Coupling remote sensing and crop growth model to estimate national wheat yield in Ethiopia

Awetahegn Niguse Beyene, Hongwei Zeng, Bingfang Wu, Liang Zhu, Tesfay Gebretsadkan Gebremicael, Miao Zhang and Temesgen Bezabh

State Key Laboratory of Remote Sensing Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China; College of Resources and Environment, University of Chinese Academy of Sciences, Beijing, China; Tigray Agricultural Research Institute, Mekelle, Ethiopia; Climate Change and Rural Development, Mekelle University, Mekelle, Ethiopia

ABSTRACT
Estimation of crop yield at a regional level is essential for making agricultural planning and addressing food security issues in Ethiopia. Remote sensing observations, particularly the leaf area index (LAI), have a strong relationship with crop yield. This study has proposed an approach to estimate wheat yield at field level and regional scale in Ethiopia by assimilating the retrieved MODIS time-series LAI data into the WOFOST model. To improve the estimation of crop yield in the region, the Ensemble Kalman Filter (EnKF) was used to incorporate the LAI into the WOFOST model. The estimation accuracy of wheat crop yield was validated using field-measured yields collected during the 2018 growing season. Our findings indicated that wheat yield was more precisely estimated by WOFOST (at water-limited mode) with EnKF algorithm ($R^2 = 0.80$ and RMSE = 413 kg ha$^{-1}$) compared to that of without assimilating remotely sensed LAI ($R^2 = 0.58$, RMSE = 592 kg ha$^{-1}$). These results demonstrated that assimilating MODIS-LAI into WOFOST has high potential and practicality to give a reference for wheat yield estimation. The findings from this study can provide information to policy, decision-makers, and other similar sectors to implement an appropriate and timely yield estimation measure.

1. Introduction

Food insecurity is a long-standing issue in Ethiopia (Tadesse, Senay, Berhan, & Regassa, 2015), which is nearly a quarter of its population is malnourished and the largest proportion is food insecure and suffers from chronic hunger (Beyene, 2010). The Ethiopian agricultural system is predominantly dependent on rain-fed and as a result, the food security at the household level is vulnerable and greatly affected by low availability of moisture (Mann & Warner, 2017), rainfall fluctuation (Birhanu, Alwin, & Manfred, 2011), higher atmospheric carbon dioxide (Muluneh & Biazin, 2015) and farming activities (e.g. practices of soil and water conservation, chemical fertilizer, land and livestock holdings

CONTACT Hongwei Zeng zenghw@radi.ac.cn State Key Laboratory of Remote Sensing Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100101, China

© 2021 The Author(s). Published by Taylor & Francis Group and Science Press on behalf of the International Society for Digital Earth, supported by the CASEarth Strategic Priority Research Programme. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
and losses of nutrients and soil (Adimassu, Mekonnen, Yirga, & Kessler, 2014). As wheat is the major staple crop in the country, early and accurate monitoring of yield is a critical issue for risk control, management, and food security. Moreover, such information can provide an early warning on production reduction (Tadesse et al., 2015). Satellite-based remote sensing products with high spatiotemporal resolution are cost-effective to acquire information for crop yield estimations (Tadesse et al., 2015). A few studies focused on applying remote sensing products as a proxy indicator for yield estimation has been carried out around the world (e.g. Tadesse et al., 2015; Kassie, Rötter, Hengsdijk, Asseng, & Ittersum, 2013; Mann & Warner, 2017). Such remotely sensed information is essential to explore the relationships of vegetation indices with crop yield. For example, a significant positive correlation was found between maize yield and enhanced vegetation index (EVI), normalized difference vegetation index (NDVI), and wheat yield and EVI with wheat crop yield in different countries (Mann & Warner, 2017). Another way of using vegetation indices to build the link among crop yields and water stress factors, such as surface temperature, soil moisture, rainfall (Prasad, Chai, Singh, & Kafatos, 2006), and evapotranspiration (Tadesse et al., 2015). Tadesse et al. (2015) estimated the yield of most cereal crops (e.g. sorghum, corn/maize, teff, wheat, and barley) in Ethiopia based on the relationships between their yield and evapotranspiration. However, the yield formation process of major crops is dynamic and complicated and this makes it difficult to capture the impacts of weather, fertilizer, and management on the yield formation using relationships between remote sensing indices and crop yield (Tadesse et al., 2015).

