Article

Integrating Climate Data and Remote Sensing for Maize and Wheat Yield Modelling in Ethiopia's Key Agricultural Region

Asfaw Kebede Kassa [1,2], Hongwei Zeng [1], Bingfang Wu [1,*], Miao Zhang [1], Kibebew Kibret Tsehai [3], Xingli Qin [1] and Tesfay G. Gebremicael [4]

1 State Key Laboratory of Remote Sensing Science, Aerospace Information Research Institute, Chinese Academy of Sciences, No. 1, Beichen West Road, Chaoyang, Beijing 100101, China; asfaw.kebede@haramaya.edu.et (A.K.K.); zenghw@aircas.ac.cn (H.Z.); zhangmiao@aircas.ac.cn (M.Z.); qinxl@aircas.ac.cn (X.Q.)
2 School of Water Resources and Environmental Engineering, Haramaya University, Dire Dawa P.O. Box 138, Ethiopia
3 College of Agriculture and Environmental Sciences, Haramaya University, Dire Dawa P.O. Box 138, Ethiopia; kibebew.kibret@haramaya.edu.et
4 Bureau of Water Supply Planning, St. Johns River Water Management District P.O. Box 1429, Palatka, FL 32178-1429, USA
* Correspondence: wubf@aircas.ac.cn

Abstract: Traditional methods for crop data collection are labor-intensive, inefficient and, more costly compared to remote sensing (RS) techniques. This study aims to identify key climatic variables influencing maize and wheat yields and develop predictive models while also evaluating the performance of the CropWatch cloud yield prediction model (CW_YPM) in major agricultural regions of Ethiopia. Climate data from 54 meteorological stations spanning 2000–2021 were analyzed. RS data, including NDVI from MODIS at 250 m resolution, agroecological zones, and observed crop yield data, were utilized for model prediction and validation. Correlation analysis and a stepwise modeling approach with multiple regression models were applied. The results revealed regional variations in the effects of climatic parameters on yields, with vapor pressure deficits showing negative correlations and rainfall exhibiting positive correlations. Non-linear models generally outperformed linear models in yield prediction—using both climate-only (CO) and combined climate-NDVI data. The best CO model for maize in the Horo Guduru area achieved an RMSE of 0.392 tons/ha, an R^2 of 0.94, and an index of agreement (d) of 0.984. Incorporating NDVI improved accuracy, with the best maize model in the Illu Ababor area achieving an RMSE of 0.477 tons/ha, an R^2 of 0.91, and d of 0.976. CW_YPM also performed effectively across the study area. This research highlights the value of integrating critical climatic variables with the NDVI to enhance crop yield forecasting in Ethiopia, thereby-supporting agricultural planning and food security initiatives.

Keywords: crop yield; NDVI; non-linear regression; Oromia region; CropWatch

1. Introduction

Accurate forecasting of agricultural yields for staple cereal crops, such as maize and wheat, is vital for ensuring food security, stabilizing markets, and managing supply chains. Robust forecasting models are particularly important for these crops, which are central to the diets of a significant portion of the world's population. However, developing such models presents challenges in many developing countries due to limited data availability. Conventional methods, such as on-site visits and manual reporting, remain the primary

means of monitoring crop yields in these regions [1]. While reliable, these traditional methods are time-consuming, resource-intensive, and often lack the scalability required to address modern agricultural challenges.

Crop yield prediction is inherently complex, as yield formation depends on a wide range of factors, including soil conditions, meteorology, environmental influences, and crop genetics. Therefore, reliable crop yield prediction is a daunting task in agriculture [2–4]. Recent studies recommend that incorporating diverse input parameters, such as vegetation indices, soil moisture, leaf area index (LAI), climate variables, and the Normalized Difference Vegetation Index (NDVI), into crop yield prediction models can enhance decision-making for policymakers and improve crop management strategies [5–7].

There are several methods for predicting crop yields. Bingfang et al. [5] summarize them into four main categories: regression, biomass and harvest index, crop growth models, and data-driven/machine learning methods. Traditionally, the linear regression model has been widely applied to establish relationships between crop yield and predictors, such as meteorological variables (e.g., rainfall and temperature) and vegetation indices [8]. While useful, linear regression often suffers from weak generalization capabilities. In contrast, data-driven models, combining machine learning techniques with multi-dimensional features, have shown promise for crops like maize and soybean in developed regions, such as the Midwestern United States [9,10]. However, these methods depend heavily on extensive datasets and advanced computational infrastructure, which are difficult to access in developing countries. Hybrid approaches, such as the CropWatch yield prediction model developed in China, integrate multiple indicators across global, national, and regional scales, offering an alternative framework for yield estimation [11–13]. Ultimately, the choice of crop yield prediction strategy must consider data availability, model performance, and the intended purpose.

Ethiopia, the second-largest wheat producer in Africa, cereal crops dominate agricultural production, covering 81% of the land used for grain farming and contributing 88% to the total grain production [14,15]. In 2022/2023 production year, Ethiopia produced 5.5 million metric tons of wheat, accounting for 21.7% of Africa's wheat production and 18.3% of the continent's harvested wheat area [16]. Despite these contributions, Ethiopia faces persistent food insecurity, exacerbated by prolonged droughts, climate change, and a predominantly rain-fed agricultural system [6,17]. Wetter years generally correlate with higher food production, while dry years result in lower yield production [18–21]. Compounding these challenges are issues such as population pressure, inadequate disease control, outdated farming technologies, and significant pre and post-harvest losses [22]. Addressing these vulnerabilities requires early and accurate monitoring of climate variability and crop production to mitigate food insecurity risks.

Many attempts have been made to predict crop yield in Ethiopia. For example, Zinna and Suryabhagavan [23] utilized time remote sensing (RS) data from SPOT VEGETATION and other data to forecast maize crop yields in South Tigray, while Reda [24] applied RS and GIS techniques to predict wheat yields in the Arsi zone. Debalke and Abebe [1] developed a linear regression model incorporating eMODIS NDVI data to estimate maize yields in the Kafa zone. Awetahegn et al. [6] demonstrated the efficiency of integrating LAI data into the MOFOST crop model for large-scale wheat yield estimation in Ethiopia. While these studies provide valuable insights, they often lack comprehensive analysis of key climatic variables across Ethiopia's diverse agricultural zones or fail to incorporate these variables into predictive models effectively.

Moreover, many studies recommend testing polynomial and non-linear regression models and enhancing RS and GIS approaches to identify additional factors contributing to yield prediction variability. Milkessa and Amsalu [14] highlighted the positive impact

of precipitation on cereal crop production and the adverse effects of rising temperatures. However, there remains a significant gap in studies testing multiple non-linear models, leveraging long-term climate data, or spatially addressing Ethiopia's key agricultural production areas. Generalizability also remains a challenge, as many models are constrained by region-specific and data-dependent limitations.

RS presents a powerful alternative, offering timely, precise, and scalable data for crop yield estimation. Despite its potential, RS-based approaches have been underutilized in the Ethiopian crop yield prediction efforts. Addressing these gaps is essential for improving the accuracy and applicability of yield forecasting models for Ethiopia's agricultural systems.

This study focuses on maize and wheat in Ethiopia to evaluate and compare the performance of regression-based approaches using stepwise multiple regression and RS-based CropWatch cloud platform models for crop yield prediction, which use the Carnegie–Ames–Stanford Approach (CASA) to calculate the net primary productivity.

It aims to identify the most effective method for predicting maize and wheat yields in key agricultural zones of Ethiopia. The primary objectives are as follows: (1) to identify key climatic variables that significantly influence maize and wheat yields and develop predictive models for these crops in selected Ethiopian regions, leveraging climate data and RS inputs, either independently or in combination, and (2) to evaluate the performance of RS-based crop yield modules provided by the CropWatch cloud platform under Ethiopia's condition through customizing its application in the area.

By developing robust predictive models utilizing the CropWatch yield prediction model, which integrates an extensive range of RS indicators at global, national, and regional scales, this research aspires to support data-driven agricultural decision-making. The ultimate goal is to bolster the resilience and sustainability of Ethiopia's agricultural systems.

2. Materials and Methods

2.1. Study Area

Ethiopia, covering a total area of 1,104,300 km^2, is divided into 72 administrative zones (Figure 1). It is located in the Horn of Africa and spans from 32°42′ E to 48°12′ E longitude and 3°30′ N to 14°50′ N latitude. The country experiences diverse climate, with mean annual rainfall ranging from 550 mm in the northern and eastern regions to over 2000 mm in the western and southwestern areas. The mean annual temperature ranges from 15 to 20 °C in high-altitude regions, while it varies from 25 to 30 °C in the lowlands [6].
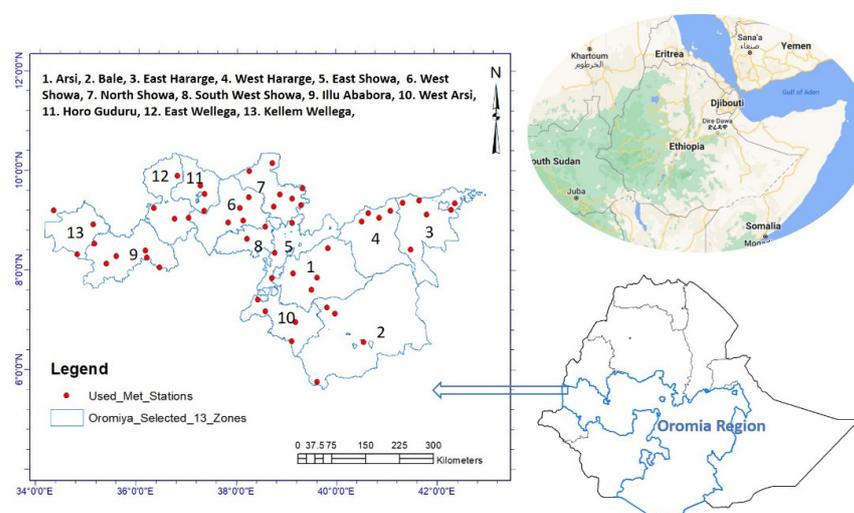
Figure 1. Location map of the study area, highlighting 13 selected administrative zones known for wheat and maize cultivation, along with the distribution of meteorological stations used in the study.

This research focused on 13 zones within the Oromia region, a key area for maize and wheat production in Ethiopia. These zones are situated at 5°33′59″ and 10°21′34″ N latitude and 34°10′48″ to 43°04′12″ E longitude, encompassing a total area of 203,174.27 km² (Figure 1). Oromia is Ethiopia's most densely populated regional state and a critical agricultural hub, contributing to 50% of the country's total production of major food crops [25]. Out of the 5.86 million hectares allocated to grain crops in Oromia, maize cultivation accounts for 1.2 million hectares, representing 21% of the total area. Remarkably, 72% of smallholder farmers in the region engage in maize cultivation [25].

