Debalke and Abebe
*Environmental Systems Research* (2022) 11:1
https://doi.org/10.1186/s40068-022-00249-5

Environmental Systems Research

RESEARCH Open Access

# Maize yield forecast using GIS and remote sensing in Kaffa Zone, South West Ethiopia

Dereje Biru Debalke[1*] and Jemal Tefera Abebe[2]

**Abstract**

**Background:** Ethiopian policy makers, government planners, and farmers all demand up-to-date information on maize yield and production. The Kaffa Zone is the country's most important maize-producing region. The Central Statistical Agency's manual gathering of field data and data processing for crop predictions takes a long time to complete before official conclusions are issued. In various investigations, satellite remote sensing data has been shown to be an accurate predictor of maize yield. With station data from 2008 to 2017, the goal of this study was to develop a maize yield forecast model in the Kaffa Zone using time series data from the Moderate Resolution Imaging Spectroradiometer Normalized Difference Vegetation Index, actual evapotranspiration, potential evapotranspiration, and Climate Hazards Group Infrared Precipitation. The indicators' correctness in describing the production was checked using official grain yield data from Ethiopia's Central Statistical Office. Crop masking was applied on cropland, and agro ecological zones suited for the crop of interest were used to change the crop. Throughout the long wet season, correlation studies were utilized to investigate correlations between crop productivity, spectral indices, and agro climatic factors for the maize harvest. There were indicators established that demonstrated a strong relationship between maize yield and other factors.

**Results:** The Normalized Difference Vegetation Index Average and Climatic Hazards Group Infrared Precipitation with station data rainfall exhibit substantial associations with maize productivity, with correlations of 84 percent and 89 percent, respectively. To put it another way, these variables have a significant beneficial impact on maize yield. The derived spectro-agro meteorological yield model ($r^2 = 0.89$, RMSE = 1.54qha$^{-1}$, and 16.7% coefficient of variation) matched the Central Statistical Agency's expected Zone level yields satisfactorily.

**Conclusion:** As a result, remote sensing and geographic information system-based maize yield forecasts improved data quality and timeliness while also distinguishing yield production levels/areas and simplifying decision-making for decision-makers, demonstrating the clear potential of spectro-agro meteorological factors for maize yield forecasting, particularly in Ethiopia.

**Keywords:** CHIRPS, eMODIS NDVI, Maize yield, Remote sensing

## Background

Crop yield forecasting is critical for policy planning and decision-making. For crop monitoring and production forecasts, many countries rely on traditional data collection methods such as ground-based visits and reports. Due to insufficient ground observation, these reporting procedures are subjective, costly, time-consuming, and prone to major errors, resulting in inaccurate crop production evaluations and a delay in reporting critical measures (Greatrex 2012). Before the emergence of remote-sensing techniques like the Normalized Difference Vegetation Index (NDVI), crop-weather models were used for crop monitoring and yield forecasts (Rojas 2007). In the Kaffa Zone, crop data was collected on the ground, which is a time-consuming, expensive, and

*Correspondence: derejebiru22@gmail.com
[1] Department of Geography and Environmental Studies, College of Social Sciences, Bonga University, Bonga, Ethiopia
Full list of author information is available at the end of the article

labor-intensive task. In terms of resolving these concerns, re mote sensing is more important than ground surveys. Because remote sensing can give precise and timely data for crop production estimation, most studies have identified a link between the Normalized Difference Vegetation Index (NDVI), agro meteorological data, green biomass, and yield (Rojas 2007).

Many research on agricultural production forecasting at various zonal levels have been undertaken in Ethiopia utilizing these methodologies; Zinna and Suryabhagavan (2016) used time series data from SPOT VEGETATION, actual and potential evapotranspiration, and rainfall estimate satellite data from 2003 to 2012 to conduct a maize crop forecast study in the south Tigray Zone. Reda (2015) used time series data from SPOTVEGETATION, actual and potential evapotranspiration, rainfall estimate, and satellite data from 2004 to 2013 to predict wheat crop yield in the Arsi zone using remote sensing and GIS approaches. However, both investigations employed SPOT VEGETATION NDVI and RFE 2.0, which cover vast areas with low-resolution (1 km) and (10 km), respectively, rather than Moderate Resolution Imaging Spectroradiometer Normalized Difference Vegetation Index (eMODIS NDVI), which is a better data set for crop monitoring due to the length of the time series (since 2000) and spatial resolution (250 m), as well as the fact that it is freely available and easy to access. For Climatic Hazards Group Infrared Precipitation (CHIRPS) rainfall, data from 1981 dekedal is accessible, and products with a spatial resolution of 0.05° can be obtained in near-real time. As a result, the researchers wanted to solve this research gap by developing a model that uses Moderate Resolution Imaging Spectroradiometer Normalized Difference Vegetation Index (eMODIS NDVI) and Climatic Hazards Group Infrared Precipitation (CHIRPS) satellite rainfall to forecast maize yield for the year 2018 in the Kaffa Zone utilizing Remote Sensing and GIS approaches.

## Materials and methods

### Description of the Study Area

This research was carried out in the Kaffa Zone, which is located in the South, Nation, Nationalities and Peoples Region, between 6°24' and 8°13' north latitude and 35°30' to 36°46' east longitude. The Zone covers a total area of 10,602.7 km$^2$, accounting for 7.06 percent of the region's total area. Based on altitude and temperature variances, the Kaffa Zone is divided into twelve administrative districts and categorized into three traditional climate zones. Highland (2500–3000 m), midland (1500–2500 m), and lowland (1500–2500 m) are the three types (500–1500 m). Highland, midland, and lowland areas make up 11.6 percent, 59.5 percent, and 28.9% of the Zone's total area, respectively. According to the National Meteorology Agency (NMA), the average annual temperature in the area is between 10.1 and 27.5 degrees Celsius February, March, and April are the hottest months, while July and August are the coolest. The annual rainfall varies between 1001 and 2200 mm. Ethiopia's Kaffa Zone is located in the country's southwest, where it receives the most rainfall. This is due to the existence of an evergreen forest cover on top of the wet monsoon winds' windward site (Fig. 1).

### Data and data sources

#### *Expedited MODIS (eMODIS) -TERRA NDVI*

For agricultural production assessments and crop yield estimation, many studies employing data from intermediate spatial resolution satellite sensors such as the Moderate Resolution Imaging Spectroradiometer (MODIS) are recommended (Becker-Reshef et al. 2010; Mkhabela et al. 2011; Vintrou et al. 2012; Kouadio et al. 2014; Johnson 2014 and Faisal et al. 2019). MODIS data is freely available and has a high temporal resolution but a low spatial resolution, which could explain some of the interest (Kouadio et al. 2014). The Normalized Difference Vegetation Index (NDVI), which indicates the contrast between the highest absorption in the red section of the spectrum and the highest reflection in the near-infrared portion, has long been used in agriculture for crop monitoring and other uses (Hatfield and Prueger 2010; Basso et al. 2013). When the MODIS NDVI was compared to the NOAA-AVHRR (National Oceanographic and Atmospheric Administration-Advanced Very High-Resolution Radiometer) NDVI temporal profiles for a number of biome types, the MODIS-based index outperformed the NOAA-AVHRR in terms of defining seasonal phenology (Kouadio et al. 2014). MODIS VIs is useful for crop monitoring in agricultural settings that are fragmented (sphere size nearing pixel scale) (Duveiller et al. 2012). As a result, the planting season in the research area began in mid-June, as seen by the zone livelihood profile. The maize crop will be sown in the study region in June, according to the document. According to local farmers, maize crops in the kaffa zone are planted in June, biomass growth occurs from July to August, and blossoming occurs in September.

