-groups were tested, including linear, interactions linear, robust, and stepwise linear. Among these, the robust linear regression showed high RMSE with the lowest RMSE value at 6,841 kg ha<sup>-1</sup>, indicating a strong non-linear relationship between yields and predictors. In addition, the stepwise linear regression failed to produce results, and therefore, no RMSE value was reported for that sub-group. The regression tree method was evaluated with fine, medium, and coarse tree sub-groups. The coarse tree sub-group demonstrated the lowest RMSE value at 1,667 kg ha<sup>-1</sup>, followed by the medium tree at 1,702 kg ha<sup>-1</sup>, and the fine tree at 1,937 kg ha<sup>-1</sup>. The SVM method was tested with linear, quadratic, and cubic SVM sub-groups. The linear SVM sub-group achieved the lowest RMSE

**Table 2**  
The RMSE scores of five major methods for maize yield estimation.

Machine learning method	Sub-group	RMSE (kg ha <sup>-1</sup> )
Linear regression	Linear	7,470
	Interactions liner	17,938
	Robust	6,841
	Stepwise linear	Failed
Regression tree	Fine tree	1,937
	Medium tree	1,702
	Coarse tree	1,667
Support vector machine	Linear	1,627
	Quadratic	1,873
	Cubic	2,123
Random forest	Ensemble (Boosted trees)	1,528
	Ensemble (Bagged trees)	1,522
Gaussian	Gaussian process regression	1,769

value at 1,627 kg ha<sup>-1</sup>, followed by the quadratic SVM at 1,873 kg ha<sup>-1</sup>, and the cubic SVM at 2,123 kg ha<sup>-1</sup>. Random forest models were examined using ensemble techniques, including boosted trees and bagged trees. Both sub-groups showed similar performance, with an RMSE value of 1,528 for boosted trees and 1,522 kg ha<sup>-1</sup> for bagged trees. Finally, the Gaussian process regression method yielded an RMSE value of 1,769 kg ha<sup>-1</sup>, indicating its predictive accuracy. Among all the methods, the random forest outperformed other yield estimation methods in this study.

In the subsequent analysis, we delved into the application of neural networks in our study. Similar to the other machine learning algorithms, we utilized datasets of 2017 for the neural networks model. The parameters of the neural networks algorithms were carefully selected to optimize their performance. The data division was performed randomly, ensuring an unbiased representation of the data in the training and validation sets. The training process utilizes the scaled conjugate gradient algorithm, which efficiently adjusts the network weights to minimize the mean squared error. This choice of performance metric allowed us to assess the accuracy of the model predictions. The calculations were performed using the MEX implementation, which enhanced the computational efficiency of the neural networks.

Our neural networks architecture consists of a single hidden layer with 128 neural nodes. This configuration strikes a balance between model complexity and computational feasibility. To train the neural networks, we employed the Levenberg-Marquardt algorithm, known for its rapid convergence compared to alternative optimization algorithms. This selection ensured that the training process was efficient and effective in finding the optimal set of weights for the neural networks. To prevent overfitting and ensure generalization, we monitored the performance of the neural networks on the validation dataset. The training process automatically halted when there was no further improvement in the model's generalization, as indicated by an increase in the mean squared error of the validation dataset. Despite setting the maximum number of epochs to 1,000, the training process stopped earlier if the desired level of generalization was achieved. This approach helped us strike a balance between model complexity and generalization, ensuring that the neural networks provided accurate and reliable predictions.

In Fig. 4, the R<sup>2</sup> values were 0.79 for training data and 0.62 for testing data. Among all the machine learning algorithms, the neural networks did not show superior performance compared to other machine learning algorithms. The model exhibited a slightly improved R<sup>2</sup> compared to other machine learning algorithms, with an increase from 0.54 to 0.62. One notable aspect of the neural networks model is its ability to generate predictions that are reasonably close to the 1:1 line, indicating its effectiveness in capturing the underlying patterns in the data. However, it is important to note that there were some discrepancies in the predictions. It is worth noting that despite the better

performance of the neural networks, one drawback is the computational time required for training and inference. Neural networks are known to be computationally intensive, particularly for large datasets and complex architectures. Therefore, this aspect should be acknowledged in the paragraph. However, given the improvements in prediction accuracy achieved by the neural networks, the slightly longer computational time may be considered a worthwhile trade-off.