2.2. Data Used

2.2.1. Climate Variables

In this study, we analyzed climate data that included seasonal minimum (Tm), maximum (Tx), and mean (Tmean) temperatures in degrees Celsius, areal rainfall in millimeters (ArealRF), and both minimum (VPDm) and maximum (VPDx) vapor pressure deficit in kilopascals (KPa). The data were sourced from the National Meteorological Agency of Ethiopia, with observations collected from 54 meteorological stations spanning from 2000 to 2021. To accurately estimate areal rainfall distribution across the study area, we adopted the Theisen polygon approach. Missing rainfall and temperature data were supplemented using the PDIR-Now (Dynamic Infrared Rain rate near real-time) system, available at CHRS Data, and TAMSAT (Tropical Applications of Meteorology using SATellite) data, accessible at TAMSAT, both at 4 km resolution. Vapor pressure deficit was calculated using an equation established by Tetens [26] (as shown in Equations (1) to (3)). We aggregated these climate variables over the growing seasons, which run from late May to early November for wheat and from late April to September for maize, across each administrative zone within the study area. Figure 2 illustrates the historical climate variables for the primary maize and wheat producing regions, covering the period from 2000 to 2021.

$$e_s = 0.6108 \times exp\left(\frac{17.27 \times T}{T + 237.3}\right) \tag{1}$$

$$e_a = \left(\frac{RH}{100}\right) \times e_s \tag{2}$$

$$Vapor\ Pressure\ Deficit = e_a - e_s \tag{3}$$

where e_s = saturation vapor pressure (Kpa), e_a = actual vapor pressure, T = temperature in °C. If *RH* is in percentage, e_s = 60%; otherwise, $e_a = RH * e_s$.
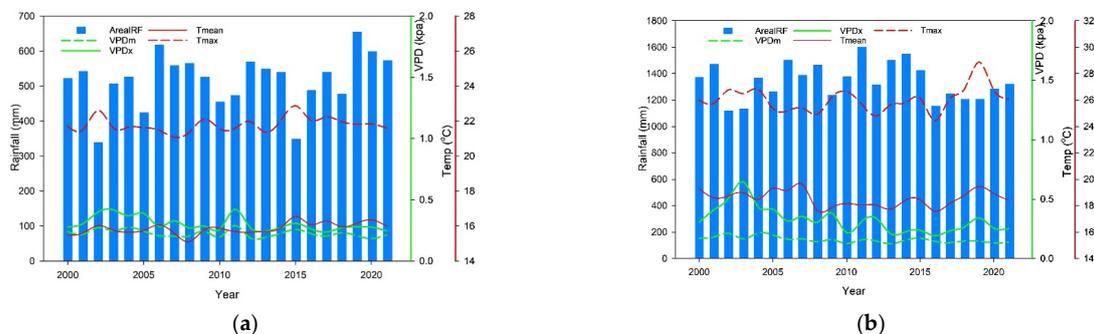
Figure 2. Historical (2000–2021) climate variables (seasonal areal rainfall, average temperature (Tmean), and vapor pressure deficit) of selected administrative zones: (**a**) Arsi from wheat growing area (June to October) and (**b**) Illu Ababora from maize growing area (May to September).

2.2.2. Satellite Imaging Data

NDVI data, calculated using near-infrared (NIR) and red bands from Landsat sensors multispectral images, served as reliable indicator of crop growth conditions (Equation (4)). For crop yield estimation, NDVI data from intermediate spatial resolution sensors, such as the Moderate Resolution Imaging Spectroradiometer (MODIS), have been widely utilized. Key studies, including those by Becker-Reshef, Vermote et al. [27], Mkhabela et al. [28], Vintrou et al. [29], Kouadio et al. [30], and Johnson [31], highlight the significance of this technology in advancing agricultural monitoring and yield forecasting. In this research, MODIS NDVI dataset with a 250 m spatial resolution (available at https://lpdaac.usgs.gov/products/mod13q1v061/ (accessed on 20 March 2024)) was utilized to align with meteorological data. This dataset provided monthly NDVI values, from which maximum eMODIS NDVI values for maize (May to September) and wheat (June to October) were extracted for the years 2000 to 2021. The Google Earth Engine (GEE) platform facilitated the processing. Cropping areas were masked using agroecological zones and the European Space Agency (ESA) WorldCereal 10m 2021 model (https://developers.google.com/earth-engine/datasets/catalog/MODIS_061_MOD13Q1, (accessed on 20 March 2024)) as described in the subsequent section.

$$NDVI = (NIR - Red)/(NIR + Red) \tag{4}$$

where near-infrared (*NIR*) bands typically range from 700 nm to 2500 nm, and red bands are in the visible spectrum, typically ranging from 620 nm to 750 nm.

For calibrating the CropWatch cloud yield prediction model, one of the required inputs is to determine the starting and end of the cropping season. For this study area, Terra and Aqua used Moderate Resolution Imaging Spectroradiometer (MODIS) Land Cover Dynamics (MCD12Q2) Version 6.1 data product, which provides global land surface phenology metrics (https://doi.org/10.5067/MODIS/MCD12Q2.061, (accessed on 16 May 2024)) [32]. The dataset provides global land surface phenology metrics at yearly intervals, capturing up to two detected growing cycles per year with a spatial resolution of 500 m, dating back to 1970. Each asset contains layers detailing various vegetation metrics, including the total number of growing cycles per year, onset of greenness, Greenup midpoint, maturity, peak greenness, senescence, green-down midpoint, dormancy, and other related metrics over a vegetation cycle. Additionally, the dataset includes comprehensive quality assessments and phenology metric-specific quality indicators, which can be analyzed for designated areas of interest. In this study, the areas of interest were defined, and the mid-Greenup and senescence periods for each growing season were subsequently extracted using Google Earth Engine (GEE) platform.

2.2.3. Crop Masks Data

To mask the cropped area, the agroecological zones (AEZs) were developed using the elevation map of the study area. Gorfu and Ahmed [33] described that maize in Ethiopia is generally grown between elevations of 1500 and 2200m, and wheat between 1500 and 3000 m. Equations (5) and (6) were used in ArcMap, employing the Digital Elevation Model (DEM) to develop these AEZs. Additionally, the cropping areas were further defined using the World Cereal 2021 map from the ESA, which provides global-scale annual and seasonal crop maps at a 10 m resolution (https://esa-worldcereal.org/en, (accessed on 20 March 2024). An intersection of the AEZs with the World Cereal 2021 map was performed to create a shapefile outlining the areas covered by crops in the study area.

$$\text{Maize AEZ based on elevation} = (\text{Value} \geq 1500) \text{ AND } (\text{Value} \leq 2200) \tag{5}$$

$$\text{Wheat AEZ based on elevation} = (\text{Value} \geq 1500) \text{ AND } (\text{Value} \leq 3000) \tag{6}$$

2.2.4. Crop Yield and Harvest Index (HI) Data

The crop yield data collected from the Central Statistics Agency of Ethiopia (CSA) for the years 2000 to 2021 provide a comprehensive overview of cereal yields at the zonal (district) administrative level (Figure 3). These data are crucial for model calibration as emphasized by Rijks et al. [34], who highlighted the importance of utilizing historical crop yield records to develop accurate quantitative yield estimates. The CSA employs standard statistical data collection and analysis methods to produce these estimates, which are reported annually. In this study, the Oromia region has been selected as representative study area, given its status as Ethiopia's leading grain producing area, as indicated in the CSA's reports. The data presented in Figure 3a illustrate the variations in crop yields over the specified period, reflecting trends and fluctuations that can inform agricultural practices and policy decisions in the region. By focusing on such a significant agricultural zone, this study aims to enhance understanding of yield dynamics and contribute to improved agricultural productivity strategies in Ethiopia.
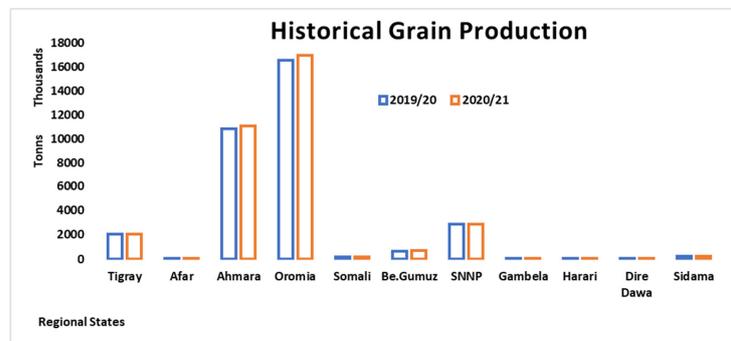
(**a**)
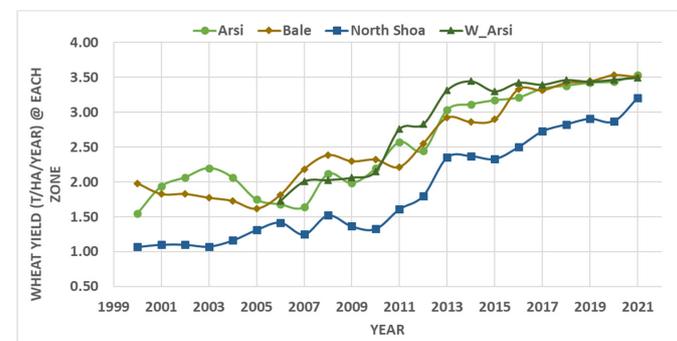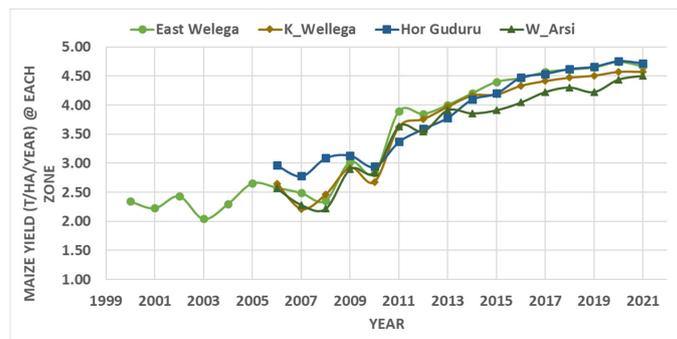
(**b**)

Figure 3. Historical grain yield: (**a**) total grain production in regional states, Ethiopia (2019/2020 and 2020/2021); (**b**) maize and wheat yield data at selected administrative zones in Oromia region (2000 to 2021), Ethiopia [15].

The harvest index (HI) for both wheat and maize in Ethiopia demonstrates notable variability influenced by the specific crop varieties and agricultural practices employed. Research has shown that the HI for maize ranges significantly, with values at farmers' fields falling between 25% and 37%, while, at the research fields, it is observed to range from 31% to 45% [35,36] This variability underscores the impact of management practices and the choice of crop variety on yield outcomes. In a similar vein, the HI for wheat varies at farmers' fields, showing a range from 13% to 25%. Given this variability, the HI is utilized as a critical parameter in the calibration process of the CropWatch yield prediction model.

2.3. Methods of Analysis

2.3.1. Method for Predicting Crop Yield

This study was undertaken following two principal analytical steps. First, Pearson's correlation analysis was carried out to explore the associations among climate variables, NDVI, and crop yield. Next, a stepwise modeling approach was used, applying multiple regression models that include linear, non-linear, and polynomial modes, to determine their effectiveness in accurately predicting crop yield. The second approach involved using crop yield prediction model that was calibrated based on the crop phenology and harvest index for dominant maize and wheat growing areas of the study region. The general approaches followed in this research are depicted in Figure 4.
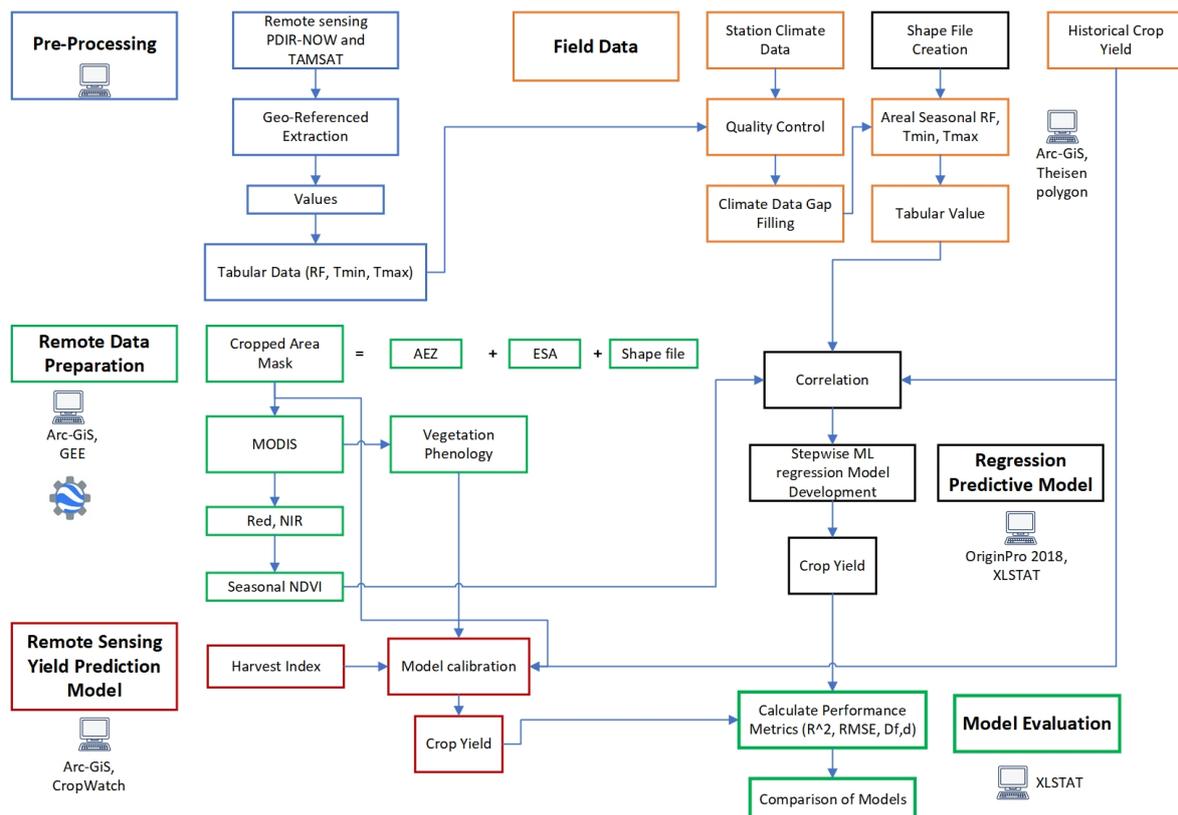
Figure 4. General methodology flow chart. (RF = rainfall).