As a result, images of the Moderate Resolution Imaging Spectroradiometer Normalized Difference Vegetation Index (eMODIS NDVI) decadal were obtained from https://earlywarning.usgs.gov/fews/datadownloads/East%20Africa/eMODIS%20NDVI20C6 from June to September, beginning in 2008 and ending in 2017. (Statistics from a ten-year period). The NDVI was calculated analytically as follows (Eq. 1):

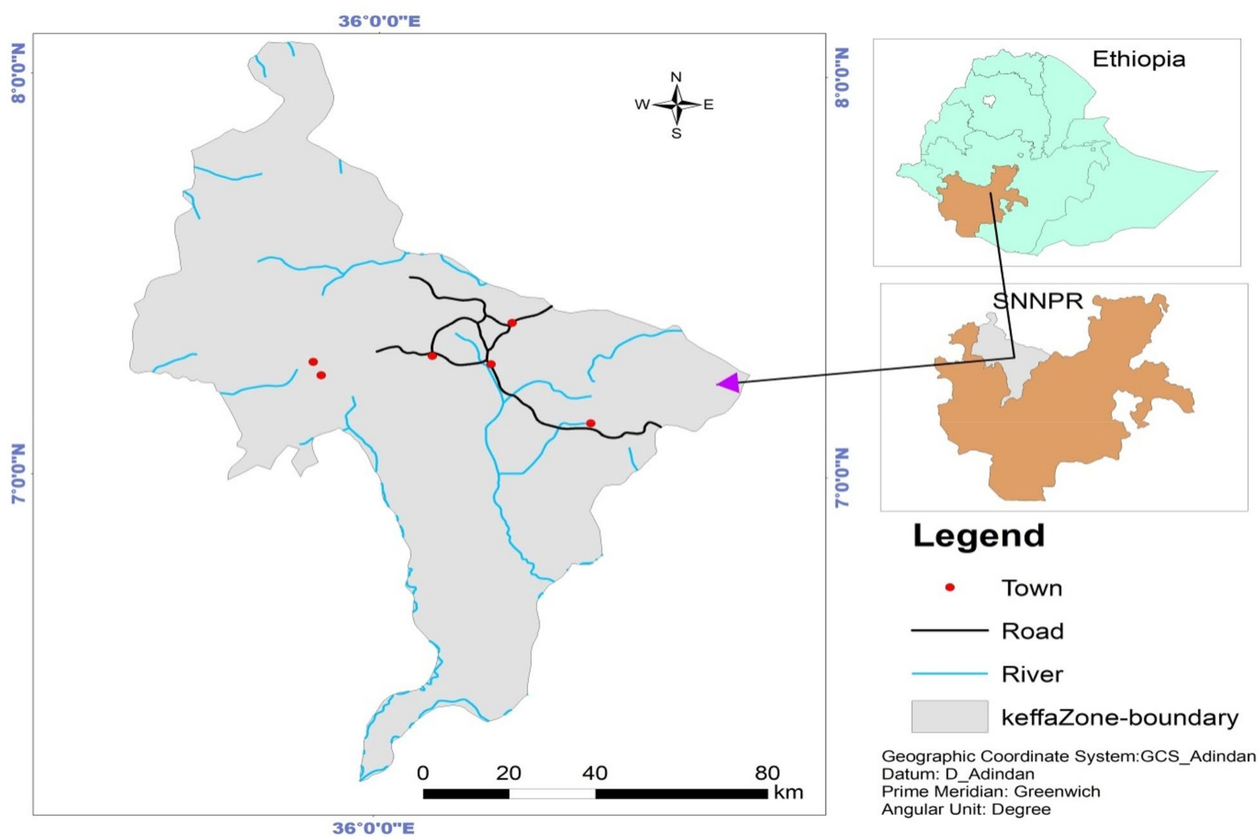


**Fig. 1.** Location map of the Kaffa Zone

$$\mathrm{NDVI} = (NIR - RED)/(\mathrm{NIR} + RED) \tag{1}$$

where NIR = near-infrared reflectance and RED = visiblered reflectance.

The row eMODIS data were processed, rescaled, and analyzed in the ArcGIS 10.5 program to produce the real NDVI value of the study area (Eq. 2):

$$\text{eMODIS NDVI} = \text{Float}(\text{Smoothed eMODIS NDVI} - 100)/100 \tag{2}$$

The Climate Hazards Center Infrared Precipitation with Station Data (CHIRPS) data set is quasi-global in scope and spans 30 years. Climate Hazards Center Infrared Precipitation with Station Data (CHIRPS) creates gridded rainfall time series for trend analysis and seasonal drought monitoring by combining 0.05° resolution satellite images with in-situ station data, spanning 50°S-50°N (and all longitudes) from 1981 to near-present. From June to September, 2008 to 2017 (ten-year statistics), which were freely downloaded from https://data.chc.ucsb.edu/products/CHIRPS-2.0/africadekad/tifs/.

Actual Evapotranspiration (ETa) is calculated using data from the Aqua satellite and the Operational Simplified Surface Energy Balance (SSEBop) model (Senay et al. 2013). The SSEBop configuration is based on (Senay et al. 2013) original Simplified Surface Energy Balance (SSEB) approach, but with updated and improved parameterizations for practical usage. It combines ET fractions derived from remotely sensed MODIS thermal imaging, which are summed every ten days (dekadal) at a resolution of one kilometer. The data was used to examine vegetation and landscape conditions in order to detect early warning droughts. Which were freely downloaded from https://earlywarning.usgs.gov/fews/datadownloads/Continental%20Africa/Monthly%20ET%20. Anomaly from June to September, 2008 to 2017 (ten years' time series data).

Another input for the model computation was Potential Evapotranspiration (PET), which was estimated using the modified Hargreaves equation, and the maize crop coefficient from the livelihood early assessment protection (LEAP) software was used to correct for the crop's growth stage. The climate variables used to create PET for this study were gathered from Ethiopia's national meteorological office from June to September, 2008 to 2017 (10 years' time series data).

#### *Water requirement satisfaction index (WRSI)*

The USGS/FEWSNET recently used a Geospatial WRSI crop model, which enables for localized crop modeling, monitoring, and forecasting at the subnational level, using locally accessible statistics as model inputs. The result of this model was also chosen as one of the parameters for developing a maize forecast model. The water requirement satisfaction index for a season is determined by the amount of water a crop receives and uses during the growing season. The water need satisfaction index was calculated using the ratio of seasonal actual evapotranspiration (ETa) to seasonal crop water requirement (WR) (Eq. 3):

$$\mathrm{WRSI} \ = (\mathrm{ETa}/\mathrm{WR}) * 100 \tag{3}$$

To account for the crop's growth stage, water requirements were calculated using the modified Hargreaves equation potential evapotranspiration (PET) and the crop coefficient (Kc) using livelihood early assessment protection (LEAP) software (Eq. 4):

$$\mathrm{WR} = \mathrm{PET} + Kc \tag{4}$$

#### *Spot6 and landsat 8 images*

The Ethiopian Geospatial Information Agency (EGIA) provided spot and Landsat images of the study area for supervised land use and land cover classification. Prior to categorization, the image's spatial resolution was increased or pans harped to 1.5 m spatial resolution for the spectral bands. A sensor fusion of a multispectral Landsat image with a panchromatic SPOT image provided the best of both image types (Lillesand et al. 2015).

#### *Crop masks data for maize*

Another input for masking maize data is crop agroecology in the research area. Maize is generally grown between the elevations of l500 and 2200 m (Eq. 5) according to Gorfu and Ahmed (2012):

$$\text{Maize elevation} = *\text{Value}* \geq 1500 AND * \text{Value}* \leq 2200. \tag{5}$$

#### *Ancillary dataset*

Ancillary dataset: The appropriate data sets, such as shape files, were received from the Central Statistics Agency of Ethiopia (CSA) for the 2007 population and housing census mapping. These shape files were used to define the study area's boundary. Topo-sheets from Ethiopian geospatial information agency were also wont to check the geometric correction of the satellite imageries.

#### *Official yield statistics*

The calibration of the model with historical crop yield records is required for the creation of quantitative yield estimates (Rijks et al. 2007). As a result, Central Statistics Agency of Ethiopia (CSA) was requested for historical grain yield data (2008–2017) at the Zonal level. The maize grain yield estimate archive was provided by Central Statistical Agency's agriculture section (Table 1). The yield statistics were derived using a list frame approach supported by a ground sample survey (Tables 2 and 3).

### Data processing and analysis

#### *Classification*

The research area's pan sharpened SPOT 6 image is processed for supervised classification in ArcGIS software. According to Yan et al. (2006), supervised categorization necessitates the user identifying the various pixel values or spectral signatures that should be linked with each class. This is done by identifying training sites or locations that are typical sample sites of well-known cover types. In order to construct a thematic map of land cover and identify the Land use land cover classification of the study area, the maximum likelihood classifier (MLC) was used to categorize land cover into two classes (agricultural and non-agriculture) (Fig. 2). It is vital to assess the precision of a map created with remote sensing data. The most popular way for presenting the accuracy of categorization findings is to use an error matrix. Overall accuracy, user and producer accuracies, and the Kappa statistic were all calculated using the error matrices. After reducing the fraction of agreement that may occur by chance, the Kappa statistic integrates the off diagonal portions of the error matrices and indicates agreement.

