Maize Yield Forecast Using GIS and Remote Sensing The Case of Kaffa Zone, South Western Ethiopia

Dereje Biru (derejebiru22@gmail.com)  
Bonga University  https://orcid.org/0000-0003-1720-6219

Jemal Tefera  
Bonga university

Research

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

Posted Date: November 22nd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1057409/v1

License: ©  This work is licensed under a Creative Commons Attribution 4.0 International License.  Read Full License
Background: Policy makers, government planners and agriculturalist in Ethiopia require accurate and timely information about maize yield and production. Kaffa zone is by far the most important maize producing zone in the country. The manual collection of field data and data processing for crop forecasting by the CSA requires significant amounts of time before official reports are released. Several studies have shown that maize yield can be effectively forecast using satellite remote sensing data. The objectives of this study were to develop a maize yield forecast model in kaffa Zone derived from time series data of eMODIS_NDVI, actual and potential evapotranspiration and CHIRPS for the years 2008-2017. Official grain yield data from the Central statistical Agency of Ethiopia was used to validate the strength of the indices in explaining the yield. Crop masking at crop land area was applied and refined by using agro ecological zones suitable for the crop of interest. Correlation analyses were used to determine associations among crop yield, spectral indices and agro meteorological variables for maize crop of the long rainy season (kiremt). Indices with high correlation with maize yield were identified.

Results: Average Normalized Difference Vegetation Index and rainfall have high correlation of maize yield with 84% and 89%, respectively. That means their variables are positively strong related with maize yield. The generated spectro-agro meteorological yield model was successfully tested against the Central Statistical Agency’s expected Zone level yields ($r^2 = 0.89$, RMSE = $1.54\text{qha}^{-1}$, and 16.7% coefficient of variation).

Conclusions: Thus, remote sensing and geographical information system based maize yield forecast improved quality and timelines of the data besides distinguishing yield production levels/areas and making intervention very easy for the decision makers there by proving the clear potential of spectro-agro meteorological factors for maize yield forecasting, particularly for Ethiopia.

Background

Crop yield prediction is critical for planning and making various policy decisions. Many countries rely on traditional data gathering techniques such as ground-based visits and reports for crop monitoring and yield forecast. These reporting processes are subjective, costly, time-consuming, and prone to huge errors due to insufficient ground observation, which results in inaccurate crop production assessments and a delay in reporting necessary actions (Greatrex, 2012). Before the advent of remote-sensing methods like NDVI, crop-weather models were utilized for crop monitoring and yield forecasts (Rojas, 2007). Crop data was collected on the ground in the Kaffa zone, which is a challenging task that requires human resources, money, and time. Remote sensing is more significant than ground survey in reducing these issues. Because remote sensing can give precise and timely data for crop production estimation, most studies have found a link between NDVI, agro meteorological data, green biomass, and yield (Rojas, 2007).

Many studies in Ethiopia have been conducted using this techniques for crop yield forecasting at various zonal levels; Abiy (2014) conducted a study in the south Tigray Zone for maize crop forecasting using time series data from SPOT VEGETATION, actual and potential evapotranspiration, and rainfall estimate satellite data for the years 2003-2012, and Akililu (2015) conducted a study in the Arsi zone for wheat crop forecasting using time series data from SPOT VEGETATION, However, both studies used SPOT VEGETATION NDVI and RFE 2.0, which cover large areas with low-resolution (1km) and low-resolution (10km), respectively, rather than eMODIS NDVI, which is a relatively better data set used in crop monitoring, due to the length of the time series (since 2000) and spatial resolution (250 meters), as well as the fact that it is freely available and easily accessible, and CHIRPS rainfall, which data are available from 1981 dekedal As a result, the researchers wanted to solve this research vacuum by developing a model that uses eMODIS NDVI and CHIRPS satellite rainfall to anticipate Maize output for the year 2018.