2.3.2. Correlation Analysis

In this study, we used a correlation inspection to assess the collinearity among all predictors and crop yield. Our goal was to find the most influential variables for constructing robust prediction models. We assessed linear dependencies between pairs of variables using Pearson's correlation coefficient (C). A correlation value near 1 indicates a strong interdependence, suggesting that changes in one variable are proportionally reflected in the

other. We scrutinized seven potential variables using both Pearson's correlation coefficient and principal component analysis (PCA), as described by Friendly [37]. PCA helped identify key variables by explaining variance through principal components and highlighting those variables that are highly correlated with each other and with crop yield. We used the software package OriginPro 2018 to perform both PCA and Pearson's correlation analysis. After identifying these variables, we evaluated them individually and collectively to develop effective regression models for yield prediction. The correlations between the predictor variables and the crop yield were ranked in descending order according to their correlation coefficients. This ranking system helped in systematically selecting variables for constructing multiple regression models. Starting with the variable that had the highest correlation, we included the next most significant climate variables one at a time in a stepwise manner to predict crop yields. This process aligns with the stepwise regression approach commonly used to select relevant climate variables. The general approaches followed in this research are depicted in Figure 4.

2.3.3. Regression Models

In the linear regression models, as described by James et al. [38], single or multiple predictors can be applied to predict yield, as indicated in Equation (7). Additionally, Equation (8) shows how the multiple non-linear regression model is used for yield prediction.

$$y_i = \sum_{i=1}^{n} \beta_i x_i \tag{7}$$

where y_i represents crop yield, β_i is the regression coefficient of variables (predictors) x_i, and n is the total number of variables used.

$$y_i = Pr_1 + Pr_2 X_1 + Pr_3 X_1^2 + Pr_4 X_2 + Pr_5 X_2^2 + \dots \tag{8}$$

where y_i represents crop yield, Pr_n are regression parameters, and X_n is variable (predictor).

Multiple linear regression and non-linear regression models were constructed using data from the period 2000 to 2021. Variables were selected based on their correlation with crop yield and each other, prioritized by their significance. These models were then evaluated throughout the growing season to assess their predictive accuracy. Furthermore, we tested all possible combinations of models using climate data alone, and combinations of climate data and NDVI, to determine the most effective approach.

2.3.4. CropWatch Crop Yield Prediction Model Overviews

A light-use efficiency model was employed to predict crop yield at the pixel level via the CropWatch cloud platform [13]. The crop yield (kg ha-1) can be described as follows (Equation (9)):

$$Y = \frac{NPP \times T \times p \times Hi}{1 - \omega} \times 10 \tag{9}$$

Here, NPP is net primary productivity; T is the conversion coefficient between plant carbon content and plant dry matter mass; P is the proportion of biomass in the aboveground part of the crop relative to the whole plant; ω is the moisture content coefficient of the crop during the storage period following harvest with 13~14%; and HI is the harvest index. In this equation, T, P, and ω were set to 2.34, 0.9, and 0.135, respectively. The Carnegie–Ames–Stanford Approach (CASA) model was used in the CropWatch cloud platform to calculate the NPP (gC m^{-2}). The CASA model is driven by meteorological data and RS data. The precipitation, minimum and maximum temperature of air at 2 m above the surface of land, and amount of solar radiation reaching the surface of the Earth of ERA5-Land were utilized in the model. The surface reflectance bands 1–7 and those

from the MOD09GA and MYD09GA, along with the NDVI, were employed in the model. Further details can be found in [39]. The HI has a value between 0 and 1, and it should be adjusted by the user according to statistical yield data or observed data. Similarly, vegetation phenology extracted from MODIS Land Cover Dynamics (MCD12Q2) Version 6.1 data product was used for calibrating the yield prediction model as cropped area mask (Figure 4).

2.3.5. Model Evaluation

We assessed the effectiveness of our developed regression models and the crop yield prediction from the remote sensing model using statistical metrics. First, we utilized the coefficient of determination (R^2) to gauge how well the models explain the variance in observed data (Equation (12)), effectively measuring how accurately the simulations predict actual crop yields. Additionally, we employed the root mean square error (RMSE, Equation (10)) to measure the precision of our predictions. We also used the modified index of agreement (*d*, Equation (11)) as proposed by Yang et al. [40]. This index, which ranges from 0 to 1, standardizes the measurement of model prediction errors. A value of 1 indicates a perfect match between predictions and observations, whereas a value of 0 signifies no agreement at all, according to Willmott [41].

$$RMSE = \left[n^{-1} \sum_{i=1}^{n} (y_i - x_i)^2 \right]^{0.5} \tag{10}$$

$$d = 1 - \frac{\sum (y_i - x_i)^2}{\sum (|y_i - \overline{x}| + |x_i - \overline{x}|)^2} \tag{11}$$

$$R^2 = 1 - \frac{SST}{SSE} \tag{12}$$

$$SST = \sum_{i=1}^{n} (y_i - \overline{y})^2$$

$$SSE = \sum_{i=1}^{n} (y_i - \overline{y_i})^2$$

where SST = total sum of squares, SSE = residual sum of squares, n is the number of observations, Y_i is model result/simulation, $\overline{y}$ is mean observed yield, $\overline{y_i}$ is mean of simulated yield, and X_i is the observed yield.

3. Results

3.1. Meteorological Variables

3.1.1. Variables That Significantly Influenced Yield Prediction

The analysis conducted to identify the most influential variable for each crop type revealed significant insights, as illustrated in Figure 5. This figure highlights the varying correlations between predictor variables and crop yields across different zones, suggesting that a single regression model may be inadequate for the entire study area. The diverse impacts of climate variables on crop yields necessitate a more tailored approach to modeling. To address this variability, this study incorporated the most significant climate variables in combination with the NDVI. This combination was evaluated for its performance against models that used only climate data.

(**a**)   (**b**)
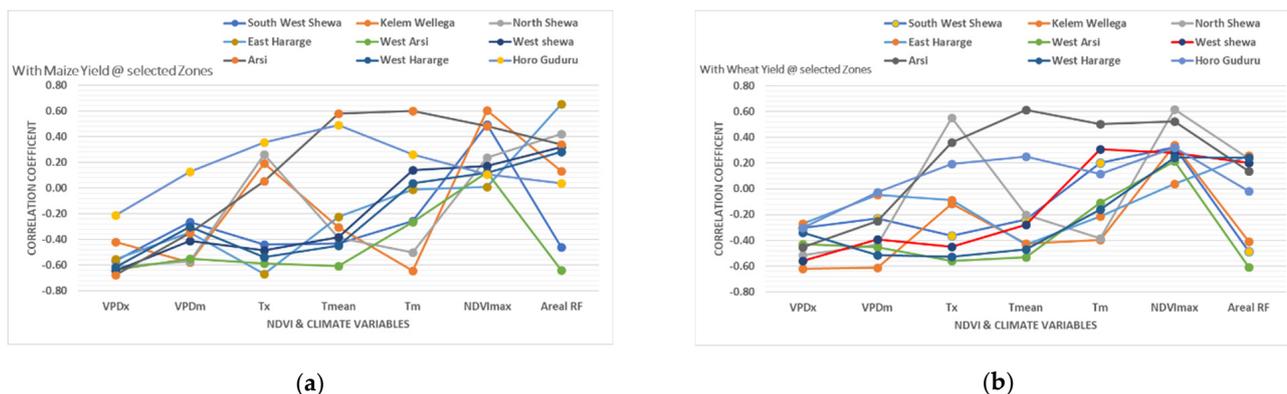
Figure 5. Climate variables and Normalized Difference Vegetation Index (NDVI) correlation analysis with (**a**) maize and (**b**) wheat yield at selected administrative zones in the study area.

The climate variables exhibited varying effects on crop yields across various zones within the study area. Specifically, the maximum VPD (VPDx) during the growing season was negatively correlated with both maize and wheat yields across all zones, with correlation coefficients ranging from 0.68 to 0.61, as shown in Figure 5. This indicates that higher maximum VPD levels are associated with lower crop yields, highlighting the detrimental impact of water stress during critical growth periods. Similarly, the minimum VPD (VPDm) also demonstrated a negative correlation with crop yields, although with varying strengths across different zones. In the Illu Ababora zone, which is known for its significant maize production, the C was notably high at 0.68. Conversely, in the East Wellega zone, the C was weaker, with a coefficient of 0.22. This variability suggests that the relationship between VPDm and crop yields may be influenced by local climatic conditions and agricultural practices, underscoring the need for region-specific strategies to mitigate the effects of vapor pressure deficits on crop production.

The analysis of areal rainfall and temperature effects on crop yields revealed significant trends across the study zones for both maize and wheat. Areal rainfall demonstrated a mostly positive correlation with crop yields in the majority of zones, with correlation coefficients ranging from 0.65 to 0.13. This positive relationship underscores the critical role that adequate rainfall plays in enhancing crop productivity. In contrast, the impact of temperature on yield displayed considerable variability. The maximum temperature (Tx) positively correlated with yields in six zones (C = 0.55 to 0.19), indicating that higher maximum temperatures may benefit crop growth in these areas. However, in seven zones, the correlation was negative, with coefficients ranging from 0.69 to 0.11, suggesting that excessive heat can be detrimental to crop yields. Minimum temperature (Tm) generally had a negative correlation with yields across most zones, with coefficients ranging from 0.65 to 0.11. However, exceptions were noted in the Arsi zone, where Tm had positive correlations of 0.6 and 0.5, and in the West Shewa zone with coefficients of 0.31 and 0.14. The mean temperature (Tmean) followed a similar pattern, exhibiting negative correlations in all zones except in Arsi (C = 0.61 and 0.58) and East Shewa (C = 0.31 and 0.37), where positive correlations were observed. These patterns highlight the complex interplay between temperature and crop yields.

The analysis of the maximum NDVI during the growing season reveals a generally positive correlation with maize and wheat yields across most agricultural zones. However, it is worth noting that the influence of the NDVI on yields appears to be significantly lower in specific regions. In East Hararge, the correlation coefficient for both maize and wheat yields is notably low at 0.04 and 0.01, respectively. Similarly, in Illu Ababora, the NDVI's impact is also minimal, with correlation values of 0.19 for maize and 0.01 for wheat. This

suggests that other factors may play a more critical role in determining crop yields in these areas, and further investigation may be necessary to understand the underlying causes.

Seven climate variables were analyzed and ranked from highest to lowest based on their correlation with crop yield, as determined by the correlation coefficient (C). In the Bale zone, the climate variables influencing crop yield were ranked accordingly. For maize yield, the rankings were as follows: NDVIx: (C = −0.59), VPDx: (C = −0.51), Tx: (C = −0.49), Areal RF: (C = 0.48), Tmean: (C = −0.48), VPDm: (C = −0.47) and Tm: (C = −0.39). For wheat yield, the variables were ranked as follows: Tx: (C = −0.69), Tmean: (C = −0.64), VPDx: (C = −0.53), Tm: (C = −0.45), Areal RF: (C = 0.39), VPDm: (C = −0.39), and NDVIx: (C = −0.06). A summary of the correlation rankings for other zones across the study area is presented in Figure 5. Notably, in most zones, VPDx and VPDm emerged as highly influential variables for crop yield, underscoring their critical role in agricultural performance.