**Table 1** Trend in maize crop yield in the Kaffa Zone from 2008 to 2017. Source of data: Annual agricultural report from the CSA (2018)

| Year | Holder | Hectares(HA) | Production(QA) | QT/HA |
|---|---|---|---|---|
| 2008 | 33,915 | 9711.58 | 169,975.06 | 17.5 |
| 2009 | 51,716 | 20,223.18 | 408,441.42 | 20.2 |
| 2010 | 60,616 | 20,225.04 | 440,056.4.92 | 21.76 |
| 2011 | 59,194 | 24,008.53 | 606,474.42 | 25.26 |
| 2012 | 71,171 | 27,552.16 | 601,615.51 | 21.84 |
| 2013 | 87,562 | 36,080.52 | 925,916.99 | 25.66 |
| 2014 | 44,075 | 20,691.03 | 589,424.15 | 28.49 |
| 2015 | 64,063 | 25,181.68 | 753,784.78 | 29.93 |
| 2016 | 80,203 | 29,370.34 | 866,635.91 | 29.51 |
| 2017 | 90,228 | 31,812.48 | 932,105.66 | 29.3 |

**Table 2** Data used in the study, along with its description, source, and purpose

| Types of data | Description | Source | Purpose |
|---|---|---|---|
| DEM | Aster 30 m resolution | United States Geological Survey(USGS) | For maize elevation generates |
| Pan sharpened SPOT 6 image | Spatial resolution of 1.5 m,Panchromatic and 6 m Multispectral,<br>12 bit Radiometric resolution,<br>1 days Temporal resolution &Path/raw—170/55 | Ethiopian Geospatial Information Agency (EGIA) | In order to perform supervised LULC classification, |
| CHERIPS rainfall | Spatial resolution:(0.050*0.050), Frequency: dekadal, Archive:2008–2017&<br>Format:netCDF | Downloading from https://data.chc.ucsb.edu/products/CHIRPS-2.0/africa_dekad/tifs/ | Used to calculate yearly average rainfall to correlate it with maize yield |
| potential evaporation(pet) | 1 km*1 km Spatial resolution, Frequency: dekadal, Archive: 2008–2017& Format:netCDF | National Meteorology Agency of Ethiopia(NMA) | Used to calculate WRSI to correlate it with maize yield |
| eMODIS NDVI | 250 m*250 m special resolution,<br>12bits radiometric resolution, 1–2 days Temporal resolution, Frequency: dekadal& Archive:2008–2017 | Downloading from https://earlywarning.usgs.gov/fews/datadownloads/East%20Africa/eMODIS%20NDVI%20C6 | Used To calculate yearly average NDVI to correlate it with maize yield |
| Eta (actual evaporation) | 1 km*1 km special resolution,8 bit spectral resolution, 16 days Temporal resolution, Frequency: dekadal & Archive: 2008–2017 | downloaded freely from https://earlywarning.usgs.gov/fews/datadownloads/Continental%20Africa/Monthly%20ET%20Anomaly | Used to calculate yearly average Eta,Eta total and WRSI to correlate these variables with maize yield |
| Maize yield(qt/ht) | Archive data from 2008 to 2017 | Central Statistical Agency annual agricultural report, 2018 | To calibrate the developed model with historical crop yield statistics |
| Ground Truth and Accuracy Assessment Points | Using HH GPS Garmin62, generate random coordinates from each land use | Bonga University | For accuracy assessment of the supervised classification |
| Study area boundary | Shape file | Central Statistics Agency of Ethiopia (CSA), 2007 | Used to demarcate the study area boundary |

**Table 3** Summary of the data collecting and analysis equipment and materials used

| Software used | Purpose |
|---|---|
| GPS(Global Position System) | For the purpose of gathering ground control points (GCPs), which will be used to assess accuracy |
| Erdas2015, ArcMap10.3, LEAP 2.7.1, SPSS statistical tool | GIS and statistical software for image and vector processing and data analysis |
| Google Earth | Used as supplementary for checking and correcting area of doubt about accuracy of the classification |
| CDT (Climate Data Tool) | To calculate potential evapotranspiration |

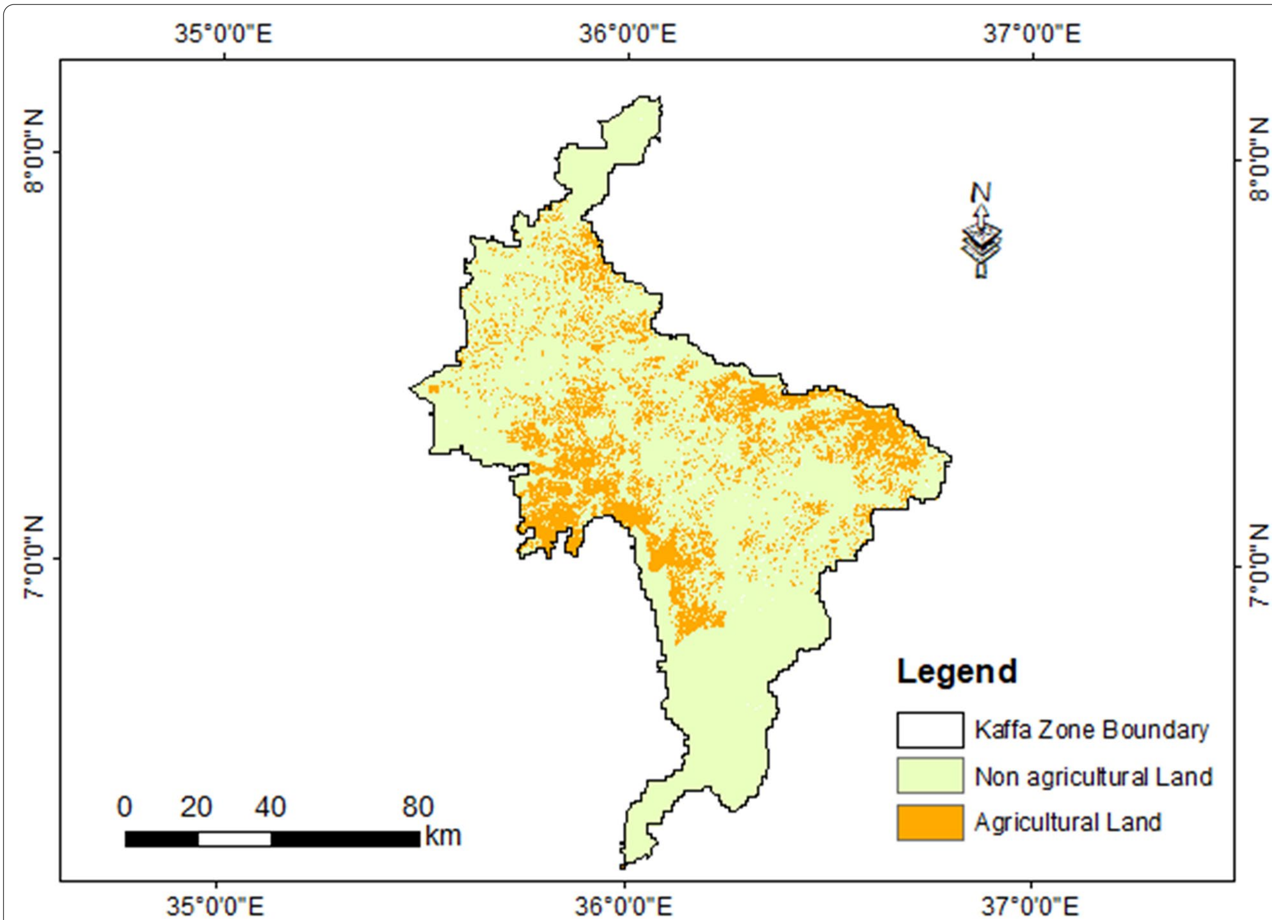


**Fig. 2** Maps of Land use/land cove of the study area

As a result, both agricultural and non-agricultural classes were evenly represented. A significant number of samples that represent the thematic classes and are scattered uniformly across the map are required to test attribute accuracy. As a general rule, Congalton and Green (2019), recommend at least 50 samples each class. At least 75–100 samples per class should be taken if the area is higher than 500 km$^2$ or the number of categories is greater than 12. As a result, the accuracy assessment sample size was set at 200, with 100 sample points for each class. These points were verified in two ways: those that were visible and reachable in the field, and those that could be verified using Google Earth as a reference. As a result, for the 200 sample points, the following error matrix (Table 4) is displayed. The overall accuracy of the data was 90%, with a kappa coefficient of 0.80, and the interpretation may be accepted for further study based on the result.

**Table 4** Accuracy assessment

| Map data | Ground truth data | | Total | User accuracy |
|---|---|---|---|---|
| | Agricultural | Non agricultural | | |
| Agricultural | 88 | 8 | 96 | 91.7 |
| Non agricultural | 12 | 92 | 104 | 88.5 |
| Total | 100 | 100<br>92 | 200 | |
| Producer accuracy | 88 | | | |

### *Mask data derivation*

Agricultural agro-ecology is another input for masking crop data in the research area. Maize is generally grown between the elevations of l500 and 2200 according to Gorfu and Ahmed (2012). Figure 3 presents crop masking data for maize.