#### 4.4. Out-of-sample prediction performance: new year

In the final stage of our analysis, we employed the neural networks model to forecast future yield outcomes. To accomplish this, we leveraged the model that was developed using data from the previous year (2017) to predict the yield outcomes for the following year (2018). The neural networks model, chosen for its better performance compared to other machine learning algorithms in predicting maize yields, serves as a reliable tool for this task. The neural model was developed only using the data from 2017, which included a range of predictor variables such as NDVI, climate, and soil data. These variables provided valuable insights into the growing conditions and environmental factors that influence crop productivity. Once the model was developed, its predictive ability was tested on the data from 2018. This independent dataset allowed us to assess the model performance in accurately forecasting the actual yield outcomes for the following year. By comparing the predicted yield values with the observed values in 2018, we could evaluate the model's effectiveness in capturing the variations and trends in maize yields.

In summary, the neural networks model was utilized to forecast future maize yield outcomes. By calibrating the model using data from the previous year and testing its performance on data from the subsequent year, we assessed its accuracy in predicting yield outcomes. The incorporation of key predictor variables enhanced the model's ability to capture the complex interactions between environmental factors and crop productivity, making it an asset for agricultural forecasting.

The obtained R<sup>2</sup> value of 0.43 signified that the model can account for approximately 43% of the variability in yield for a different year. This suggested that the model had a moderate ability to predict future yields, which could be valuable in offering farmers an estimation of potential crop yields in the upcoming season. To visually represent the performance of the model, Fig. 5 displays a scatter plot depicting the predicted maize yield versus the observed maize yield in 2018. This scatter plot was generated using the model that was calibrated with the data from 2017.

#### 4.5. Prediction variables and the order of relative importance

To assess the importance of different variables for maize yield estimation in Ethiopia, we examined the impact of excluding each variable from the model by measuring the increase in mean square error. By

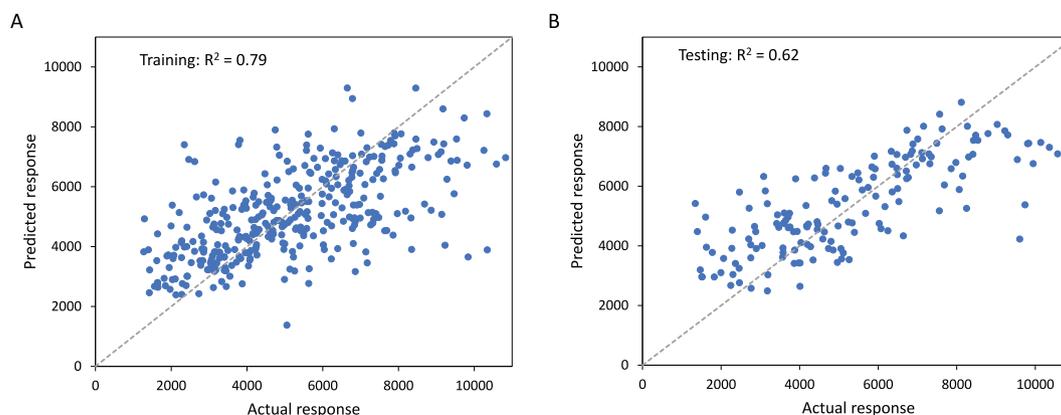


Fig. 4. The scatter plot of observed and predicted yields (kg ha<sup>-1</sup>) using the neural networks for both training (A) and testing (B).

analyzing the change in accuracy associated with the exclusion of each variable, we were able to rank them in order of importance. Our findings showed that the most important predictor variable for maize yield estimation was the NDVI, which was consistent with previous studies. Temperature and rainfall variables, specifically T<sub>min</sub> and precipitation, also proved to be important predictors of maize yield. In contrast, soil moisture was found to be less important to yield estimation when compared to other variables. This could be due to the coarser spatial resolution of the soil moisture data or a lack of sensitivity to variation among the yield plots.

Based on the expertise in the field, it is well recognized that the land management practices implemented in the previous season could have an impact on crop growth and yield. The land management practices carried out in the previous season, such as tillage, fertilization, and pest control, could influence soil properties and nutrient availability. Additionally, certain environmental factors might exhibit lag effects, meaning their influence on crop growth and yield might not be immediately evident but could manifest over time. We decided to initiate the study using time series data of the entire year. This approach allowed us to capture the complete seasonal variations in the satellite data, including variables such as NDVI, T<sub>min</sub>, precipitation, and T<sub>max</sub>. Our analysis showed that NDVI in late September had the highest importance score, indicating its critical role in maize yield estimation. Additionally, T<sub>min</sub> and precipitation in September were found to be the most important among the 12 months. T<sub>max</sub>, on the other hand, showed the greatest importance in November. These findings suggested that T<sub>min</sub> and rainfall at the beginning of the growing season were critical to maize production, while T<sub>max</sub> values were more important at the end of the growing season.