3.1.2. Selecting Important Climate Variables

Effectively predicting crop yield typically requires four to six variables. However, in data-limited regions, a smaller set of climate variables can often yield reliable results. One significant challenge in expanding the number of climate variables in multilinear regression models is the risk of collinearity, which can complicate analysis and reduce model reliability. To address this, we employed several statistical criteria to enhance model performance. Specifically, our focus was on minimizing the MSE and RMSE, reducing the degree of freedom (DF) and the number of predictor variables, and preventing overfitting.

During the model selection process, we carefully analyzed residual patterns between observed and model-fitted crop yields to ensure no discernible trends were present, as such patterns could indicate potential model deficiencies. Figure 6 presents a residual plot for maize and wheat yields across all zonal administrations in the study area. This analysis was crucial in identifying the most effective predictor variables, typically ranging from four to seven per study area, based on the evaluation of the MSE, RMSE, and DF.

3.2. Regression-Based Yield Prediction Models
3.2.1. Models Using Climate Predictors

This study explored the relationship between crop yield and climate variables using both linear and non-linear multiple regression models. Notably, only the regression coefficients from the non-linear models exhibited significant correlations, prompting us to focus our analysis on these findings. The most effective model utilizing climate variables alone was identified in the Horo Guduru zone for maize, achieving an RMSE of 0.392 tons/ha, an R^2 of 0.94, and an index of agreement (d) of 0.984, reflecting a high degree of accuracy (see Table 1 for details). In contrast, the least effective model was observed in the West Hararge zone for wheat, with an RMSE of 0.562 tons/ha, an R^2 of 0.46, and a d value of 0.79. Despite this variation, the majority of the models based solely on climate predictors demonstrated above-average performance, underscoring the potential of climate data for crop yield modeling.

Table 1. The best selected multiple non-linear regression models constructed using climate variables or Normalized Difference Vegetation Index (NDVI) and climate variables for maize and wheat for a time scale of 22 years (2000–2021).

Crop	Study Area	Climate Variables/NDVI and Climate Variables	R^2	RMSE	MSE	DF	d	Regression Model Code *
Maize	Arsi	ArealRF, VPDm, VPDx, Tmean, Tm, NDVIx	0.79	0.478	0.228	9	0.937	Arsi_M
Wheat		VPDm, VPDx, Tmean, Tx, Tm, NDVIx	0.83	0.432	0.187	9	0.952	Arsi_W
Maize	Bale	ArealRF, VPDm, VPDx, Tmean, NDVIx	0.62	0.657	0.432	11	0.868	Bale_W
Wheat		ArealRF, VPDm, VPDx, Tmean, Tx, NDVIx	0.79	0.455	0.207	9	0.939	Bale_W
Maize	W_Shewa	ArealRF, VPDm, VPDx, Tmean, Tm, Tx	0.67	0.792	0.628	9	0.888	West Shewa_M
Wheat		ArealRF, VPDx, Tx, Tm, NDVIx	0.71	0.468	0.219	11	0.907	West Shewa_W
Maize	E_Shewa	ArealRF, VPDm, VPDx, Tmean, Tx	0.72	0.502	0.252	11	0.912	East Shewa_M
Wheat		ArealRF, VPDm, VPDx, Tmean, Tm, NDVIx	0.74	0.515	0.266	9	0.92	East Shewa_W
Maize	Ill Ababora	ArealRF, VPDm, VPDMx, Tmean, Tm, NDVIx	0.91	0.477	0.228	9	0.976	Illu Ababora_M
Wheat		ArealRF, VPDm, VPDMx, Tmean, Tm, Tx, NDVIx	0.81	0.589	0.347	7	0.945	Illu Ababora_W
Maize	W_ Arsi	ArealRF, VPDx, Tmean, Tm, NDVIx	0.89	0.440	0.193	5	0.971	West Arsi_M
Wheat		ArealRF, VPDm, VPDx, Tmean, Tx, NDVIx	0.92	0.410	0.168	3	0.98	West Arsi_W
Maize	N Shewa	ArealRF, VPDm, VPDMx, Tmean, Tm, NDVIx	0.83	0.459	0.211	9	0.952	North Shewa_M
Wheat		ArealRF, VPDMx, Tx, Tm, NDVIx	0.77	0.476	0.227	11	0.929	North Shewa_W
Maize	S W Shewa	ArealRF, VPDx, Tmean, Tx, NDVIx	0.74	0.746	0.557	8	0.92	South west Shewa_M
Wheat		ArealRF, VPDm, VPDx, Tmean, Tx, Tm, NDVIx	0.74	0.696	0.484	4	0.918	South west Shewa_W
Maize	E Hararge	ArealRF, VPDm, VPDx, Tmean, Tx, Tm	0.82	0.288	0.083	9	0.949	East Hararge_M
Wheat		ArealRF, VPDm, VPDMx, Tmean, Tm, Tx, NDVIx	0.52	0.615	0.378	7	0.813	East Hararge_W
Maize	W Hararge	ArealRF, VPDm, Tmean, Tx, NDVIx	0.76	0.329	0.109	11	0.926	West Hararge_M
Wheat		ArealRF, VPDm, Tmean, Tx, NDVIx	0.56	0.219	0.048	11	0.834	West Hararge_W
Maize	E Wellega	ArealRF, VPDm, VPDx, Tmean, Tx, Tm	0.77	0.742	0.551	9	0.929	East Wellega_M
Wheat		ArealRF, VPDm, VPDMx, Tmean, Tm, NDVIx	0.72	0.291	0.085	4	0.913	East Wellega_W

Table 1. *Cont.*

Crop	Study Area	Climate Variables/NDVI and Climate Variables	R²	RMSE	MSE	DF	d	Regression Model Code *
Maize	K_Wellega	VPDm, VPDx, Tm, NDVIx	0.89	0.411	0.169	7	0.969	Kellem Wellega_M
Wheat		ArealRF, VPDm, VPDx	0.93	0.154	0.024	4	0.982	Kellem Wellega_W
Maize	Horo Guduru	ArealRF, VPDm, VPDMx, NDVIx	0.82	0.446	0.199	7	0.948	Horo Guduru _M
Wheat		VPDMx, Tmean, Tx, NDVIx	0.79	0.408	0.167	7	0.939	Horo Guduru _W

* the detail regression model found as annex.

South West Shewa

Kelem Wellega

North Shewa

East Wellega

Bale

East Hararge
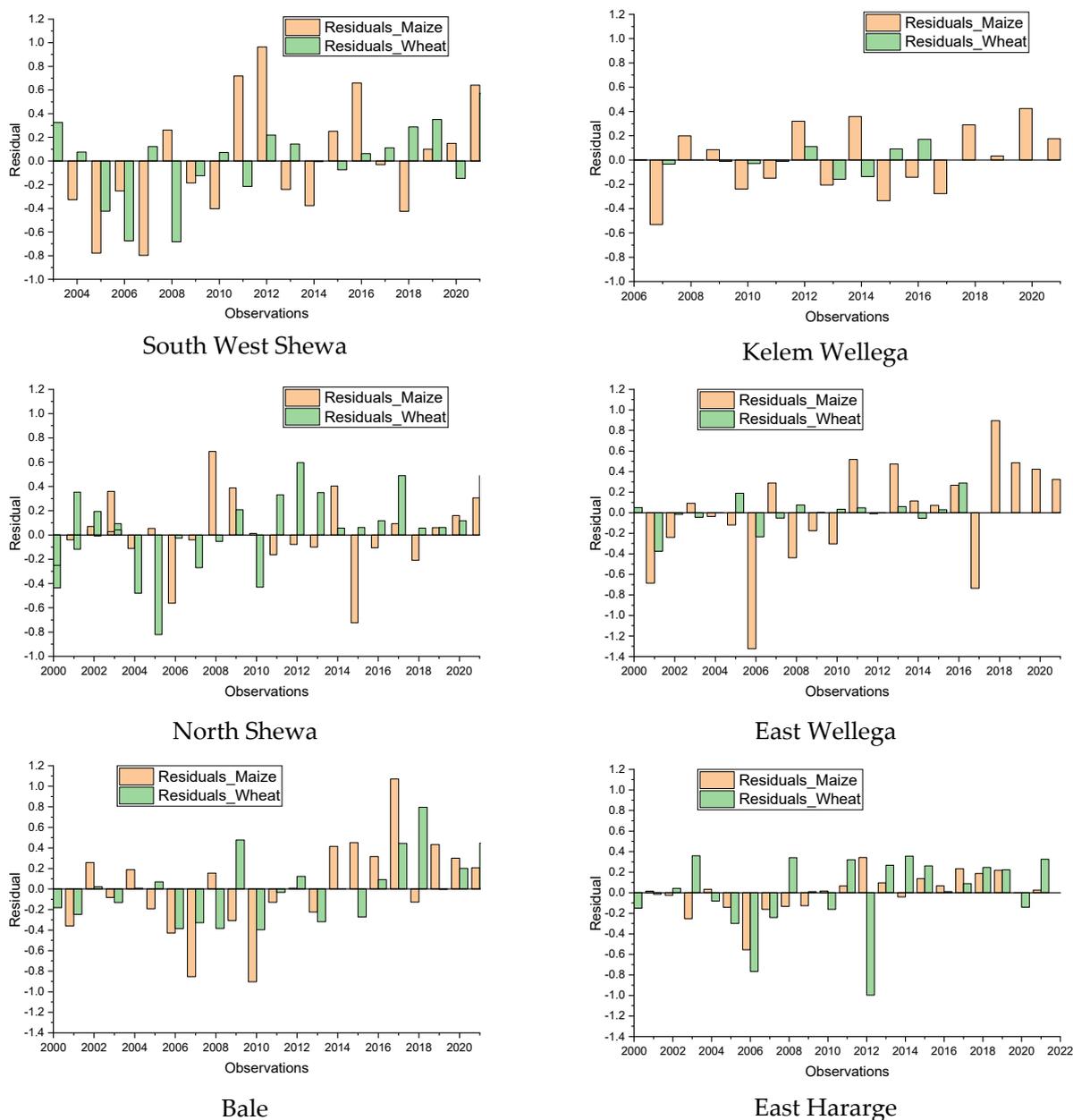
Figure 6. *Cont.*

Figure 6. Residual plot for observed and model-fitted crop yield (maize and wheat) for all the study areas in the zonal administrations.

3.2.2. Models Using NDVI and Climate Predictors

From 2000 to 2021, regression models combining climate and NDVI data were developed to predict crop yields, with their performance summarized in Table 1. For both maize and wheat within the study area, the most effective models were constructed using multiple non-linear regression techniques. The top-performing model for maize was observed in Illu Ababora, achieving an RMSE of 0.477 tons/ha, an R^2 of 0.91, and a d value of 0.976. Conversely, the least effective model for wheat was found in East Hararge, with an RMSE of 0.615 tons/ha, an R^2 of 0.52, and a d value of 0.813.

When comparing models that relied solely on climate data to those incorporating both climate data and the NDVI, significant performance improvements were observed. For instance, in the Arsi zone, the integration of the NDVI increased the R^2 value for maize from 0.71 to 0.79 and for wheat from 0.73 to 0.83. Similarly, in the South West Shewa zone, models using only climate data achieved R^2 values of 0.56 for maize and 0.58 for wheat.

With the inclusion of the NDVI, these values rose to 0.74 for both crops. This marked improvement in predictive performance underscores the value of integrating the NDVI with climate data, as detailed further in Table 1.

3.2.3. Selected Yield Prediction Models

Table 1 presents the top-performing regression models, comparing those relying solely on climate predictors with models incorporating both climate data and the NDVI. The results clearly show models combining the NDVI with climate predictors consistently achieve higher R^2 and d values, along with lower RMSEs, compared to models using only climate data (Table 1). This underscores the substantial benefit of integrating NDVI data in improving the accuracy of crop yield prediction.

Most of the models selected in Tables 1 and A1 utilize combined climate-NDVI predictors, highlighting their superior performance. However, in specific zones such as West Shewa, East Shewa, East Hararge, and East Wellega for maize, and Kelem Wellega for wheat, models based solely on climate predictors remain valuable. These models are particularly effective for estimating yield gaps in these areas, demonstrating their continued importance in yield prediction strategies.