### *Using maize mask data to create independent variables*

To establish the independent variables' predictive power, all variables were retrieved using crop mask data for further correlation analysis and to discover significantly linked ones with maize yield. The time series data for the Normalized Differential Vegetation Index (NDVI) (120 decadal) were image preprocessed in one step and were ready for monthly maximum value compositing (MVC).In ArcGIS, a tool called 'Cell Statistics' is found in the Spatial Analyst toolbox. You will be adding a lot of rasters, including the MODIS NDVI (Moderate Resolution Imaging Spectroradiometer Normalized Difference Vegetation Index) for June–September. The'maximum'

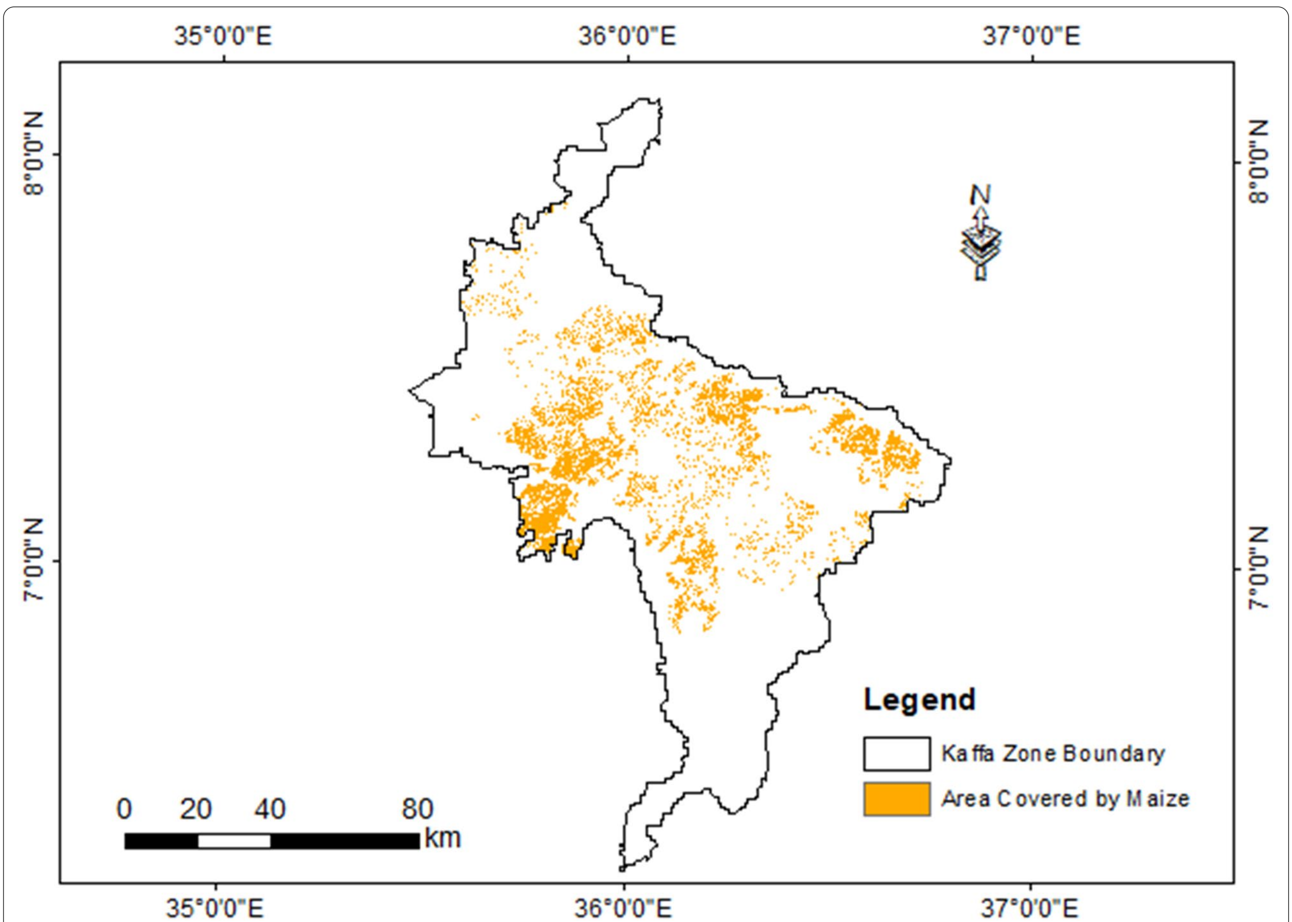


**Fig. 3** Crop mask data for Maize

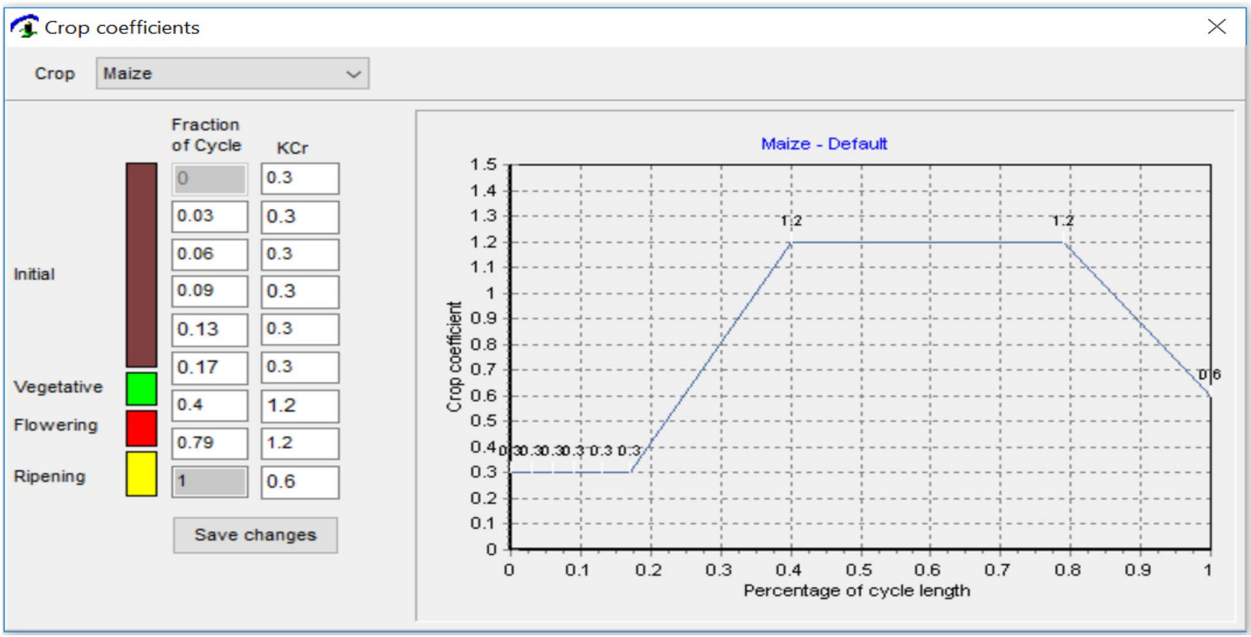


**Fig. 4** Maize crop coefficients at various stages (Planting-Flowering) ( Source: LEAP software)

option was chosen, resulting in 40 monthly composited normalized difference vegetation index (NDVI) images. These monthly Normalized Difference Vegetation Index (NDVI) images were then removed using crop mask data to focus just on the crop of interest, and an Average Normalized Difference Vegetation Index (NDVIa) value was calculated for each year. The calculated result was in raster format, ranging from 0 to 255, and had to be converted to normalized difference vegetation index (NDVI) format. As a result, Gidey et al. (2018) utilized the formula eMODIS NDVI = Float (Smoothed eMODIS NDVI—100)/100, and the results were ready to be associated with maize production (Table 5). These monthly Normalized Difference Vegetation Index (NDVI) images were extracted with crop mask data to focus only on the crop of interest, and Climate Hazards Group Infrared Precipitation With Station Data (CHIRPS) time series data of decadal image was composited at monthly level using monthly maximum value compositing (MVC) and extracted with crop mask data for further analysis (Table 5). The Water Requirement Satisfaction Index (WRSI) model is a ratio of seasonal actual crop evapotranspiration (ETA) to seasonal crop water requirement, which is the same as potential crop evapotranspiration (PETc). For the phonological from planting to flowering, the maize crop coefficient from the livelihood early

**Table 5** Observed yields and independent variables

| NO | Year (meher season) | maize Yield in(qt/ht) | NDVIa | Eta | Eta total | WARSI | CHERIPS |
|---|---|---|---|---|---|---|---|
| 1 | 2008 | 17.5 | 0.78 | 38.35 | 135.94 | 136.77 | 49.35 |
| 2 | 2009 | 20.2 | 0.84 | 39.59 | 135.07 | 138.43 | 50.99 |
| 3 | 2010 | 21.76 | 0.84 | 39.57 | 136.47 | 157.01 | 59.86 |
| 4 | 2011 | 25.26 | 0.94 | 38.36 | 130.34 | 144.19 | 60.89 |
| 5 | 2012 | 21.84 | 0.85 | 37.99 | 132.64 | 155.15 | 64.87 |
| 6 | 2013 | 25.66 | 0.95 | 39.17 | 133.72 | 152.43 | 63.48 |
| 7 | 2014 | 28.49 | 0.95 | 39.13 | 136.26 | 151.09 | 65.40 |
| 8 | 2015 | 29.93 | 0.97 | 39.80 | 137.52 | 154.37 | 70.28 |
| 9 | 2016 | 29.51 | 0.95 | 37.69 | 128.50 | 144.94 | 76.18 |
| 10 | 2017 | 29.3 | 0.87 | 38.42 | 133.7 | 140.51 | 69.97 |

assessment protection (LEAP) software was used (Initial 0.3, Vegetative1.15, Flowering1.15, and Ripening 0.55) (Fig. 4).