## 5. Discussion

This study aimed to test all possible combinations of variables, including individual band, satellite-derived VI, soil properties, and climate variables, to determine their impact on the accuracy of maize yield estimation (Burke and Lobell, 2017). Our findings indicate that simply including more features did not always result in better yield estimates, particularly when ground truth data were limited. Therefore, it is crucial to carefully select the features used in estimation, even though machine learning algorithms have the capability to minimize the impact of low-quality observations. We have found that the individual bands exhibited poor performance as predictors and considering the time series of individual bands would lead to a substantial increase in the number of independent variables. Given the limited dataset of approximately 1,000 observations, incorporating a large number of independent variables poses

challenges to the model performance, potentially resulting in issues such as overfitting, multicollinearity, and decreased stability. Therefore, a more comprehensive approach to feature selection is required, which considers both the quality of the data and the relevance of the variables to maize yield estimation. By doing so, we can improve the accuracy of yield predictions.

Through careful selection of variables, including VI, climate, and soil variables, our model was able to explain over half of the variation in observed maize yields. Importantly, we found that using VI calculated from individual bands was more effective than using individual bands themselves. Both NDVI and GCI performed comparably in yield estimation, likely due to their shared use of near-infrared band values in their calculations. The fact that our models using NDVI and GCI had similar performance suggests that either of these indices can be used effectively for maize yield estimation. Our findings highlight the potential benefits of using satellite-derived VI such as NDVI and GCI and the accuracy of yield prediction can be improved by leveraging the sensitivity of near-infrared values to yield estimates.

The presence of clouds in satellite images can significantly affect the accuracy of crop monitoring using NDVI and GCI. Clouds can cause missing data and spectral contamination in satellite images, which can lead to unreliable estimates of crop yield. Therefore, it is essential to develop methods that can effectively handle the cloud effects in satellite images and improve the accuracy of crop yield estimation (Guerini Filho et al., 2020). The SG function is commonly used to smooth time series data by fitting low-degree polynomials to successive subsets of adjacent data points through least squares. However, our analysis revealed limitations with the SG function provided in the GEE platform, particularly when dealing with long periods of cloud coverage. Despite applying the SG function, the NDVI profiles still showed sudden drops, indicating unsatisfactory results. We observed this problem in one example where the cloud coverage persisted throughout August in Fig. 6. To address this issue, we suggest using a third-party SG function that provides more parameterization options, such as deriving weighted values for each NDVI value based on the Quality Assurance (QA) layer from the original satellite image bands. The QA values indicate the quality of the observation from the satellite and label the pixel with high quality when there is no contamination. By applying QA weights, the original NDVI values with low QA will be treated as missing values in the curve fitting process, potentially improving the accuracy of the smoothing algorithm.

It is also worth noting that the choice of machine learning algorithm depends on the specific problem being solved and the data available (Hastie et al., 2009). Neural networks are known to capture intricate patterns and relations in the data. While neural networks have shown slightly better performance in the yield estimates, it is essential to carefully choose the architecture and hyperparameters of the model to achieve optimal performance when working with neural networks. This requires extensive experimentation and tuning, which can be time-consuming and challenging.

Using a model developed from data from a prior year to predict future yield outcomes is a common approach in agriculture and other fields (Battude et al., 2016). It can be an effective way to estimate potential crop yields for the upcoming season and help farmers make informed decisions about crop management. The model from neural networks provides reasonable results when estimating the yield of different years but the R<sup>2</sup> is lower than the estimates of same year. It indicates that the ability of the model to predict future yields accurately may depend on various factors such as weather conditions, soil quality, and farming practices, which may vary from year to year. Therefore, it is essential to regularly evaluate the model performance and retrain it as needed to ensure its accuracy and reliability in predicting future yield outcomes. Overall, the use of machine learning models in agriculture to predict future yield outcomes is a promising approach.