3.3. Predicted and Observed Maize and Wheat Yield

3.3.1. Yield Estimate with Climate Predictors Only

Yield predictions based solely on climate data were generated for the entire dataset, covering the years 2000 to 2021. As shown in Figure 7, these predictions occasionally overestimated or underestimated yields in specific years. Despite these deviations, the overall estimates closely matched the observed yields throughout the study period. The model demonstrated particularly higher accuracy in predicting maize and wheat yields in the Illu Ababora, Horo Guduru, and West Arsi zones. Additionally, it provides especially effective for wheat yield predictions across nearly all the study areas, as highlighted in Figure 7.

3.3.2. Yield Estimate with NDVI and Climate Predictors

An evaluation of the observed and predicted yields for maize and wheat, using two model combinations—climate-only and climate-NDVI—is presented in Figure 7. Models that integrated both climate data and the NDVI (as detailed in Table 1) consistently outperformed those relying solely on climate data. This improvement is further illustrated in Figures 8 and 9, where scatter plots of the predicted versus observed yields for the climate-NDVI models show points closely clustered along the 1:1 line, indicating strong predictive accuracy. In contrast, the climate-only models display a less precise alignment with observed yields in the scatter plots. Overall, the combined climate and NDVI model exhibited a significantly closer match to observed yields across the study period, demonstrating enhanced predictive reliability. While a few outliers were observed, particularly in West Shewa and North Shewa for maize and in East Shewa and West Hararge for wheat, the combined models provided robust predictions across most areas.
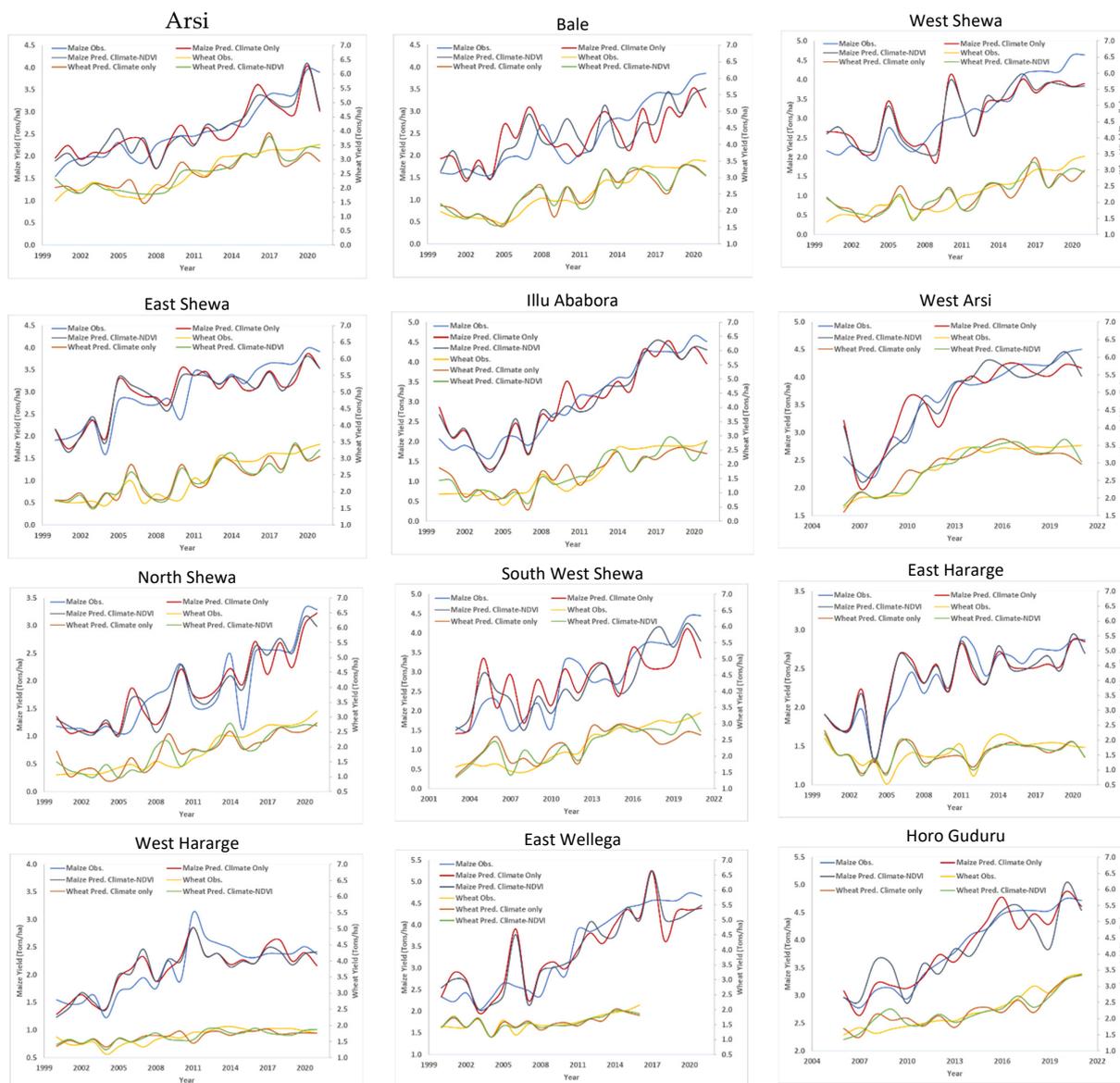
Figure 7. Comparison between observed maize and wheat yield and their corresponding predicted yields generated by the top-performing "Climate only" and "Climate and NDVI" models across the study region.
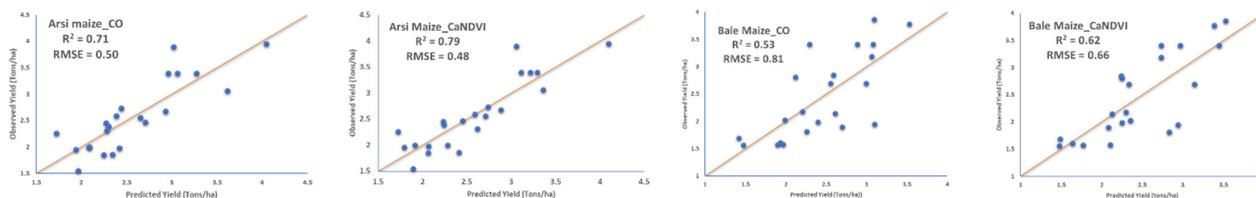
Figure 8. *Cont.*
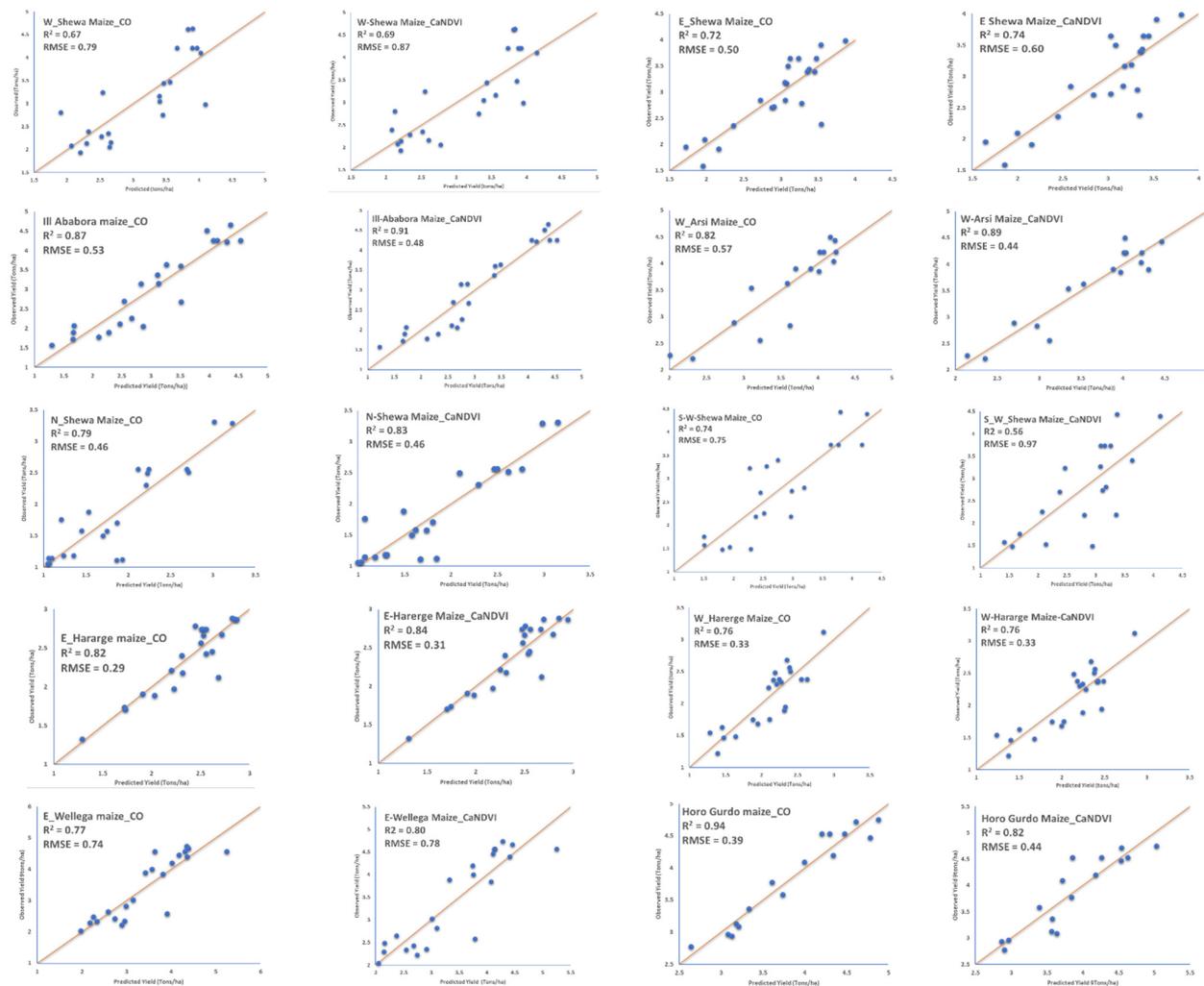
Figure 8. Scatter plots for predicted versus observed maize yield, "CO = Climate only and CaNDVI = Climate-NDVI variables" across the study area.
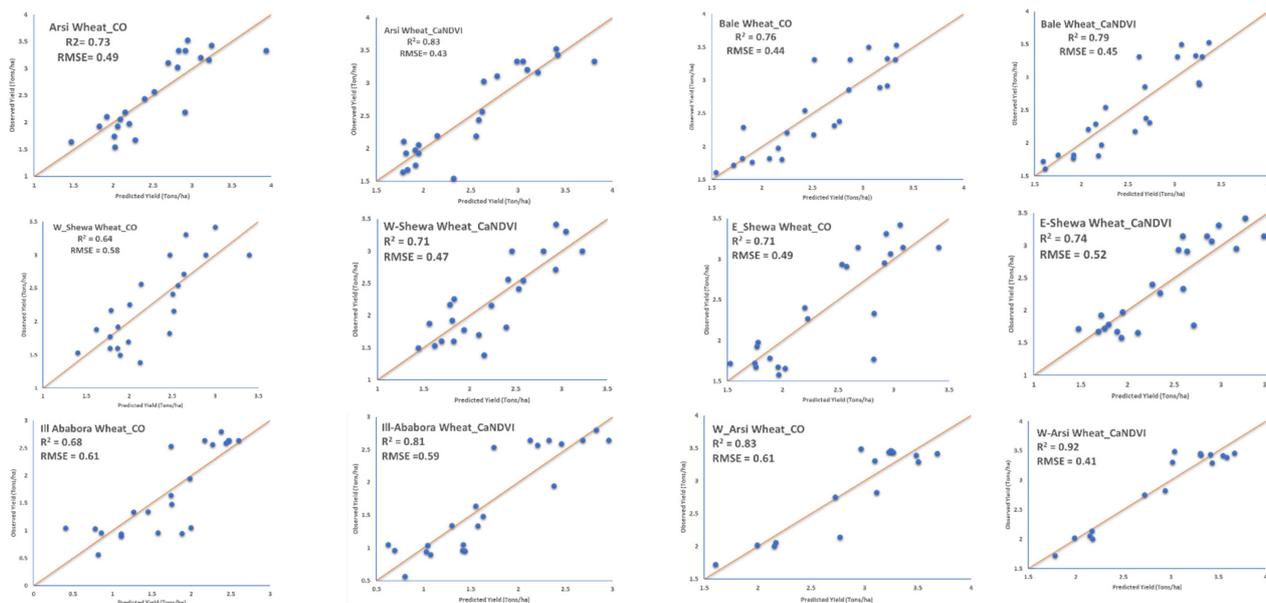
Figure 9. *Cont.*
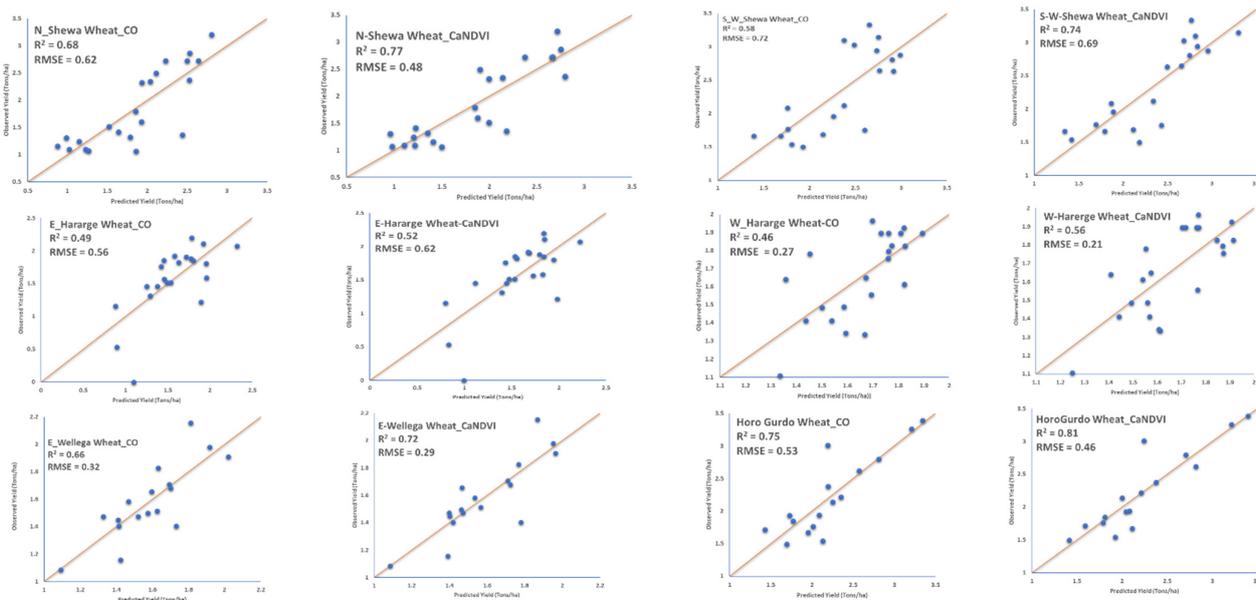
Figure 9. Scatter plots for predicted versus observed wheat yield, "CO = Climate only and CaNDVI = Climate-NDVI variables" across the study area.