#### *Multiple linear regression analysis*

The statistical method of regression analysis is used to estimate the relationships between variables. Establishing a link between an independent variable (indicator or predictor) and a dependent variable is typical practice in forecasting (crop yield). This study aids us in identifying the indicator that best explains the behavior of agricultural yields. A statistical strategy for predicting a dependent variable from a set of independent variables is known as multiple regression analysis (Bekele 2015). The data from Table 5 was used to run Multiple Linear Regression.

There have been some assumptions using during this statistic: -

a. The regression analysis method relies on the availability of lengthy and consistent time series of remote sensing data and agricultural statistics. The latter are frequently merged at the national/subnational administrative unit level, allowing for the generation of average NDVI values.

b. The criterion variable was believed to be a random variable.

c. Instead of a functional relationship, a statistical relationship (estimation of the average value) would be established (calculating an exact value).

d. The relationship between the dependent and each independent variable is deemed linear in multiple linear regressions. The linearity assumption can be tested using scatter plots (Osborne and Waters 2002). As a consequence of the multiple regression analysis, the prediction equation (Eq. 6) is as follows:

$$\mathrm{Y} = \beta 0 + \beta 1\mathrm{x}1 + \beta 2\ \mathrm{x}2 + \cdots + \beta \mathrm{n}xn + \epsilon \tag{6}$$

where, β0 is constant; β1, β2… βn is beta coefficient or standardized partial regression coefficients (reflecting the relative impact on the criterion variable), × 1, × 2, x n is scores on different predictors. When the associated independent variable changes by one unit, the regression coefficients are the quantities by which the dependent variable y changes. When all of the independent variables are zero, the dependent y will be 0 and the regression line will intercept the y axis. The ratio of the beta coefficients is the ratio of the independent variables' relative predictive power, while the beta weights are a standardized form of the coefficients (Linear regression analysis, Yan and Su 2009). The developed model predicts the average value of one variable (Y) based on the value of another variable (X). The X variable is also known as a predictor. A regression model is the name given to this type of model (Fig. 5).

## Results and discussions

### Developing multiple linear regression model equation for maize yield forecasting in the study area

The monthly maximum value composite (MVC) averages of normalized difference vegetation index average (NDVIa) from the planting date to the end of the crop cycle have a correlation coefficient of 0.84 with a significant P value of 0.002 at 95 percent confidence level, while rainfall has a correlation coefficient of 0.89 with a significant P value of 0.0001 at 95 percent confidence level. Actual crop evapotranspiration (ETA), with a correlation value of 0.024 and a significant P value of 0.942 at 95 percent confidence level, Eta total, with a correlation value of 0.22 and a significant P value of 0.537 at 95 percent confidence level, and water requirement satisfaction index (WRSI) ($r = 0.258$) with a P value of 0.472, which is beyond the acceptable range at 95 percent confidence level, were all rejected from the model development. As a result, to develop a multiple linear regression model, the two most associated variables normalized difference vegetation index average (NDVIa) and climatic hazards group infrared precipitation with station data (CHIRPS) rainfall with the dependent variable (yield) are chosen. According to numerous crop forecasting studies, linear regression modeling is the most common method for generating yield forecasts using remote sensing derived indicators and bioclimatic data. Maize yield data and data from various variables were generated for multiple linear regression analysis. Utilizing the Statistical Package for Social Science (SPSS) software, a multiple linear regression model was created using the two most associated variables. The model closely connected variables normalized difference vegetation index average (NDVIa) and climatic hazards group infrared precipitation with station data (CHIRPS)rainfall were used to construct a model as a result of all of the preceding operations. As shown in the table, the coefficient of determination ($R^2$), root mean square error (RMSE), and coefficient of variation (CV) of this model were all used to validate it (Fig. 6).

When we plot the actual yield per hectare vs the expected yield per hectare to see how well the model fits, we can see that most areas are quite close to the 45° line (exact prediction line). With a root mean square error of 1.54 quintal per hectare, the model's R square value is 0.89; adjusted R square is 0.88. The model's P value is 0.0001 at a 95% confidence level.

Based on this P value, it is unclear which independent variable is a very good predictor and which is a poor predictor. According to Table 6, the analysis of variance, the maize yield forecast model has an observed significance probability ($\mathrm{Prob} > \mathrm{F}$) of 0.0001, which is significant at the 0.05 level. Because of the p0.0001, we conclude that Yield is connected to the average normalized difference

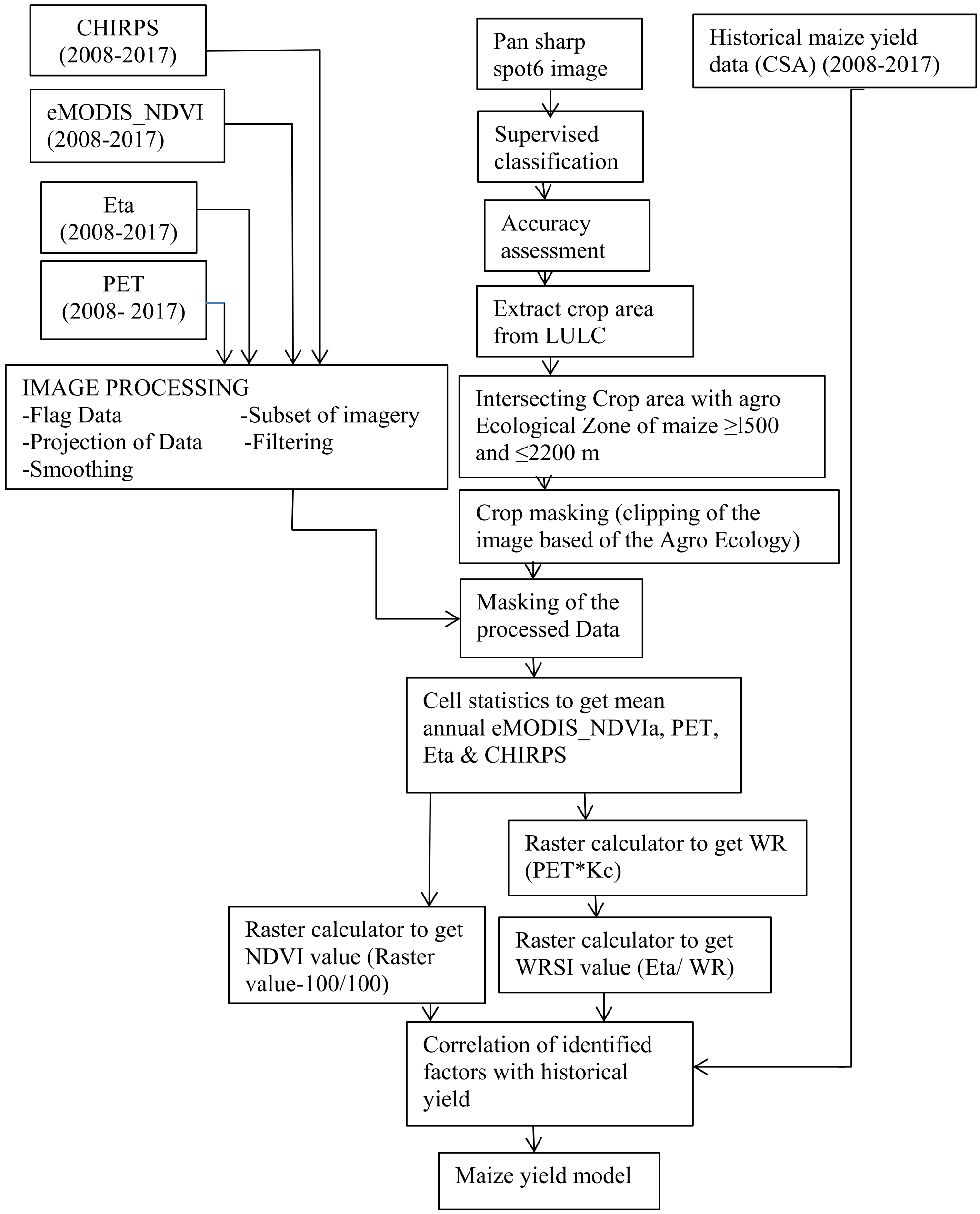


**Fig. 5.** Flow chart of the methodology

vegetation index (NDVIa) and/or climatic hazards group infrared precipitation with station data (CHIRPS). According to the normalized difference vegetation index average (NDVIa) and climatic hazards group infrared precipitation with station data (CHIRPS) rainfall have a variance inflation factor (VIF) of 1. 992. There is no multicollinearity between these two variables because the Variance Inflation Factor (VIF) is less than 10 (Table 6). As a result, normalized difference vegetation index average (NDVIa) and climate hazards group infrared precipitation with station data (CHIRPS) rainfall were chosen for model development in this study. Table 6 shows values of -20.375, 28.360, and 0.316 for the intercept (constant term), normalized difference vegetation index average (NDVIa), and climatic hazards group infrared precipitation with station data (CHIRPS) rainfall, respectively. Rainfall, normalized difference vegetation index average (NDVIa), and Intercept become extremely significant