Materials And Methods

Description of study area

This research was carried out in the Kaffa Zone, which is located in the South Western region of 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 km2, accounting for 7.06 percent of the region’s total area. Kaffa Zone is divided into twelve administrative districts and contains three traditional climate zones based on altitude and temperature variances. Highland (2500-3000 m), midland (1500-2500 m), and lowland (500- 1500 m) are the three types. Highland, midland, and lowland occupy 11.6 percent, 59.5 percent, and 28.9 percent of the Zone’s total area, respectively. According to national meteorology agency, the average yearly 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 yearly rainfall ranges from 1001-2200mm. The Kaffa Zone is located in Ethiopia’s South West area, which receives the most rainfall. This is due to the existence of an evergreen forest cover on top of the wet monsoon winds’ windward location.

Source of data and software’s used

It is critical to determine the sources and types of data in order to meet the study’s objectives. The information for this study was gathered from both primary and secondary sources. Primary data is made up of information gathered from satellite imagery and observations on the ground. Books, topographic and thematic layers, periodicals, Meteorological Agency and Central Statistical Agency reports, as well as other publications and scholarly works, are all examples of secondary data sources. Different softwares were also employed to analyze these data sets.
| No | Types of data               | Source | Description | Purpose                                                      |
|----|-----------------------------|--------|-------------|--------------------------------------------------------------|
| 1  | Study area boundary         | CSA, 2007 | Shape file  | Used to demarcate the study area boundary                   |
| 2  | DEM                         | USGS   | Aster 30m resolution | For maize elevation generates.                               |
| 3  | Pan sharpened SPOT 6 image  | EGIA   | Landsat8    | For supervised LULC classification                           |
|    |                             |        | Spatial resolution of 30 meters (visible, NIR, SWIR); 100 meters (thermal); and 15 meters (panchromatic). | For supervised LULC classification |
|    |                             |        | Spatial resolution of 1.5m Panchromatic and 6m Multispectral | For supervised LULC classification |
|    |                             |        | 12-bit Radiometric resolution | For supervised LULC classification |
|    |                             |        | 16 days Temporal resolution | For supervised LULC classification |
|    |                             |        | Path/row - 170/55 Spot 6 | For supervised LULC classification |
| 4  | CHERIPS rainfall            | NMA    | Spatial resolution: (0.05°×0.05°) | Used to calculate yearly average rainfall to correlate it with maize yield. |
|    |                             |        | Frequency: dekadal | For supervised LULC classification |
|    |                             |        | Archive:2008-2017 | For supervised LULC classification |
|    |                             |        | Format:netCDF | For supervised LULC classification |
| 5  | potential evaporation(pet)  | NMA    | 1km×1km Spatial resolution | Used to calculate WRSI to correlate it with maize yield. |
|    |                             |        | Frequency: dekadal | For supervised LULC classification |
|    |                             |        | Archive:2008-201 | For supervised LULC classification |
|    |                             |        | Format:netCDF | For supervised LULC classification |
No | Types of data | Source | Description | Purpose |
---|---|---|---|---|
6 | eMODIS NDVI | Downloading from https://earlywarning.usgs.gov/fews/datadownloads/East%20Africa/eMODIS%20NDVI%20C6 | • 250m*250m special resolution  
• 12bits radiometric resolution  
• 1-2days Temporal resolution  
• Frequency: dekadal  
• Archive:2008-2017 | Used To calculate yearly average NDVI to correlate it with maize yield. |
7 | Eta (actual evaporation) | Downloaded freely from EWSNET http://earlywarning.usgs.gov/fews/downloads/index.php? | • 1km*1km special resolution  
• 8 bit spectral resolution  
• 16 days Temporal resolution  
• Frequency: dekadal  
• Archive: 2008-2017 | Used to calculate yearly average Eta, Eta total and WRSI to correlate these variables with maize yield |
8 | Maize yield(qt/ht) | CSA annual agricultural report, 2018 | • Archive data from 2008-2017 | To calibrate the developed model with historical crop yield statistics |
9 | Ground Truth and Accuracy Assessment Points | Bonga University | Random coordinates from each land use using HHGPS Garmin62 | For accuracy assessment of the supervised classification |