3.4. Remote Sensing Cloud Platform Yield Prediction

The spatial distribution of maize and wheat yields from 2013 to 2021 in two key cereal crop production zones of the Oromia region was estimated using the CropWatch crop yield prediction model. The model was specifically calibrated for wheat in the Bale zone and maize in the Illu Ababora zone. The results demonstrated a strong correlation between the recorded yields and model predictions, with an R^2 value of 0.65 and an RMSE value of 0.332 tons/ha. For 2021, the maize yield prediction (Figure 10) showed that most yields per unit area ranged spatially between 4000 and 6000 kg/ha, with an average predicted yield of 4418 kg/ha. This closely aligned with the recorded yield of 4518 kg/ha, highlighting the model's accuracy in capturing spatial yield variations.
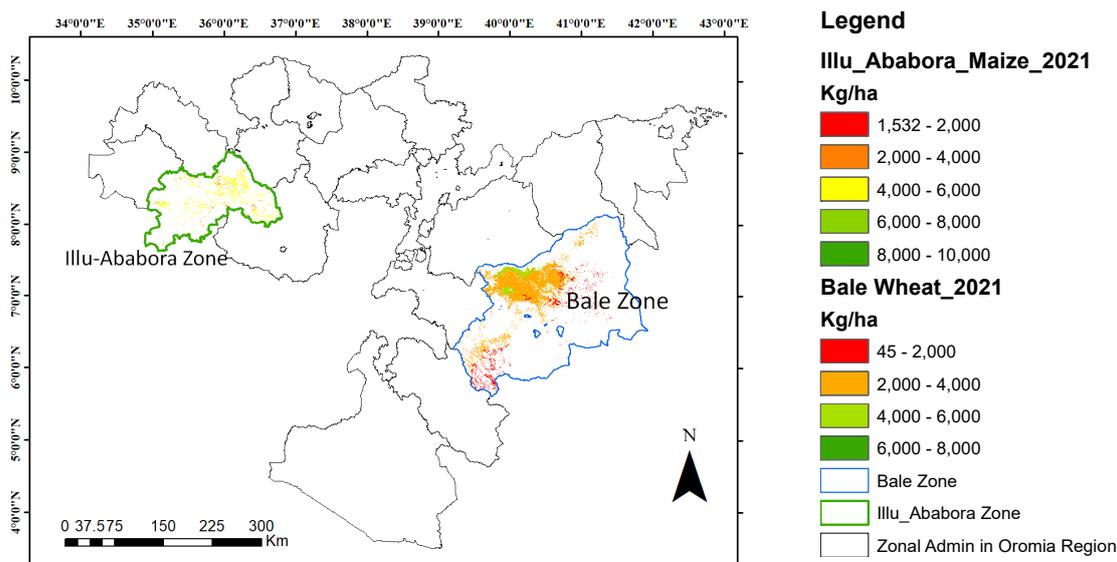
Figure 10. Spatial distribution of maize and wheat crop yield in 2021 in two zones predicted using CropWatch yield prediction model.

The wheat yield prediction for the Bale zone from 2013 to 2021 was also simulated using the same CropWatch crop yield prediction model. The results showed that the spatial distribution of the recorded wheat yields aligned closely with the model's predictions, achieving an R^2 value of 0.67 and an RMSE of 0.16 tons/ha. The 2021 wheat yield prediction (Figure 10) indicated that most yields per unit area ranged between 3000 and 4000 kg/ha, with an average predicted yield of 3299 kg/ha, closely matching the recorded yield of 3505 kg/ha. Figure 11 further illustrates that the CropWatch yield prediction model accurately captured both maize and wheat yields during the 2013 to 2021 period. These results underscore the model's potential for simulating crop yields in the study area. However, its accuracy depends on a thorough assessment of crop phenology and harvest index, as the model is highly sensitive to these parameters. Users should consider the actual agronomic practices and systems in the area to ensure reliable results.

(**a**) (**b**)
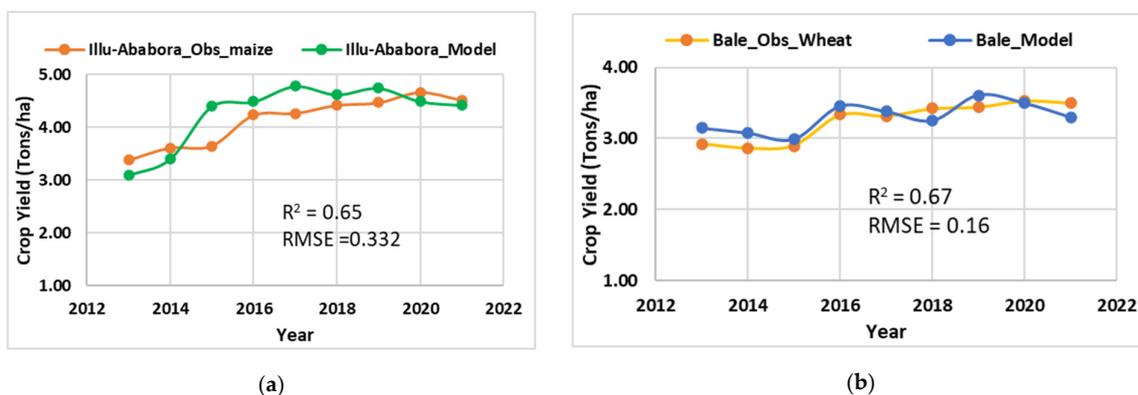
Figure 11. Comparison of observed and predicted (CropWatch crop yield prediction model) yield for the period 2013 to 2021: (**a**) maize, Illu Ababora zone; (**b**) wheat, Bale zone.

4. Discussion

Understanding how various climate variables influence crop growth and yield is crucial. Climate conditions before and during the growing season such as high temperature and water scarcity can delay planting, stress plants, and ultimately reduce yields [7]. Therefore, selecting the appropriate timeframe for analyzing climate variables in yield prediction is essential. Capturing the variability of these conditions is critical for building accurate yield prediction models [42]. Studies have used different periods, such as the entire growing season [1,7,43,44] or monthly climate variable averages, for this purpose. In this study, we focused on the growing season: late May to early November for wheat and from late April to September for maize. This ensured alignment between data collection, analysis periods, and the crops' critical growth phases. Our results highlighted that both the minimum and maximum vapor pressure deficits (VPDx and VPDm), averaged over the growing season, were significant predictors of maize and wheat yields across all the study areas (Figure 5). Furthermore, the mean temperature proved more significant for predicting yields than the individual minimum or maximum temperature assessments. This aligns with findings that temperature effects depend on crops' optimal growth temperature. While moderate warming may benefit certain crops, excessive heat beyond a crop's threshold reduces yields [45], potentially explaining inconsistent temperature yield correlation in our analysis.

According to Debalke and Abebe [1] and Yadav and Geli [7], rainfall was identified as a significant predictor of crop yield in multiple linear regression models, explaining up to 88% of yield variability. Consistent with these findings, our study also included areal rainfall in the models, which resulted in high coefficients of determination (R^2) ranging from 79 to 93% across the study area. The predictive accuracy of our multiple regression

model was notably strong, with RMSE values ranging from 0.154 to 0.792 tons/ha and R^2 values from 93 to 79%, as shown in Figures 8 and 9. These results are comparable to other studies. For instance, Zinna and Suryabhagavan [23] reported an RMSE of 1.41 tons/ha and an R^2 of 88% for maize yield prediction in the South Tigray zone. Similarly, Reda [24] observed an RMSE of 0.99 tons/ha and an R^2 of 93% for wheat yield predictions in the East Arsi zone.

Additionally, this study demonstrated the effectiveness of the seasonal maximum NDVI in predicting yields for major crops such as maize and wheat in Ethiopia's Oromia region. The monthly maximum NDVI captured variations in climate effects on crop yields, yielding low RMSE values and high R^2 values, signifying strong predictive accuracy. In our study, the inclusion of the maximum NDVI (NDVIx) significantly enhanced model prediction accuracy as detailed in Table 1 and Figures 8 and 9. For instance, incorporating the NDVIx into our models improved the R^2 from 71% to 79% for maize and 73% to 83% for wheat in the Arsi zone. Similar improvements were observed in West Arsi zone, where the R^2 increased from 82% to 89% for maize and 83% to 91% for wheat, highlighting the substantial benefits of including the NDVIx in yield predictions. We compared our results with studies that utilized NDVI data for yield prediction. Zinna and Suryabhagavan [23] found that the average NDVI was a significant predictor in their multiple linear regression model, suggesting that the average NDVI is a highly effective parameter for field-level yield predictions. Similarly, Rojas [46] developed a multiple linear regression using the NDVI to forecast maize yield in Kenya, explaining 87% of variations and achieving an RMSE value of 0.333 tons/ha.

The results of our multiple regression model were compared with the CropWatch model, further validating our approach. For maize yield predictions in the Illu Ababora zone, our model, using only climate parameters, achieved an R^2 of 87% and an RMSE of 0.53 tons/ha. Adding the NDVI improved these metrics to an R^2 of 91% and an RMSE of 0.48 tons/ha. In contrast, the CropWatch yield prediction model for maize in the same zone had an R^2 of 65% and an RMSE value of 0.332 tons/ha, Similarly, in the Bale zone, wheat yield predictions improved from an R^2 of 76% (RMSE: 0.44 tons/ha) using climate data alone to an R^2 of 79% (RMSE: 0.45 tons/ha) with the NDVI, outperforming CropWatch's R^2 of 67% (RMSE: 0.16 tons/ha), indicating a good performance with less error compared to the regression model.