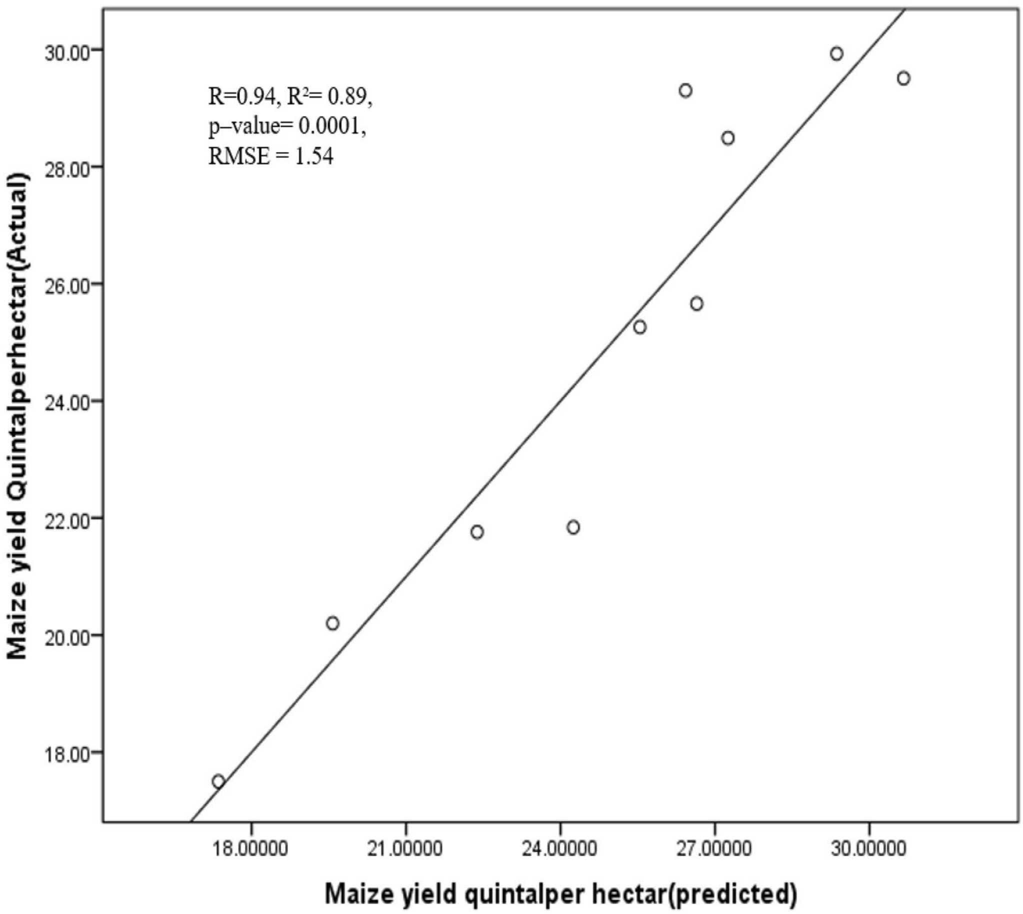


**Fig. 6** Comparison of maize yields predicted by the agro meteorological model and actual yields in the study area

**Table 6** Results of the variance analysis; the Variance Inflation Factor; and the parameter estimations for the model

| Model | | Sum of Squares | | Df | Mean Square | F | Sig |
|---|---|---|---|---|---|---|---|
| Maize yield forecast model variance analysis (ANOVA) | | | | | | | |
| Regression | | 156.49 | | 1 | 156.49 | 65.643 | 0.000 |
| Residual | | 19.072 | | 8 | 2.384 | | |
| Total | | 175.561 | | 9 | | | |
| Variance Inflation Factor (VIF) between NDVIa and CHIRPS | | | | | | | |
| Constant | | 0 | | | | | |
| NDVIa | | 1.992 | | | | | |
| CHIRPS | | 1.992 | | | | | |
| **Model** | | **Unstandardized Coefficients** | | | **Standardized Coefficients** | **T** | **Sig** |
| | | **B** | **Std. Error** | | **Beta** | | |
| Maize forecast model's parameter estimates | | | | | | | |
| 1 | Constant | -20.375 | 7.696 | | | -2.647 | 0. 033 |
| | CHIRPS | 0.316 | 0. 0.093 | | 0. 599 | 3.407 | 0. 048 |
| | NDVIa | 28.360 | 11.851 | | 0. 4211 | 2.393 | 0. 011 |

within the model. A unit change in normalized difference vegetation index average (NDVIa) and climatic hazards group infrared precipitation with station data (CHIRPS) rainfall resulted in yield changes of 28.360 and 0.316 unit times, respectively. As a result, the multi linear regression model equation for maize yield forecasting is (Eq. 7):

$$\begin{aligned}&\text{Predicated Maize yield (qt/ha)}\\&\quad= -20.375 + (28.360 * \text{ NDVIa})\\&\qquad+ (0.316 * \text{CHIRPS})\end{aligned} \tag{7}$$

According to the Agricultural Production report of the Central Statistical Agency, the coefficient of variation of maize yield is 17.7%, which is within the allowed range of validation values.

## Comparing the accuracy level of maize crop yield forecast using model and Central Statistics at the ground level in the study area

When the subjectivity of traditional and remote sensing yield forecasts is compared, the remote sensing approach succeeds. According to a report by the Central Statistical Agency (CSA), the forecast data, which is a result of the conventional approach, has a coefficient of variance of 17.7% and is a subjective approach. The remote sensing-based model, on the other hand, forecasts 16.7% with a high level of confidence (95%) and a high probability value. Furthermore, because September is the maize crop's flowering stage, the remote sensing-enabled methodology's forecast result might be supplied as early as early October, whereas the traditional method's data release date is normally in December and includes all cereal crops. Despite the fact that we did not consider all grains covered by the Central Statistical Agency (CSA) in my research, this shows that the timeliness issue could be addressed more effectively by using a remote sensing-aided strategy rather than the traditional approach.

Another benefit of the remote sensing-based approach is that it provides location information, as the forecast can be verified by taking GPS measurements and going to the locations once it is prepared. As a result, whilst standard methods fail horribly, this method gives for a precise indication of which locations have a high and low yield in a tangible manner. As a result, it is clear that using remote sensing and a geographic information system (GIS) to anticipate maize production improves data quality and timeliness while lowering subjectivity. This research and other similar studies have proven that a remote sensing-enabled approach can reveal locales (lower administrative areas) where there is comparatively high, medium, and low production, making decision-making much easier. A comparison of standard yield estimations and the Remote Sensing aid technique is shown in Fig. 7.

## Testing the model for predicting maize yields for the year 2018 in the study area

The 2018 maize crop forecast was created using the developed model. With a mean of 20 $qha^{-1}$, the maximum maize yield for 2018 is expected to be 25 $qha^{-1}$ and the lowest 15 $qha^{-1}$. Maize yields are expected to be 10–15 $qha^{-1}$ in 6.1 percent of the study area, 15–19 $qha^{-1}$ in 50.3 percent of the area, and 20–25 $qha^{-1}$ in the remaining 43.6 percent of the study area, according to the prediction (Table 7). Certain pockets of the study area, such as Gesha, Sayilem, Gimbo, Gewata, and Menjwo district, are most productive with 20—25 $qha^{-1}$ of yield, while the western, south-eastern, and central parts of the Zone, Bita, Cheta, Talo, and Bonga town zuria weredas, are intermediately productive with 15–19 $qha^{-1}$ of output. The rest of the study area has low-yielding patches that produce only 10–15 $qha^{-1}$ of grain. As a result, the zone's northwestern, north-eastern, northern, and eastern parts were more productive than the rest of the study area (Fig. 8).