**Table 2**

Summary of equipment and materials used for data collection and analysis.

| Software used | Purpose |
|---|---|
| GPS(Global Position System) | For collecting of GCP points which will be created at random for the study area using Arc GIS 10.3 used mainly for accuracy assessment area measurement  
GPS is from Bonga University. |
| 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 |

**Data processing and analysis**

**Classification**

The pan sharpened SPOT 6 image is processed for supervised classification in ArcGIS software. According to Yan et al. (2006), supervised categorization necessitates the user specifying the various pixels values or spectral signatures that should be associated with each class. This is performed by identifying Training Sites or Areas, which are typical sample sites of known cover types. To construct the thematic map of Land cover and to identify the Land use land cover classification of the research area, the maximum likelihood classifier (MLC) was used to classify land cover into two classes (agricultural and non-agriculture) (Figure 2).

The accuracy of a map created from remote sensing data must be assessed. The most popular technique to communicate the accuracy of categorization results is using an error matrix. The error matrices were used to calculate overall accuracy, user and producer accuracies, and the Kappa statistic. The Kappa
statistic integrates the error matrices’ off diagonal portions and represents agreement after reducing the fraction of agreement that may be anticipated to happen by chance. As a result, the above-mentioned classifications (agricultural and non-agriculture) were represented evenly. The enough number of samples that represent the thematic classes and ensure good distribution across the map is important to test the attribute accuracy. As a rule of thumb Congalton et al. (2008) recommends at least 50 samples per class. If the area exceeds 500km² or the number of categories is more than 12, then at least 75-100 samples should be taken per class. These recommendations coincide with those recommended by Fenstermaker (1991). The number of samples for each category might be adjusted based on the relative importance of that category for a particular application. To verify attribute correctness, there must be a sufficient number of samples that represent the thematic classes and are distributed evenly across the map. Congalton et al. (2008) suggests at least 50 samples each class as a general guideline. If the region is greater than 500km² or the number of categories is greater than 12, at least 75-100 samples per class should be taken. These suggestions are similar to those made by Fenstermaker (1991). Depending on the relative relevance of each category for a given application, the quantity of samples for each category may be changed. Furthermore, sampling could be assigned based on the degree of variation within each category (Congalton et al., 2008). As a result, the accuracy assessment sample size was determined to be 200, with 100 sample points created for each class. Then, for each class, these spots were produced at random and their GPS readings were placed onto a GPS for field accuracy testing (Figure 3).

These points were verified in two ways: those that were visible and accessible in the field, and those that were verified using Google Earth as a reference. As a result, for the 200 sample points, the following error matrix (Table 3) is presented. The overall accuracy and kappa analysis were used to complete a classification accuracy evaluation, and the overall accuracy of the data is 90.0 percent, with a kappa coefficient of 0.80, and the interpretation may be taken as correct for further analysis based on the result.

| Ground Truth data | User Accuracy |
|-------------------|---------------|
| Map data          |               |
| Agricultural      | 8             | 96  | 91.7 |
| Non agricultural  | 12            | 92  | 104 | 88.5 |
| Total             | 100           | 100 | 200 |
| Producer accuracy | 88            | 92  |     |

Maize Crop Mask Data Derivation

Crop agro-ecology in the research area is another input for disguising crop data. According to Gorfu and Ahmed (2012), maize is primarily grown between the elevations of 1500 and 2200masl, i.e. Figure 4 shows crop masks data for maize.

Preparing Independent Variables Using Mask Data of Maize.