Lastly, the results showed that NDVI in late September had the highest importance score throughout the year. In terms of T<sub>min</sub> and precipitation, the data for September showed the most relative importance out of the 12 months. On the other hand, T<sub>max</sub> showed the greatest

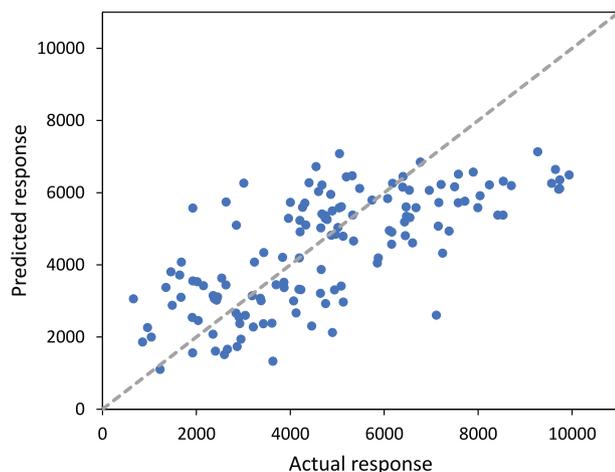


Fig. 5. Scatter plot of predicted and observed maize yields ( $\text{kg ha}^{-1}$ ) in 2018 using model calibrated with 2017 data.

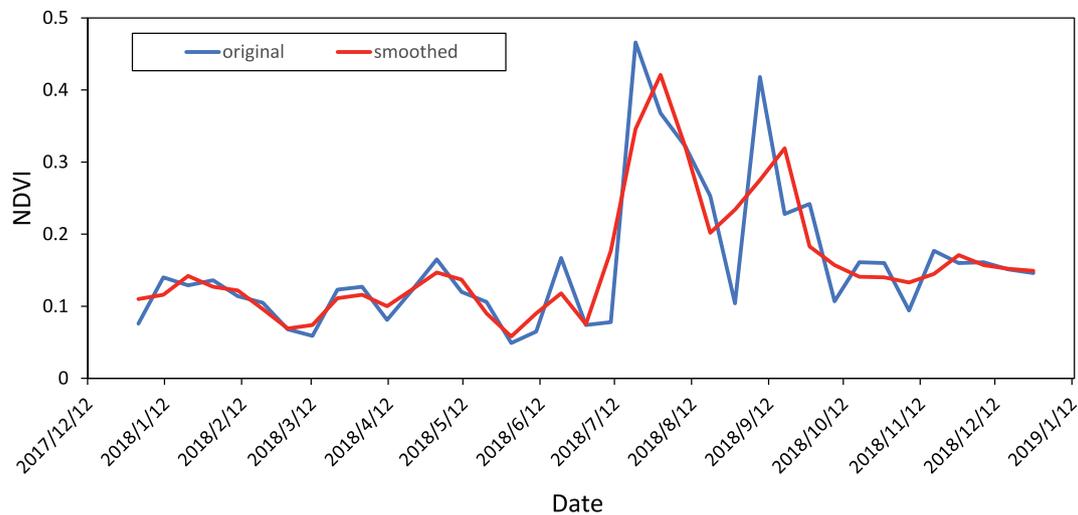


Fig. 6. The comparison of the time series of original NDVI and smoothed NDVI.

importance in November. The significance of these findings is that different periods play varying roles in predicting maize yield. The importance of T<sub>min</sub> and precipitation in September, for example, suggests that these factors are crucial for the beginning of the growing season, while the importance of T<sub>max</sub> in November suggests that it is more relevant to the end of the growing season. These findings could be useful for farmers and policymakers in Ethiopia, as they can provide insights into the timing of key environmental factors that influence crop growth and yield.

## 6. Conclusions

In this study we described a machine learning based approach to estimate smallholder maize yield using geospatial data from various sources, using data from Ethiopia as a case study. NDVI, together with climate data and soil indicators, produced the maize yield estimates with the highest accuracy. We found the neural networks model to perform slightly better than other machine learning algorithms. Importantly, our work indicates that a model developed and calibrated on a previous year's data could be used to reasonably estimate maize yield in the subsequent year, although more work is needed to further explore the robustness of such out-of-sample predictions in other settings. Our analysis of feature importance indicated that NDVI was the most important predictor when estimating maize yield in the field. Furthermore, the relative importance of NDVI, T<sub>min</sub>, precipitation, and T<sub>max</sub> varied. Our study suggests the feasibility of developing national programs for the routine generation of broad-scale, high-resolution estimates of smallholder maize yield, including seasonal forecasts, on the basis of machine learning algorithms, well-measured ground control data, and currently available time series satellite data.