## Discussion

Despite the crop is uniqueness, we attempted to compare our results to those of earlier studies in terms of relevant research. According to Zinna and Suryabhagavan (2016), in the multiple linear regression model, the Normalized Difference Vegetation Index Average (NDVIa) and rainfall parameters were retained as significant variables for field level yield prediction, explaining 88 percent of the yield variability, implying that rainfall and Normalized Difference Vegetation Index Average (NDVIa) are the best parameters for yield prediction. Meanwhile, according to this study, the Normalized Difference Vegetation Index Average (NDVIa) and rainfall are kept in the model, accounting for 89 percent of yield variation.

Rojas (2007) conducted a maize yield forecast in Kenya for a maize crop, and the most important components for building a multiple linear regression model were evapotranspiration total and NDVIc. The ETa total model explained 83 percent of the yield variance (RMSE = 0.333 $tha^{-1}$ and CV = 21 percent), while the NDVIc model explained 87 percent (RMSE = 0.333 $tha^{-1}$ and CV = 21 percent), demonstrating that spectro-agro meteorological models can be used to model even fragmented agricultural lands like this one. Due to different geo-climatic circumstances, the evapotranspiration total was not qualified for inclusion in the model (because to its insignificant p value association with maize yield and its minute correlation coefficient).

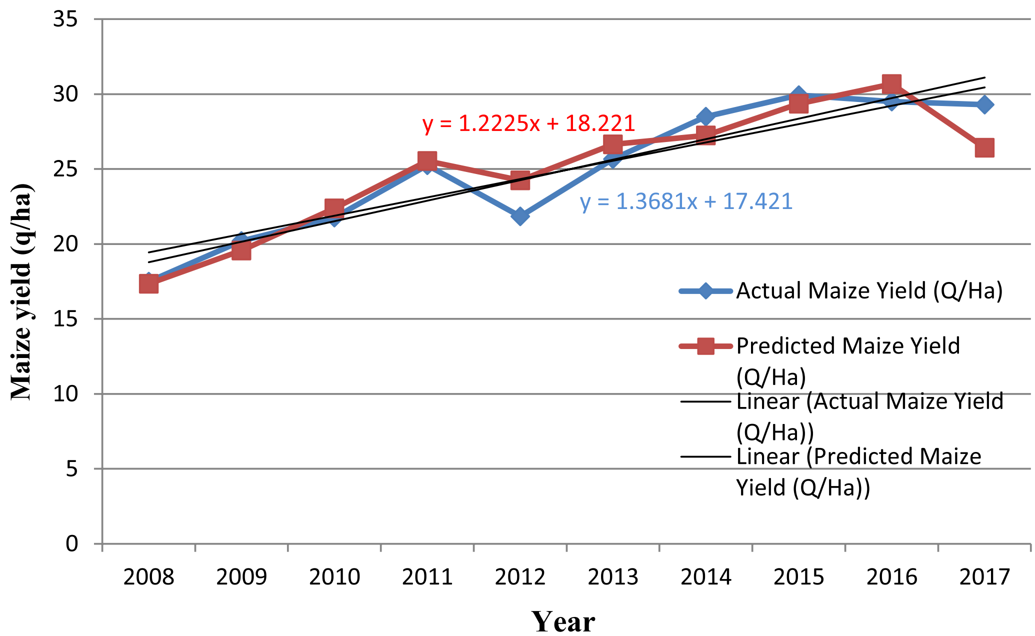


**Fig. 7** Comparison of the model's estimated maize yield (quintal/ha) with the actual yield

The prediction power of the model in this investigation was high (Root mean square error = 1.54 and $R^2$ = 89%). The magnitude of the results is nearly identical when compared to Zinna and Suryabhagavan (2016) (Root mean square error = 1.41 and $R^2$ = 88 percent) and Reda (2015) (Root mean square error = 0.99 and $R^2$ = 93 percent) South Tigray Zone maize yield forecast and east Arsi Zone wheat yield forecast. Rainfall (r = 0.89) is the most strongly correlated independent variable with yield, followed by the Normalized Difference Vegetation Index Average (NDVIa) (r = 0.84). Nonetheless, in Reda, the 2015 normalized difference vegetation index average (NDVIa) and rainfall (r = 0.89) are significantly related (r = 0.96). This demonstrates that yield prediction parameters range from one agro ecological zone to the next, implying that our model takes into account a variety of factors in determining varied correlation outcomes.

The Water Requirement Satisfaction Index (WRSI) and Actual Evapotranspiration (Eta) were not connected to yield in this study, comparable to Zinna and Suryabhagavan (2016) and Reda (2015). Similar to Zinna and Suryabhagavan's work, the normalized difference vegetation index average (NDVIa) and rainfall are chosen for the final model based on Statistics results, however rainfall is deleted from Reda's (2015) article based on the Variance Inflation Factor (VIF) result. Following Zinna and Suryabhagavan (2016) maize crop yield forecast research and Reda (2015) wheat crop yield forecast research, the findings of this study reveal that agro metrological characteristics have a definite potential for maize yield forecasting in the kaffa zone.

## Conclusions

Crop yield forecasting is essential for addressing the challenges provided by climate change's impact on agriculture. By improving the timeliness and accuracy of yield forecasting, we can improve our ability to respond effectively to these challenges. The major purpose of this study was to develop a maize crop model using remote sensing and Geographic Information Systems (GIS). Crop statistical data was employed as a dependent variable, and many predictor factors derived from remotely sensed imageries were calculated, with the variables with the highest correlation and significant P values chosen for model construction. The investigation's findings revealed that the Normalized Difference Vegetation Index Average (NDVIa) and Climate Hazards Group Infrared Precipitation with Station Data (CHIRPS) rainfall for the study area have excellent correlations of r = 0.84 and r = 0.89, respectively, with a significant P value confirming the result. Using these correlation results, agro meteorological yield forecasting using a multiple linear regression was developed using a table of data containing yields as a dependent and a series of agro meteorological and remote sensing variables that have a high correlation with the yield. The created agrometric model has a prediction capability of 0.89 quintal per hectare and an RMSE of 1.54 quintal per hectare, which is a good result. In an area