To determine the predictive capability of the independent variables, all variables were extracted with crop mask data for further correlation analysis and to identify highly correlated ones with maize yield. The time series data (120 decadal) of NDVI have undergone image preprocesses in one goes were ready for monthly maximum value compositing (MVC). In ArcGIS there is a tool called ‘Cell Statistics’ under Spatial Analyst toolbox. You will add multiple rasters, which during this case is MODIS NDVI June-sept. Select the ‘maximum’ option and 40 monthly composited NDVI images were prepared. These monthly NDVI images were then extracted using the crop mask data to focus only on crop of interest then average NDVI value for every year was computed. The calculated value is in raster value, which ranges from 0 to 255 and needed to be changed to NDVI value. Thus, the formula, \( \text{emodis NDVI} = \frac{\text{modis NDVI} - 100}{100} \) (Gidey et al., 2018), was run and also the result were ready for correlation with sorghum yield (Table 4). CHIRPS time series data of Decadal image was also composited at monthly level using MVC and were extracted with crop mask data and yearly average was computed from the extracted results for further analysis (Table 4). The WRSI model is a ratio of seasonal actual crop evapotranspiration (ETA) to the seasonal crop water requirement, the same as the potential crop evapotranspiration (PETc). Here, sorghum crop coefficient from LEAP software was adopted for the phonological from planting to flowering (initial 0.3, vegetative1.15, flowering1.15, Ripening0.55) (Figure 5).
Table 4 showing observed yield and independent variables.

| NO | Year (meher season) | 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   |

Multiple Linear Regression Analysis.

To run Multiple Linear Regression we use the data of Table 4. There were some assumptions using in this statistics: - (a) the basic assumption of the regression analysis approach is that sufficiently long and consistent time series of both remote sensing data and agricultural statistics are available. The latter are normally aggregated at the level of national/sub national administrative units, from which average NDVI values be extracted (b) The criterion variable was assumed to be a random variable (c) There would be statistical relationship (estimating the average value) rather than functional relationship (calculating an exact value) (d) Multiple linear regressions assume the relationship between the dependent and each independent variable to be linear. The linearity assumption can be tested with scatter plots (Osborne & Waters, 2002). Multiple regression analysis provides a predictive equation:

\[ Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n + \varepsilon \]

Where, \( \beta_0 \) = constant
\( \beta_1, \beta_2, \ldots, \beta_n \) = beta coefficient or standardized partial regression coefficients (reflecting the relative impact on the criterion variable)
\( x_1, x_2, x_n \) = scores on different predictors

The \( \beta \)'s are the regression coefficients, representing the amount the dependent variable \( y \) changes when the corresponding independent changes 1 unit. The \( \beta_0 \) is the constant, where the regression line intercepts the y axis, representing the amount the dependent \( y \) will be when all the independent variables are 0. The standardized version of the \( \beta \) coefficients is the beta weights, and the ratio of the beta coefficients is the ratio of the relative predictive power of the independent variables (Linear regression analysis, Yan and Su, 2009). The developed model predicts the average value of one variable (\( Y \)) from the value of another variable (\( X \)). The \( X \) variable is also called a predictor. Generally, this model is called a regression model.

Results And Discussions

Developing Multiple Linear Regression Model Equation for Maize Yield Forecasting in the Study Area.

Normalized Difference Vegetation Index Average, which is a result of monthly maximum value composite (MVC) averages of NDVI from the planting date to the end of the crop cycle gives a correlation coefficient of 0.84 with significant P value of 0.002 at 95 % confidence level and highly correlated independent variable was rainfall with a correlation coefficient of 0.89 with significant P value of 0.0001 at 95% confidence level. While others like ETA, which has a correlation value of 0.024 with significant P value of 0.942 at 95% confidence level, ETA total which has a correlation value of 0.22 with significant P value of 0.537 at 95% confidence level and WRSI (\( r =0.258 \)) with a P value of 0.472, which is beyond the acceptable range at 95 % of confidence level were rejected from the model development. Hence the two most correlated variables (NDVIa and rainfall) with the dependent variable (yield) are selected to create a multiple linear regression model.