## Abbreviations

Not applicable.

## Availability of data and materials

Data will be shared upon request by the readers.

## Authors' contributions

Conceptualization, Z.G., J.C., and L.Y.; Writing original draft, Z.G.; Review and editing, Z.G., J.C., and L.Y.; Data curation, Z.G. and J.C.; Investigation, Z.G., J.C., and L.Y.; Funding acquisition, Z.G. and J.C.; All authors have approved the final version of the manuscript.

## Declaration of competing interest

No conflicts of interest exist in the submission of this manuscript. Author Jordan Chamberlin and Liangzhi You (Editorial Board members) were not involved in the journal's review nor decisions related to this manuscript.

## Acknowledgements

We gratefully acknowledge support for this study from the Bill & Melinda Gates Foundation, through the Taking Maize Agronomy to Scale in Africa (TAMASA) project (investment number INV-008260), and from the CGIAR Research Program MAIZE, led by the International Maize and Wheat Improvement Center (CIMMYT). This work also received financial support from the German Federal Ministry for Economic Cooperation and Development (BMZ) commissioned and administered through the Deutsche Gesellschaft für INTERNATIONALE ZUSAMMENARBEIT (GIZ) Fund for International Agricultural Research (FIA) (grant number 81260864).

## References

- Abate, T., Shiferaw, B., Menkir, A., Wegary, D., Kebede, Y., Tesfaye, K., Kassie, M., Bogale, G., Tadesse, B., Keno, T., 2015. Factors that transformed maize productivity in Ethiopia. *Food Secur.* 7, 965–981. <https://doi.org/10.1007/s12571-015-0488-z>.
- Battude, M., Al Bitar, A., Morin, D., Cros, J., Huc, M., Marais Sicre, C., Le Dantec, V., Demarez, V., 2016. Estimating maize biomass and yield over large areas using high spatial and temporal resolution Sentinel-2 like remote sensing data. *Remote Sens. Environ.* 184, 668–681. <https://doi.org/10.1016/j.rse.2016.07.030>.
- Becker-Reshef, I., Barker, B., Humber, M., Puricelli, E., Sanchez, A., Sahajpal, R., McGaughey, K., Justice, C., Baruth, B., Wu, B., Prakash, A., Abdolreza, A., Jarvis, I., 2019. The GEOGLAM crop monitor for AMIS: Assessing crop conditions in the context of global markets. *Glob. Food Secur.* 23, 173–181. <https://doi.org/10.1016/j.gfs.2019.04.010>.
- Becker-Reshef, I., Franch, B., Barker, B., Murphy, E., Santamaria-Artigas, A., Humber, M., Skakun, S., Vermote, E., 2018. Prior season crop type masks for winter wheat yield forecasting: a US case study. *Remote Sens.* 10, 1659. <https://doi.org/10.3390/rs10101659>.
- Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A., 1984. *Classification and Regression Trees*. Chapman & Hall, New York, U.S.A.
- Burke, M., Lobell, D.B., 2017. Satellite-based assessment of yield variation and its determinants in smallholder African systems. *Proc. Natl. Acad. Sci. U.S.A.* 114, 2189–2194. <https://doi.org/10.1073/pnas.1616919114>.
- Cao, J., Zhang, Z., Tao, F., Zhang, L., Luo, Y., Zhang, J., Han, J., Xie, J., 2021. Integrating multi-source data for rice yield prediction across China using machine learning and deep learning approaches. *Agric. For. Meteorol.* 297, 108275. <https://doi.org/10.1016/j.agrformet.2020.108275>.
- Carletto, C., Jolliffe, D., Banerjee, R., 2015. From tragedy to renaissance: improving agricultural data for better policies. *J. Dev. Stud.* 51, 133–148. <https://doi.org/10.1080/00220388.2014.968140>.
- Davis, B., Di Giuseppe, S., Zezza, A., 2017. Are African households (not) leaving agriculture? Patterns of households' income sources in rural Sub-Saharan Africa. *Food Policy* 67, 153–174. <https://doi.org/10.1016/j.foodpol.2016.09.018>.