Our model's accuracy was validated against Ethiopian Central Statistics Agency (CSA) ground data (Figures 8, 9 and 11). Integrating remote sensing data with climate variables provided superior predictions, closely aligned with observed yields. This remote sensing approach offers several advantages over traditional methods. Notably, it delivers location-specific yield estimates earlier—by October—compared to the conventional December timeline, aiding in timely crop management. This enhanced forecasting capability can empower local administrations, the Central Statistics Agency, and farmers to address food security challenges more effectively.

5. Conclusions

This study aimed to identify key climate variables that significantly influence maize and wheat yields, enabling the development of predictive models for selected regions in Ethiopia. The models used both climate data alone and a combination of climate data with RS imagery. The observed crop yield data served as the dependent variable, while six climatic predictors and NDVI values derived from RS were analyzed. Variables with the strongest correlation, highest index of agreement, and lowest RMSE values were selected to construct multiple non-linear regression models. The findings highlighted the

vapor pressure deficit (minimum and maximum) as a critical factor affecting yield in the study area.

The results demonstrated that combining climate data with the NDVI provided more accurate predictions than using climate data alone. This aligns with the existing literature, which suggests that integrating RS data with climate variables enhances yield prediction accuracy. Additionally, models developed at smaller spatial scales better captured climatic variability, leading to improved predictive performance. The calibrated result of the RS-based CropWatch yield prediction model for maize and wheat in selected areas showed a strong correlation between recorded yields and model predictions, offering reasonably high accuracy comparable to other methods.

This research underscores the importance of identifying critical climate variables and improving the timeliness and accuracy of yield forecasts in Oromia's agricultural systems. Accurate and early yield predictions can significantly aid Ethiopia in developing strategies for crop management and food security. The regression models and the calibrated CropWatch yield prediction model enable yield forecasting well before harvest, supporting better agricultural planning and response strategies. Scaling up spatial calibration and testing of the CropWatch model across other regions of the country is necessary. The model's pixel-level predictions are valuable and easy to use, although they depend on the availability of optical RS data.

Further research is required to establish a representative harvest index value in the area, as this factor influences the accuracy of the CropWatch yield prediction model. Exploring additional crop models that incorporate biomass, harvest indices, or growth models would be beneficial for practical applications. Future studies should also investigate yield forecasting using advanced methods such as machine learning and data-driven algorithms. Expanding the analysis to include a longer time series of data would enhance practical implementation, while incorporating factors like soil characteristics would provide a more comprehensive understanding of yield determinants.

Data Availability Statement: Raw data like crop yield data were generated at Central Statistical Authority of Ethiopia (CSA). Derived data supporting the findings of this study are available from the corresponding author [Asfaw] on request.

Conflicts of Interest: The authors report there are no competing interests to declare.

Appendix A

Table A1. Developed multiple non-linear regression models and code using NDVI and climate variables, or climate variables only for maize and wheat for a time scale of 22 years (2000–2021).

Regression Model Code	Developed Regression Model
Arsi_M	$Y = -954.22 + 0.04 \times ArealRF - 2.41 \times VPDm - 0.46 \times VPDx + 63.15 \times Tmean - 18.2 \times Tm + 1211.04 \times NDVIx - 0.23 \times VPDm^2 + 0.33 \times VPDx2 - 1.87 \times Tmean^2 + 0.89 \times Tm^2 - 724.43 \times NDVIx^2$
Arsi_W	$Y = -410.94 - 68.59 \times VPDm - 48.68 \times VPDx - 24.28 \times Tmean + 31.53 \times Tx + 588.45 \times NDVx + 140.71 \times VPDm^2 + 69.86 \times VPDx^2 + 1.96 \times Tmean^2 - 1.14 \times Tx^2 - 0.93 \times Tm^2 - 321.83 \times NDVIx^2$
Bale_W	$Y = 487.69 + 0.002 \times ArealRF - 4.11 \times VPDm + 24.60 \times VPDx - 21.05 \times Tmean - 546.47 \times NDVIx + 1.59 \times VPDm^2 - 14.1 \times VPDx^2 + 0.442 \times Tmean^2 + 304.23 \times NDVIx^2$
Bale_W	$Y = 65.35 - 0.014 \times ArealRF + 29.44 \times VPDm + 41.20 \times VPDx + 9.41 \times Tmean + 1.69 \times Tx - 488.31 \times NDVIx - 22.79 \times VPDm^2 - 20.66 \times VPDx^2 - 0.197 \times Tmean^2 - 0.04 \times Tx^2 + 275.70 \times NDVIx^2$
West Shewa_M	$Y = 134.9 - 0.08 \times ArealRF + 57.81 \times VPDm - 50.02 \times VPDx - 151.9 \times Tmean + 101.46 \times Tx - 191.96 \times VPDm^2 + 40 - 44 \times VPDx^2 + 0.46 \times Tmean^2 - 0.68*Tx^2 + 3.29*Tm^2$
West Shewa_W	$Y = -1060.7 + 0.01 \times ArealRF - 39.14 \times VPDx + 2.14 \times Tx - 18.6 \times Tm + 2595.5 \times NDVIx + 55.3 \times VPDx^2 - 0.1 \times Tx^2 + 0.98 \times Tm^2 - 1488.8 \times NDVIx^2$
East Shewa_M	$Y = -210.25 + 0.02 \times ArealRF - 52.59 \times VPDm + 2.42 \times VPDx + 56.96 \times Tmean - 24.34 \times Tx + 65.49 \times VPDm^2 - 3.99 \times VPDx^2 - 1.6 \times Tmean^2 + 0.51 \times Tx^2$
East Shewa_W	$Y = 368.91 + 0.02 \times ArealRF - 34.6 \times VPDm - 1.53 \times VPDx + 9.01 \times Tmean - 42.24 \times Tm - 523.8 \times NDVIx + 43.9 \times VPDm^2 - 0.62 \times VPDx^2 - 0.26 \times Tmean^2 + 1.95 \times Tm^2 + 320.4 \times NDVIx^2$
Ill Ababora_M	$Y = -2101.03 - 0.03 \times ArealRF + 28.51 \times VPDm + 5.1 \times VPDx + 39.42 \times Tmean - 0.28 \times Tm + 3995.97 \times NDVIx - 147.54 \times VPDm^2 - 14.9 \times VPDx^2 - 1.1 \times Tmean^2 - 2269.04 \times NDVIx^2$
Ill Ababora_W	$Y = -3670.51 - 0.03 \times ArealRF + 76.6 \times VPDm - 21.7 \times VPDx + 14.97 \times Tmean - 2.73 \times Tmax - 0.8 \times Tm + 8063.5 \times NDVIx - 281.6 \times VPDm^2 + 25.8 \times VPDx^2 - 0.5 \times Tmean^2 + 0.1 \times Tx2 + 0.07 \times Tm^2 - 4525.63 \times NDVIx^2$
West Arsi_M	$Y = -7.67 + 0.03 \times ArealRF + 28.17 \times VPDx - 3.33 \times Tmean - 21.32 \times Tm + 205.41 \times NDVIx - 88.58 \times VPDx^2 + 0.10 \times Tmean^2 + 1.71 \times Tm^2 - 110.19 \times NDVx^2$
West Arsi_W	$Y = -596.99 + 0.1 \times ArealRF - 10.5 \times VPDm - 94.6 \times VPDx + 10.8 \times Tmean + 4.5 \times Tx + 11 \times NDVIx + 105.4 \times VPDm^2 + 399.9 \times VPDx^2 - 0.44 \times Tmean^2 - 0.2 \times Tx^2 - 623.2 \times NDVIx^2$
North Shewa_M	$Y = 108.9 + 0.01 \times ArealRF - 14.6 \times VPDm - 5.4 \times VPDx - 7.4 \times Tmean + 6.6 \times Tm - 237.01 \times NDVIx + 19.82 \times VPDm2 + 1.95 \times VPDx2 + 0.22 \times Tmean2 - 0.3 \times Tm2 + 158.60 \times NDVIx2$
North Shewa_W	$Y = -188 + 0.002 \times ArealRF - 5.7 \times VPDx - 3.2 \times Tx + 2.45 \times Tm + 463.14 \times NDVIx + 2.62 \times VPDx^2 + 0.1 \times Tx^2 - 0.09 \times Tm^2 - 256.9 \times NDVx^2$
South west Shewa_M	$Y = -570.5 + 0.02 \times ArealRF + 30.72 \times VPDx + 229.8 \times Tmean - 133.5 \times Tx + 31.5 \times NDVIx - 24.6 \times VPDx2 - 5.7 \times Tmean^2 + 2.5 \times Tx^2 + 7.2 \times NDVIx^2$
South west Shewa_W	$Y = -0.6 \times ArealRF - 140.5 \times VPDm - 79.7 \times VPDx + 85.53 \times Tmean - 23.1 \times Tx - 1037.7 \times NDVIx + 245 \times VPDm^2 + 89 \times VPDx^2 + 2.02 \times Tmean^2 - 1.13 \times Tx^2 - 3.2 \times Tm^2 + 634 \times NDVIx^2$
East Hararge_M	$Y = -194.07 + 0.017 \times ArealRF - 23.25 \times VPDm + 26.65 \times VPDx + 4.71 \times Tmean + 17.09 \times Tx - 3.99 \times Tm + 55.98 \times VPDm2 - 37.89 \times VPDx2 - 0.09 \times Tmean2 - 0.44 \times Tx2 + 0.14 \times Tm2$
East Hararge_W	$Y = 252.8 - 0.03 \times ArealRF - 6.72 \times VPDm + 28.41 \times VPDx + 10.62 \times Tmean - 18.3 \times Tx - 4.7 \times Tm - 297.5 \times NDVIx + 10.2 \times VPDm^2 - 28.3 \times VPDx^2 - 0.3 \times Tmean^2 + 0.42 \times Tx^2 + 0.2 \times Tm^2 + 179.7 \times NDVIx^2$
West Hararge_M	$Y = 57.9 - 0.003 \times ArealRF - 7.02 \times VPDm - 492.75 \times NDVIx - 15.33 \times Tmean + 20.8 \times Tx + 11.2 \times VPDm^2 + 317.1 \times NDVIx^2 + 0.35 \times Tmean^2 - 0.35 \times Tx^2$
West Hararge_W	$Y = -121.94 + 0.002 \times ArealRF - 2.4 \times VPDm - 18.4 \times Tmean + 8.1 \times Tx + 471.03 \times NDVIx + 2.2 \times VPDm^2 + 0.41 \times Tmean^2 - 0.13 \times Tx^2 - 262.3 \times NDVIx^2$
East Wellega_M	$Y = -734.89 - 0.04 \times ArealRF - 122.71 \times VPDm - 16.39 \times VPDx - 73.93 \times Tmean + 112.25 \times Tx + 66.83 \times VPDm^2 + 8.15 \times VPDx^2 + 1.18 \times Tmean^2 - 1.84 \times Tx^2 + 0.59 \times Tm^2$
East Wellega_W	$Y = 204.5 + 0.004 \times ArealRF + 100.3 \times VPDm - 35.8 \times VPDx + 16.2 \times Tmean - 32.2 \times Tm - 383.2 \times NDVIx - 346.5 \times VPDm^2 + 57.2 \times VPDx^2 - 0.41 \times Tmean^2 + 1.3 \times Tm^2 + 229.4 \times NDVIx^2$
Kellem Wellega_M	$Y = -1841.6 + 466.8 \times VPDm - 60.4 \times VPDx + 18 \times Tm + 3872.1 \times NDVIx - 1203.9 \times VPDm^2 + 71.0 \times VPDx^2 - 0.65 \times Tm^2 - 2226.9 \times NDVIx^2$
Kellem Wellega_W	$Y = 7.49 + 0.03 \times ArealRF - 137.62 \times VPDm - 21.27 \times VPDx + 348.6 \times VPDm^2 + 26.74 \times VPDx^2$
Horo Gurdro_M	$Y = 3790.8 - 0.07 \times ArealRF + 1224.1 \times VPDm + 15.5 \times VPDx - 9350.9 \times NDVIx - 3027.4 \times VPDm^2 - 14.5 \times VPDx^2 + 5660 \times NDVx^2$
Horo Gurdro_W	$Y = -2973 - 96.5 \times VPDx + 18.6 \times Tmean - 7.8 \times Tx + 6824.6 \times NDVIx + 103.5 \times VPDx^2 - 0.42 \times Tmean^2 + 0.13 \times Tx^2 - 4002.3 \times NDVIx^2$