As many studies on crop forecast states that linear regression modeling is the most common method to produce yield predictions by using remote sensing derived indicators together with bio climatic information. Maize yield data and data derived from the different indices were prepared for multiple linear regression analysis. The Statistical Package for Social Science (SPSS) software was used to build a multiple linear regression model using the two most correlated variables. As a result of all the above processes, the model highly correlated variables (NDVIa and Rainfall) were used to develop a model. This model was validated based on its Coefficient of determination (R2), root mean square error (RMSE) and coefficient of variation (CV) as shows in (Figure 7). When we see the overall fit of the model by examining the plot of the actual yield per hectare against the predicted yield per hectare, it reveals that, most points
lie fairly close to the 45° line (exact prediction line). The R square value of the model is 0.89; R square adjusted is 0.88 with root mean square error of 1.54 quintal per hectare. The P value of the model is 0.0001 at 95% confidence level and. By observing this P value, it is unclear which independent variable is the very good predictor and which is poor. The analysis of variance as shown in Table 5 state that maize yield forecast model has an observed significance probability (Prob>F) of 0.0001, which is significant at 0.05 level. Since the p<0.0001, we conclude that Yield is related to NDVIa and/or CHIRPS. From Table 6, the Variance Inflation Factor (VIF) of NDVIa and CHIRPS is 1.992. This shows that there is no multicollinearity problem between these two variables since VIF is less than 10. Therefore, for this research NDVIa and CHIRPS rainfall are selected for the model development. Table 7 shows parameter estimates of the model which reveals that NDVIa and CHIRPS rainfall. Therefore, from the result of Table 7 the values of the intercept (constant term), the estimate of NDVIa and CHIRPS are -20.375, 28.360 and 0.316 respectively. Within the model CHIRPS, NDVI and Intercept becomes strongly significant. A unit change in NDVIa and rainfall brings 28.360- and 0.316-unit times' changes in yield respectively. Therefore, the multi linear regression model equation of Maize yield forecast is given by:

Predicated Maize yield (qt/ha) = -20.375+ (28.360 x NDVIa) + (0.316 x CHIRPS)

According to the Central Statistical Agency Agricultural Production report coefficient of variation of barley Yield is 17.7 % and it in the acceptable range of validation value:

| Table 5 | Maize yield forecast model variance analysis (ANOVA) |
|---------|-----------------------------------------------------|
| Model   | Sum of Squares | Df | Mean Square | F     | Sig  |
| Regression | 156.490   | 1  | 156.490     | 65.643| 0.000|
| Residual   | 19.072    | 8  | 2.384       |       |      |
| Total      | 175.561   | 9  |             |       |      |

Degree of freedom (df): - is the number of values in the final calculation of a statistic (estimated value) that are free to vary.

Degree of freedom (df): - is the number of values in the final calculation of a statistic (estimated value) that are free to vary.

| Table 6 | Variance Inflation Factor (VIF) between NDVIa and CHIRPS |
|---------|----------------------------------------------------------|
| Variance inflation factor (VIF)            |
| Constant   | 0                                      |
| NDVIa      | 1.992                                   |
| CHIRPS     | 1.992                                   |

| Table 7 | Parameter estimates for the maize forecast model. |
|---------|----------------------------------------------------|
| Model   | Unstandardized Coefficients | Standardized Coefficients | t     | Sig. |
|         | B      | Std. Error | Beta     |       |      |
| 1       |        |            |          |       |      |
| Constant | -20.375| 7.696      | -2.647   | 0.033 |
| CHIRPS   | 0.316  | 0.093      | 0.599    | 3.407 | 0.048 |
| NDVIa    | 28.360 | 11.851     | 0.4211   | 2.393 | 0.011 |

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

When comparing the subjectivity of the conventional and remote sensing yield forecasts, the remote sensing approach wins. The forecast data, which is a result of the conventional approach, has a coefficient of variation of 17.7%, according to the CSA report, and it is a subjective approach. However, the remote sensing-based model predicts 16.7% with a high level of confidence (95%) and a large probability value. Furthermore, because September is the flowering stage of the maize crop, the forecast result of the remote sensing enabled methodology may be delivered in early October, whereas the traditional method data release calendar is generally in December and includes all cereal crops. Even though we did not consider all cereals that CSA has covered both in my research, this suggests that the timeliness issue may be addressed by applying the remote sensing aided strategy in a better way than the conventional approach. Another advantage of the remote sensing-based approach is that it gives location information, as the forecast can be checked by getting GPS readings and navigating to the places after it is created. As a result, this method allows for a precise indication of which areas have a high and low yield in a tangible manner, whereas traditional methods fail miserably. As a result, it is obvious that maize yield forecast utilizing remote sensing and GIS increases the data quality and timeliness while also reducing subjectivity. This research and prior related studies have shown that a remote sensing enabled approach can show places (lower administrative areas) where there is comparatively high, medium, and low production, making intervention very easy for decision makers. The following (Figure 8) shows a comparison of traditional yield estimations against the Remote Sensing assistance technique.