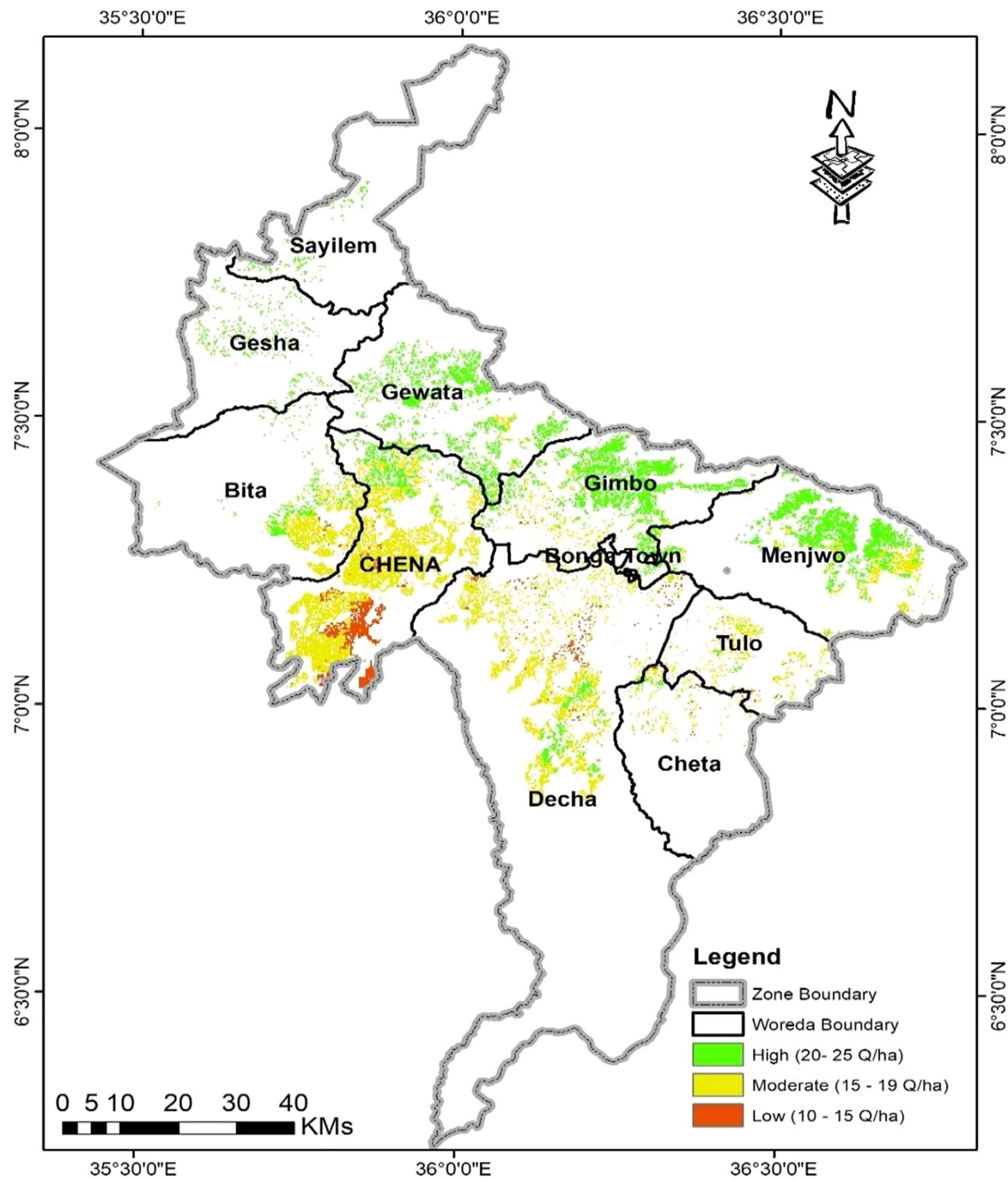


**Fig. 8** Maize yield forecast map of 2018

like the Kaffa Zone, where land is fragmented, it can be argued that using proven yield forecasting methodologies and remotely sensed data, a reasonably precise forecast can be formed. Using the regression model developed for the research area, maize production predictions may be done pretty far ahead of the harvest date. The developed

**Table 7** Maize production level of the year 2018 for kaffa zone

| Maize production level | Quintal per/ha | Area coverage in (%) |
|---|---|---|
| I | 20–25 | 43.6 |
| II | 15–19 | 50.3 |
| III | 10–15 | 6.1 |

model was also used to create a maize yield forecast map for the year 2018, with an average result of 20 quintal per hectare, indicating that the Zone's northwestern, northeastern, northern, and eastern parts have high productivity per hectare and can be used by decision makers to identify relative productive areas prior to harvest at the lower administration level. Normalized Difference Vegetation Index Average (NDVI) generated from Moderate Resolution Imaging Spectroradiometer (eMODIS) and Climate Hazards Group Infrared Precipitation with Station Data (CHIRPS) rainfall can generally be used to forecast maize yields in areas similar to the kaffa zone.

Following the methods indicated in the research methodology, the created model can be tested in areas other than the kaffa zone, however more research and testing is required. Additional work is needed to operationalize the findings of this study, which include: a longer period of time series data should be reviewed in order to reach a practical application. Other elements, such as soil, should be included in future study, and Instead of the Multiple Linear Regression Model, other models such as polynomial regression and non-linear regression should be used. More research, as well as improved remote sensing and GIS technologies, are needed to identify additional factors that contribute to production variability.

**Acknowledgements**
We are really grateful to Bonga University for providing material support in order to accomplish this research. We also received secondary data from the National Meteorology Agency (NMA), Ethiopian Geospatial Information Agency (EGIA), and the Central Statistical Agency (CSA). Our gratitude also goes to Mr. Molla Maru from geography department Addis Ababa University for his assistance in the data analysis and for their unwavering support, sharing of information, and important advice in every aspect of our project.

**Authors' contributions**
DB and JT conceived and designed the method section's proven work. The author contributes to the article's analysis, verification, and writing. Both authors read and approved the final manuscript.

**Funding**
This study received no funding from any institutions, agencies, or people.

**Availability of data and materials**
On reasonable request, the corresponding author will provide the datasets created and/or analyzed during this study.

## Declarations

**Ethics approval and consent to participate**
Not applicable.

**Consent for publication**
Not applicable.

**Competing interests**
The corresponding author declares, on behalf of all authors, that there is no conflict of interest in this scientific activity.

**Author details**
[1] Department of Geography and Environmental Studies, College of Social Sciences, Bonga University, Bonga, Ethiopia. [2] Department of Geography and Environmental Studies, College of Social Sciences, Bonga University, P.O. Box 334, Bonga, Ethiopia.

Received: 7 November 2021 Accepted: 22 February 2022
Published online: 03 March 2022

**References**

Basso B, Cammarano D, Carfagna E (2013) Review of crop yield forecasting methods and early warning systems. In Proceedings of the First Meeting of the Scientific Advisory Committee of the Global Strategy to Improve Agricultural and Rural Statistics, FAO Headquarters, Rome, Italy, 18–19 July 2013

Becker-Reshef I, Vermote E, Lindeman M, Justice C (2010) A generalized regression-based model for forecasting winter wheat yields in Kansas and Ukraine using MODIS data. Remote Sens Environ 114(6):1312–1323

Bekele F (2015) *Characterizing Current and Future Rainfall Variability and its Effect on Wheat and Barley Production in Sinana District, South Eastern Ethiopia* (Doctoral dissertation, Doctoral dissertation, M. Sc thesis (published), Haramaya University, Haramaya. English Abstract). Advances in Water Science, 14: 91–96)

Congalton RG, Green K (2019) Assessing the accuracy of remotely sensed data: principles and practices. CRC Press

Duveiller G, Baret F, Defourny P (2012) Remotely sensed green area index for winter wheat crop monitoring: 10-Year assessment at regional scale over a fragmented landscape. Agric Forest Meteorol 2012(1):156–168

Faisal B, Rahman H, Sharifee N, Sultana N, Islam M, Ahammad T (2019) Remotely Sensed Boro Rice Production Forecasting Using MODIS-NDVI: A Bangladesh Perspective. AgriEngineering 1:356–375

Gidey E, Dikinya O, Sebego R, Segosebe E, Zenebe A (2018) Modeling the spatio-temporal meteorological drought characteristics using the standardized precipitation index (SPI) in Raya and its environs Northern Ethiopia. Earth Syst Environ 2(2):281–292

Gorfu D, Ahmed E (2012) Crops and agro-ecological zones of Ethiopia. Ethiopian Institute of Agricultural Research

Greatrex H (2012) *The application of seasonal rainfall forecasts and satellite rainfall estimates to seasonal crop yield forcasting for Africa* (Doctoral dissertation, University of Reading)

Hatfield JL, Prueger JH (2010) Value of using different vegetative indices to quantify agricultural crop characteristics at different growth stages under varying management practices. Remote Sens 2010(2):562–578

Johnson DM (2014) An assessment of pre- and within-season remotely sensed variables for forecasting corn and soybean yields in the United States. Remote Sens Environ 141:116–128

Kouadio L, Newlands NK, Davidson A, Zhang Y, Chipanshi A (2014) Assessing the performance of MODIS NDVI and EVI for seasonal crop yield forecasting at the ecodistrict scale. Remote Sensing 6(10):10193–10214

Lillesand T, Kiefer RW, Chipman J (2015) Remote sensing and image interpretation. Wiley

Mkhabela MS, Bullock P, Raj S, Wang S, Yang Y (2011) Crop yield forecasting on the Canadian Prairies using MODIS NDVI data. Agric for Meteorol 151(3):385–393

Osborne JW, Waters E (2002) Four assumptions of multiple regression that researchers should always test. Pract Assess Res Eval 8(1):2

Reda AF (2015) *Wheat Yield Forecast Using Remote Sensing and GIS in East Arsi Zone, Ethiopia* (Doctoral dissertation, Addis Ababa University)

Rijks O, Massart M, Rembold F, Gommes R, Leo O (2007)Crop and rangeland monitoring in eastern Africa. In: Proceedings of the 2nd International Workshop (pp. 95–104)

Rojas O (2007) Operational maize yield model development and validation based on remote sensing and agro-meteorological data in Kenya. Int J Remote Sens 28(17):3775–3793

Senay GB, Bohms S, Singh RK, Gowda PH, Velpuri NM, Alemu H, Verdin JP (2013) Operational evapotranspiration mapping using remote sensing and weather datasets: a new parameterization for the SSEB approach. J Am Water Resourc Assoc 49(3):577–591

Vintrou E, Desbrosse A, Bégué A, Traoré S, Baron C, Seen DL (2012) Crop area mapping in West Africa using landscape stratification of MODIS time series and comparison with existing global land products. Int J Appl Earth Obs Geoinf 14(1):83–93

Yan X, Su X (2009) Linear regression analysis: theory and computing. World Scientific

Yan G, Mas JF, Maathuis BH, Xiangmin Z, Van Dijk PM (2006) Comparison of pixel-based and object-oriented image classification approaches—a case study in a coal fire area, Wuda, Inner Mongolia China. Int J Remote Sens 27(18):4039–4055

Zinna AW, Suryabhagavan KV (2016) Remote Sensing and GIS Based Spectro-Agrometeorological Maize Yield Forecast Model for South Tigray Zone Ethiopia. J Geogr Inform Syst 8(2):282–292

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.