References

1. Debalke, D.B.; Abebe, J.T. Maize yield forecast using GIS and remote sensing in Kaffa Zone, South West Ethiopia. *Environ. Syst. Res.* **2022**, *11*, 1. [CrossRef]

2. Nyéki, A.; Neményi, M. Crop Yield Prediction in Precision Agriculture. *Agronomy* **2022**, *12*, 2460. [CrossRef]

3. Ulfa, F.; Orton, T.G.; Dang, Y.P.; Menzies, N.W. Developing and Testing Remote-Sensing Indices to Represent within-Field Variation of Wheat Yields: Assessment of the Variation Explained by Simple Models. *Agronomy* **2022**, *12*, 384. [CrossRef]

4. Piekutowska, M.; Niedbała, G.; Piskier, T.; Lenartowicz, T.; Pilarski, K.; Wojciechowski, T.; Pilarska, A.A.; Czechowska-Kosacka, A. The Application of Multiple Linear Regression and Artificial Neural Network Models for Yield Prediction of Very Early Potato Cultivars before Harvest. *Agronomy* **2021**, *11*, 885. [CrossRef]

5. Wu, B.; Zhang, M.; Zeng, H.; Tian, F.; Potgieter, A.B.; Qin, X.; Yan, N.; Chang, S.; Zhao, Y.; Dong, Q.; et al. Challenges and opportunities in remote sensing-based crop monitoring: A review. *Natl. Sci. Rev.* **2022**, *10*, nwac290. [CrossRef]

6. Beyene, A.N.; Zeng, H.; Wu, B.; Zhu, L.; Gebremicael, T.G.; Zhang, M.; Bezabh, T. Coupling remote sensing and crop growth model to estimate national wheat yield in Ethiopia. *Big Earth Data* **2022**, *6*, 18–35. [CrossRef]

7. Yadav, K.; Geli, H.M.E. Prediction of Crop Yield for New Mexico Based on Climate and Remote Sensing Data for the 1920–2019 Period. *Land* **2021**, *10*, 1389. [CrossRef]

8. Lobell, D.B.; Burke, M.B. On the use of statistical models to predict crop yield responses to climate change. *Agric. For. Meteorol.* **2010**, *150*, 1443–1452. [CrossRef]

9. Li, Y.; Zeng, H.; Zhang, M.; Wu, B.; Qin, X. Global de-trending significantly improves the accuracy of XGBoost-based county-level maize and soybean yield prediction in the Midwestern United States. *GIScience Remote Sens.* **2024**, *61*, 2349341. [CrossRef]

10. Li, Y.; Zeng, H.; Zhang, M.; Wu, B.; Zhao, Y.; Yao, X.; Cheng, T.; Qin, X.; Wu, F. A county-level soybean yield prediction framework coupled with XGBoost and multidimensional feature engineering. *Int. J. Appl. Earth Obs. Geoinform.* **2023**, *118*, 103269. [CrossRef]

11. Bingfang, W.; Miao, Z.; Hongwei, Z.; Guoshui, L.; Sheng, C.; Gommes, R. New indicators for global crop monitoring in CropWatch -case study in North China Plain. In Proceedings of the IOP Conference Series: Earth and Environmental Science, Beijing, China, 22–26 April 2014.

12. Du, X.; Wu, B.; Li, Q.; Meng, J.; Jia, K. A Method to Estimated Winter Wheat Yield with the MERIS Data. In Proceedings of the Progress in Electromagnetics Research Symposium (PIERS) Proceedings, Beijing, China, 23–27 March 2009.

13. Wu, B.; Gommes, R.; Zhang, M.; Zeng, H.; Yan, N.; Zou, W.; Zheng, Y.; Zhang, N.; Chang, S.; Xing, Q.; et al. Global Crop Monitoring: A Satellite-Based Hierarchical Approach. *Remote Sens.* **2015**, *7*, 3907–3933. [CrossRef]

14. Asfew, M.; Bedemo, A. Impact of Climate Change on Cereal Crops Production in Ethiopia. *Adv. Agric.* **2022**, *2022*, 2208694. [CrossRef]

15. CSA. *Report on Area and Production of Major Crops (Private Peasant Holdings, Meher Season), Statistical Bulletin*; Annual Report; Central Statistical Agency: Addis Ababa, Ethiopia, 2021.

16. Senbeta, A.F.; Worku, W. Ethiopia's wheat production pathways to self-sufficiency through land area expansion, irrigation advance, and yield gap closure. *Heliyon* **2023**, *9*, e20720. [CrossRef] [PubMed]

17. Tadesse, T.; Senay, G.B.; Berhan, G.; Regassa, T.; Beyene, S. Evaluating a satellite-based seasonal evapotranspiration product and identifying its relationship with other satellite-derived products and crop yield: A case study for Ethiopia. *Int. J. Appl. Earth Obs. Geoinf.* **2015**, *40*, 39–54. [CrossRef]

18. Weldearegay, S.K.; Tedla, D.G. Impact of climate variability on household food availability in Tigray, Ethiopia. *Agric. Food Secur.* **2018**, *7*, 6. [CrossRef]

19. Nath, K.P.; Behera, B. A critical review of impact of and adaptation to climate change in developed and developing economies. *Environ. Dev. Sustain.* **2011**, *13*, 141–162. [CrossRef]

20. Mwaniki, A. Achieving Food Security in Africa: Challenges and Issues. Cornell University: Ithaca, NY, USA, 2004. U.S. Plant, Soil and Nutrition Laboratory. Available online: https://repositori.kpkm.gov.my/handle/123456789/85 (accessed on 23 May 2024).

21. Devereux, S.; Simon, M. *Food Insecurity in Sub-Saharan Africa*; Practical Action Publishing: London, UK, 2001.

22. Birara Endalew, B.E.; Mequanent Muche, M.M.; Tadesse, S. Assessment of Food Security Situation in Ethiopia: A Review. *Asian J. Agric. Res.* **2017**, *9*, 55–68.

23. Zinna, A.W.; Suryabhagavan, K.V. Remote Sensing and GIS Based Spectro-Agrometeorological Maize Yield Forecast Model for South Tigray Zone, Ethiopia. *J. Geogr. Inf. Syst.* **2016**, *8*, 282–292. [CrossRef]

24. Reda, A.F. Wheat Yield Forecast Using Remote Sensing and GIS in East Arsi Zone, Ethiopia. Doctoral Dissertation, Addis Ababa University, Addis Ababa, Ethiopia, 2015.

25. CSA. *Agricultural Sample Survey; Report on Farm Management Practices 2018*; Annual Report; Central Statistical Agency: Addis Ababa, Ethiopia, 2019.

26. Tetens, O. Uber einige meteorologische Begriffe. *Z. Geophys.* **1930**, *6*, 297–309.

27. Becker-Reshef, I.; Vermote, E.; Lindeman, M.; Justice, C. A generalized regression-based model for forecasting winter wheat yields in Kansas and Ukraine using MODIS data. *Remote Sens. Environ.* **2010**, *114*, 1312–1323. [CrossRef]

28. Mkhabela, M.; Bullock, P.; Raj, S.; Wang, S.; Yang, Y. Crop yield forecasting on the Canadian Prairies using MODIS NDVI data. *Agric. For. Meteorol.* **2011**, *151*, 385–393. [CrossRef]

29. Vintrou, E.; Desbrosse, A.; Bégué, A.; Traoré, S.; Baron, C.; Seen, D.L. Crop area mapping in West Africa using landscape stratiication of MODIS time series and comparison with existing global land products. *Int. J. Appl. Earth Obs. Geoinf.* **2012**, *14*, 83–93.

30. Kouadio, L.; Newlands, N.K.; Davidson, A.; Zhang, Y.; Chipanshi, A. Assessing the Performance of MODIS NDVI and EVI for Seasonal Crop Yield Forecasting at the Ecodistrict Scale. *Remote Sens.* **2014**, *6*, 10193–10214. [CrossRef]

31. Johnson, D.M. An assessment of pre- and within-season remotely sensed variables for forecasting corn and soybean yields in the United States. *Remote Sens. Environ.* **2014**, *141*, 116–128. [CrossRef]

32. Friedl, M.; Gray, J.; Sulla-Menashe, D. MODIS/Terra+Aqua Land Cover Dynamics Yearly L3 Global 500m SIN Grid V061. 2022, distributed by NASA EOSDIS Land Processes Distributed Active Archive Center. 2022. Available online: https://doi.org/10.5067/MODIS/MCD12Q2.061 (accessed on 28 May 2024).

33. Gorfu, D.; Ahmed, E. *Crops and Agro-Ecological Zones of Ethiopia*; Ethiopian Institute of Agricultural Research: Addis Ababa, Ethiopia, 2012.

34. Rijks, O.; Massart, M.; Rembold, F.; Gommes, R.; Leo, O. Crop and rangeland monitoring in eastern Africa. In Proceedings of the 2nd International Workshop, Nara, Japan, 29–31 October 2007; pp. 95–104.

35. Worku, M.; Zelleke, H. Advances in Improving Harvest Index and Grain Yield of Maize in Ethiopia. *East Afr. J. Sci.* **2007**, *1*, 112–119. [CrossRef]

36. Belay, M.K. Growth, Yield-Related Traits and Yield of Lowland Maize (*Zea mays* L.) Varieties as Influenced by Inorganic NPS and N Fertilizer Rates at Babile, Eastern Ethiopia. *Int. J. Agron.* **2020**, *2020*, 8811308. [CrossRef]

37. Friendly, M. Corrgrams: Exploratory displays for correlation matrices. *Am. Stat.* **2002**, *56*, 316–324. [CrossRef]

38. Gareth, J.; Daniela, W.; Trevor, H.; Robert, T. *An Introduction to Statistical Learning: With Applications in R*; Springer: Berlin/Heidelberg, Germany, 2014.

39. Wang, J.; He, P.; Liu, Z.; Jing, Y.; Bi, R. Yield estimation of summer maize based on multi-source remote-sensing data. *Agron. J.* **2022**, *114*, 3389–3406. [CrossRef]

40. Yang, J.; Liu, S.; Hoogenboom, G. An evaluation of the statistical methods for testing the performance of crop models with observed data. *Agric. Syst.* **2014**, *127*, 81–89. [CrossRef]

41. Willmott, C.J. On the validation of models. *Phys. Geogr.* **1981**, *2*, 184–194. [CrossRef]

42. Siebert, S.; Webber, H.; Rezaei, E.E. Weather impacts on crop yields—Searching for simple answers to a complex problem. *Environ. Res. Lett.* **2017**, *12*, 081001. [CrossRef]

43. Urban, D.; Roberts, M.J.; Schlenker, W.; Lobell, D.B. Projected temperature changes indicate significant increase in interannual variability of U.S. maize yields. *Clim. Change* **2012**, *112*, 525–533. [CrossRef]

44. You, L.; Rosegrant, M.W.; Wood, S.; Sun, D. Impact of growing season temperature on wheat productivity in China. *Agric. Forest Meteorol.* **2009**, *149*, 1009–1014. [CrossRef]

45. Hatfield, J.; Takle, G.; Grotjahn, R.; Holden, P.; Izaurralde, R.C.; Mader, T.; Marshall, E.; Liverman, D. *Ch. 6: Agriculture—Climate Change Impacts in the United States: The Third National Climate Assessment*; Melillo, J.M., Richmond, T.C., Yohe, G.W., Eds.; U.S. Global Change Research Program: Washington, DC, USA, 2014; pp. 150–174. Available online: http://nca2014.globalchange.gov/report/sectors/agriculture (accessed on 17 January 2025).

46. Rojas, O. Operational maize yield model development and validation based on remote sensing and agro-meteorological data in Kenya. *Int. J. Remote Sens.* **2007**, *28*, 3775–3793. [CrossRef]