Testing the Model for Predicting maize Yields for the Year 2018 in the Study Area.
Based on the developed model, the 2018 maize crop forecast was made. Accordingly, highest maize yield for 2018 is expected to be 25 q∙ha⁻¹ and lowest 15 q∙ha⁻¹ with a mean of 20q∙ha⁻¹. The prediction also indicates that maize yield in 6.1% of the study area will be 10-15q∙ha⁻¹ and in 50.3% of the area to be 15-19q∙ha⁻¹ while the rest of 43.6% area is likely to yield 20-25 q∙ha⁻¹ (Table 4.4). Spatial distribution of the production levels in kaffa zone reveal that certain pockets of North-western, North-Eastern, northern and eastern part of the study area such as Gesha, Sayilem, Gimbo, Gewata and Menjwo district are most productive with 20 - 25 q∙ha⁻¹ of yield while the, western, south-eastern and central part of the Zone, Bita, Cheta, Talo and Bonga town zuria weredas are intermediately productive with 15-19 q∙ha⁻¹ of output. The rest of the study area also hosts least productive pockets giving only 10-15 q∙ha⁻¹ of grains. Hence, North-western, North-Eastern, northern and eastern part of the zone was more productive than other part of the study area (Figure 9).

### Table 8
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                  |

### Discussion

In terms of relevant study, despite the fact that the crop is distinct, we have attempted to compare our findings to those of other studies. According to Abiy (2014), among all the Argo climatic and remote sensing variables, NDVia and rainfall parameters were retained as significant variables for field level yield prediction in the multiple linear regression model, explaining 88 percent of the yield variability, implying that both rainfall and NDVia are the best parameters for yield prediction, which is consistent with the research findings. Meanwhile, this study discovered that NDVia and rainfall are preserved 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 relevant components for building a multiple linear regression model were evapotranspiration total and NDVic. The author discovered that ETa total had explained 83 percent of the yield variance (RMSE = 0.333 tha⁻¹ and CV = 21 percent) and that NDVic had explained 87 percent of the yield variance (RMSE = 0.333 tha⁻¹ and CV = 21 percent), demonstrating that spectro-agro meteorological models are possible even for fragmented agricultural lands, as in this case. However, due to different geo-climatic conditions, evapotranspiration total was not suitable for inclusion in the model (due to its minute correlation coefficient paired with insignificant p value link with maize yield).

The prediction power of the model in this research was high (Root mean square error =1.54 and R² =89%). When this result is compared to Abiy’s 2014 (Root mean square error =1.41 and R² =88 percent) and Akililu’s 2015 (Root mean square error =0.99 and R² =93 percent) south Tigray zone maize yield forecast and east Arsi zone wheat yield forecast, it reveals that the magnitude of the results is nearly the same. In terms of strongly correlated independent variables with yield, rainfall (r=0.89) is the most highly correlated, followed by Normalized Difference Vegetation Index Average (NDVia) (r=0.84). However, in Akililu, the 2015 Normalized Difference Vegetation Index Average (NDVia) and rainfall (r=0.89) are substantially associated (r=0.96). This demonstrates that yield prediction parameters fluctuate from one agro ecological zone to the next, implying that our model takes into account distinct elements in different correlation results.