- Delincé, J., 2017. Recent Practices and Advances for AMIS Crop Yield Forecasting at Farm and Parcel Level: A Review. Food and Agriculture Organization of the United Nations Publication, Rome, Italy. <http://www.fao.org/3/a-i7339e.pdf>.
- 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 Sens.* 6, 9653–9675. <https://doi.org/10.3390/rs6109653>.
- Eze, E., Girma, A., Zenebe, A.A., Zenebe, G., 2020. Feasible crop insurance indexes for drought risk management in Northern Ethiopia. *Int. J. Disaster Risk Reduct.* 47, 101544. <https://doi.org/10.1016/j.ijdrr.2020.101544>.
- Fischer, G., Nachtergaele, F., Prieler, S., Van Velthuizen, H.T., Verelst, L., Wiberg, D., 2008. *Global Agro-Ecological Zones Assessment for Agriculture (GAEZ 2008)*. IIASA, Laxenburg, Austria and FAO, Rome, Italy.
- 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. *Int. J. Appl. Earth Obs. Geoinf.* 76, 112–127. <https://doi.org/10.1016/j.jag.2018.11.012>.
- Fritz, S., See, L., Bayas, J.C.L., Waldner, F., Jacques, D., Becker-Reshef, I., Whitcraft, A., Baruth, B., Bonifacio, R., Crutchfield, J., Rembold, F., Rojas, O., Schucknecht, A., Van der Velde, M., Verdin, J., Wu, B., Yan, N., You, L., Gilliams, S., Múchern, S., Tetrault, R., Moorthya, I., McCallum, I., 2019. A comparison of global agricultural monitoring systems and current gaps. *Agric. Syst.* 168, 258–272. <https://doi.org/10.1016/j.agry.2018.05.010>.
- Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., Michaelsen, J., 2015. The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Sci. Data* 2, 150066. <https://doi.org/10.1038/sdata.2015.66>.
- Gao, F., Anderson, M.C., Zhang, X., Yang, Z., Alfieri, J.G., Kustas, W.P., Mueller, R., Johnson, D.M., Prueger, J.H., 2017. Toward mapping crop progress at field scales through fusion of Landsat and MODIS imagery. *Remote Sens. Environ.* 188, 9–25. <https://doi.org/10.1016/j.rse.2016.11.004>.
- Guerini Filho, M., Kuplich, T.M., De Quadros, F.L.F., 2020. Estimating natural grassland biomass by vegetation indices using Sentinel 2 remote sensing data. *Int. J. Remote Sens.* 41, 2861–2876. <https://doi.org/10.1080/01431161.2019.1697004>.
- Hastie, T., Tibshirani, R., Friedman, J., 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition*. Springer, New York, U.S.A.
- Hunt, M.L., Blackburn, G.A., Carrasco, L., Redhead, J.W., Rowland, C.S., 2019. High resolution wheat yield mapping using Sentinel-2. *Remote Sens. Environ.* 233, 111410. <https://doi.org/10.1016/j.rse.2019.111410>.
- Jain, M., Srivastava, A.K., Balwinder-Singh, Joon, R.K., McDonald, A., Royal, K., Lisaius, M.C., Lobell, D.B., 2016. Mapping smallholder wheat yields and sowing dates using micro-satellite data. *Remote Sens.* 8, 860. <https://doi.org/10.3390/rs8100860>.
- Jin, Z., Azzari, G., Burke, M., Aston, S., Lobell, D.B., 2017. Mapping smallholder yield heterogeneity at multiple scales in Eastern Africa. *Remote Sens.* 9, 931. <https://doi.org/10.3390/rs9090931>.
- Jin, Z., Azzari, G., You, C., Di Tommaso, S., Aston, S., Burke, M., Lobell, D.B., 2019. Smallholder maize area and yield mapping at national scales with Google Earth Engine. *Remote Sens. Environ.* 228, 115–128. <https://doi.org/10.1016/j.rse.2019.04.016>.
- Lambert, M.-J., Traoré, P.C.S., Blaes, X., Baret, P., Defourny, P., 2018. Estimating smallholder crops production at village level from Sentinel-2 time series in Mali's cotton belt. *Remote Sens. Environ.* 216, 647–657. <https://doi.org/10.1016/j.rse.2018.06.036>.
- Li, N., Li, X.Y., Zou, Z.X., Lin, L.R., Li, Y.Q., 2011. A novel baseline-correction method for standard addition based derivative spectra and its application to quantitative analysis of benzo(a)pyrene in vegetable oil samples. *Analyst* 136, 2802–2810. <https://doi.org/10.1039/C0AN00751J>.