Another consistency in terms of components or characteristics generated from Remote Sensing data was that in this study, the Water Requirement Satisfaction Index (WRSI) and Actual Evapotranspiration (Eta) were not related to yield, and the same was true for Abiy’s (2014) and Akililu’s (2014) studies (2015). In this study work, NDVI and rainfall are selected for the final model based on the Statistics results, similar to Abiy’s paper, however rainfall is omitted from the model in Akililu’s(2015) paper based on the Variance Inflation Factor (VIF) result. The findings of this study, which is the third crop analyzed after Abiy’s (2014) maize crop yield forecast research and Akililu’s(2015) wheat crop yield forecast research, show that Agro metrological factors have a definite potential for maize yield forecasting in the kaffa zone.

### Declarations

#### Acknowledgements

We are considerably grateful, to Bonga University for provision of monetary and Materials support to complete this study .We also the National Meteorology Agency (NMA), Ethiopian Geospatial Information Agency (EGIA), and CSA for providing the specified secondary data. Our appreciation additionally goes to Mr.Leta Bekele (from National Meteorology Agency of Ethiopia) for his assist in the analysis of the facts for their unreserved and tireless support, sharing of information and valuable recommendation in each element all through this study.

#### Authors’ contributions

Dereje Biru and Jemal Tefera conceived and designed the proven work of the method section. Author Participate in Analysis, Verification and Write Article .The author read and approved the final manuscript.

#### Funding

This research work did not get any grant from institutions, agencies, or individuals.

#### Availability of data and materials

...
Data and materials, generated and analysed are available in this research work.

**Ethics approval and consent to participate**

Not applicable.

**Consent for publication**

Not applicable.

**Competing interest**

On behalf of all authors, the corresponding author states that there is no conflict of interest among all the authors in this research work.

**Author details**

1. Bonga University, College of Social Sciences, Department of Geography and Environmental Studies, P.O. Box 334, Bonga, Ethiopia  derejebiru23@gmail.com

2. Bonga University, College of Social Sciences, Department of Geography and Environmental Studies, P.O. Box 334, Bonga, Ethiopia abukewserjemal@gmail.com

**References**

1. Abiy Wgderes (2014) Maize Yield Forcasting in South Tigray, M.Sc. Thesis Addis Ababa University. Addis Ababa, Ethiopia
2. Akillilu Fikere (2015) Wheat yield forecasting in Arsi, oromiya, M.SC.Thesis Addis Ababa University.AddisaAbaba, Ethiopia
3. Congalton RG, Green K (2008) Assessing the accuracy of remotely sensed data:. principles and practices.CRC press
4. Fenstermaker L (1991, October) A proposed approach for national to global scale error assessments.In Proceedings GIS/LIS (Vol. 91, No. 1, pp. 293-300)
5. 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 Systems and Environment 2(2):281–292
6. Gorfu D, Ahmed E (2012) Crops and agro-ecological zones of Ethiopia. Ethiopian Institute of Agricultural Research
7. 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)
8. Rajasankar O, Toruoya F, Valentina S (2016) Remote sensing Technique in crop mointring facilitates food security in Estonia. Tartu University, Estonia
9. 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
10. Shieh G (2010) On the misconception of multicollinearity in detection of moderating effects: Multicollinearity is not always detrimental. Multivar Behav Res 45(3):483–507
11. Yan G, Mas JF, Maathuis BHP, 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
12. Yan X, Su X (2009) Linear regression analysis: theory and computing. World Scientific

**Figures**
Figure 1
Location map of Kaffa Zone.

Figure 2
Maps of Land use/land cove of the study area.
Figure 3

Random points generated for accuracy
Figure 4

Crop mask data for Maize.

Figure 5

Crop coefficient of Maize in different stage (Planting-Flowering). (Source: LEAP software)
Figure 6

Methodological flow chart.

Figure 7

Maize yield quintal/hectare (Actual) vs. Maize yield quintal/hectare (Predicted)

R = 0.94, R² = 0.80, p-value = 0.0001, RMSE = 1.54
Comparison between the maize yields estimated by the agro meteorological Model and the observed yields for the study area.

**Figure 8**
Comparison between maize yield (quintal/ha) Estimated By the model and the observed yield.

**Figure 9**
Maize yield forecast map of 2018.