- Lobell, D.B., Azzari, G., Burke, M., Gourlay, S., Jin, Z., Kilic, T., Murray, S., 2020. Eyes in the sky, boots on the ground: assessing satellite- and ground-based approaches to crop yield measurement and analysis. *Am. J. Agric. Econ.* 102, 202–219. <https://doi.org/10.1093/ajae/aaz051>.
- Lobell, D.B., Burke, M.B., 2010. On the use of statistical models to predict crop yield responses to climate change. *Agric. For. Meteorol.* 150, 1443–1452. <https://doi.org/10.1016/j.agrformet.2010.07.008>.
- Lobell, D.B., Thau, D., Seifert, C., Engle, E., Little, B., 2015. A scalable satellite-based crop yield mapper. *Remote Sens. Environ.* 164, 324–333. <https://doi.org/10.1016/j.rse.2015.04.021>.
- Nguyen, L.H., Joshi, D.R., Clay, D.E., Henebry, G.M., 2020. Characterizing land cover/land use from multiple years of Landsat and MODIS time series: A novel approach using land surface phenology modeling and random forest classifier. *Remote Sens. Environ.* 238, 111017. <https://doi.org/10.1016/j.rse.2018.12.016>.
- O'Neill, P.E., Chan, S., Njoku, E.G., Jackson, T., Bindlish, R., 2016. SMAP L3 Radiometer Global Daily 36 km EASE-Grid Soil Moisture (Version 4). <https://doi.org/10.5067/ZX7YX2Y2LHEB>.
- Pradhan, S., 2001. Crop area estimation using GIS, remote sensing and area frame sampling. *Int. J. Appl. Earth Obs. Geoinf.* 3, 86–92. [https://doi.org/10.1016/S0303-2434\(01\)85025-X](https://doi.org/10.1016/S0303-2434(01)85025-X).
- Rasmussen, C.E., 2004. Gaussian processes in machine learning. In: Bousquet, O., von Luxburg, U., Rätsch, G. (Eds.), *Advanced Lectures on Machine Learning*. Springer-Verlag Berlin Heidelberg, Germany, pp. 63–71. [https://doi.org/10.1007/978-3-540-28650-9\\_4](https://doi.org/10.1007/978-3-540-28650-9_4).
- Sagan, V., Maimaitjiang, M., Bhadra, S., Maimaitiyiming, M., Brown, D.R., Sidike, P., Fritschi, F.B., 2021. Field-scale crop yield prediction using multi-temporal WorldView-3 and PlanetScope satellite data and deep learning. *ISPRS-J. Photogramm. Remote Sens.* 174, 265–281. <https://doi.org/10.1016/j.isprsjprs.2021.02.008>.
- Tao, F.L., Zhang, Z., Liu, J.Y., Yokozawa, M., 2009. Modelling the impacts of weather and climate variability on crop productivity over a large area: A new process-based model development, optimization, and uncertainties analysis. *Agric. For. Meteorol.* 149, 1266–1278. <https://doi.org/10.1016/j.agrformet.2008.11.004>.
- Tuvdendorj, B., Wu, B., Zeng, H., Batdelger, G., Nanzad, L., 2019. Determination of appropriate remote sensing indices for spring wheat yield estimation in Mongolia. *Remote Sens.* 11, 2568. <https://doi.org/10.3390/rs11212568>.
- Wang, B., Gu, X., Ma, L., Yan, S., 2016. Temperature error correction based on BP neural network in meteorological wireless sensor network. In: Sun, X., Liu, A., Chao, H.C., Bertino, E. (Eds.), *Proceedings of the International Conference on Cloud Computing and Security*. Springer-Cham, Geneva, Switzerland, pp. 117–132.
- Zhang, L., Zhang, Z., Luo, Y., Cao, J., Xie, R., Li, S., 2021. Integrating satellite-derived climatic and vegetation indices to predict smallholder maize yield using deep learning. *Agric. For. Meteorol.* 311, 108666. <https://doi.org/10.1016/j.agrformet.2021.108666>.
- Zinyengere, N., Mhizha, T., Mashonjowa, E., Chipindu, B., Geerts, S., Raes, D., 2011. Using seasonal climate forecasts to improve maize production decision support in Zimbabwe. *Agric. For. Meteorol.* 151, 1792–1799. <https://doi.org/10.1016/j.agrformet.2011.07.015>.