The weaknesses of yield estimation through remote sensing can be eliminated by crop growth models. Crop growth models are able to represent the physical processes of crops during simulation of yields (de Wit et al., 2018). In the last few decades, physical models have been developed to forecast crop yield. For example, Decision Support System for Agrotechnology Transfer (DSSAT) (Jones et al., 1998), World Food Studies (WOFOST) (Van Diepen, Wolf, & Van Keulen, 1989), Agricultural Production System Simulator (APSIM) (McCown, Hammer, Hargreaves, Holzworth, & Huth, 1995), and AquaCrop model (Raes, Steduto, Hsiao, & Fereres, 2009; Steduto, Hsiao, Raes, & Fereres, 2009) are the most repeatedly applied to estimate cereal crop yields. These crop models were able to simulate the full crop growth process and their interactions with soil, weather, and crop management practices. Similarly, these crop models have also been applied in Ethiopia to estimate cereal crop yield. For example, Kassie et al. (2013) estimate maize yield gap and climate-induced of yield variability in the central rift valley of Ethiopia using both WOFOST and DSSAT models. Similarly, the yield of potato under future climate change projections was predicted at Mekelle Areas, in Northern Ethiopia using the AquaCrop model (Gebremedhin & Berhe, 2015). However, uncertainties of yield estimation using these models are caused by weather, climate, and model structures (Hansen, Challinor, Ines, Wheeler, & Moron, 2006).

Remote sensing observations have paramount importance in retrieving spatiotemporal crop conditions within the crop growing season (de Wit et al., 2018). Assimilating data from remote sensing at the entire growing season into crop models can potentially reduce model-related uncertainty (De Wit & Van Diepen, 2007). Data assimilation (DA) can assimilate remote sensing information for correcting errors in crop models and hence to enhance the simulation accuracy of yield estimation (Raes et al., 2009). Recalibration/re-
initialization methods and state-updating methods are the major approaches in data assimilation as it contributes to reducing the uncertainties between observations and initial assumptions of model condition (Kassie et al., 2013). Remote sensing leaf area index (LAI) representing the canopy photosynthetic capacity has a strong relationship with crop condition, biomass, grain yield, and which has been used in agricultural crop productivity estimation (Li, Chen, et al., 2017).

Several practices over the entire world proved that assimilation is a reliable method to minimize the uncertainty happening in crop yield estimation (Wang, Huang, et al., 2017). The performance of wheat yield estimation was improved after assimilating the LAI and surface soil moisture (SSM) into the CERES-Wheat model using the particle filter algorithm (Pan, Chen, Allard, & Ren, 2019). Similarly, wheat yield estimation accuracy at the regional scale was substantially improved with a four-dimensional variation data assimilation compared with unassimilated results (Huang et al., 2015). Because of its typical state-updating method, Ensemble Kalman filter (EnKF) is widely used in crop growth monitoring, yield estimation (Li et al., 2014; Ma et al., 2013) and agricultural drought monitoring (Han, Crow, Holmes, & Bolten, 2014). For example, the overall performance of yield forecasting was significantly improved after the assimilating of MODIS-LAI and AMSR-E soil moisture into the WOFOST crop model using EnKF (Ines et al., 2013). MODIS time-series LAI datasets were used as an observation data and to investigate the performance and competence of the EnKF assimilation method to increase crop yield estimation accuracy (Han et al., 2014). In summary, previous studies clearly showed that the integration of LAI and WOFOST crop model using EnKF assimilating method is an important strategy on the estimation of crop yield at higher accuracy. However, no studies were carried out in the region to estimate the wheat crop yield by assimilating remotely sensed information into a crop growth model using the EnKF assimilation techniques.

Therefore, this study gives a feasible way to predict wheat yield in Ethiopia through evaluating the performance difference of wheat yield estimation at the field and regional levels by integrating remote sensing indicators, physical crop model, and assimilating remote sensing indicator into the physical crop growth model. This study also provides better insights into the importance of the LAI in understanding how wheat yield can be predicted.

2. Materials and methods

2.1. Site description

Ethiopia is located in the Horn Africa between 32°42′E-48°12′E and 3°30′N-14°50′N (Figure 1). The average annual rainfall of the country ranges from 550 mm year⁻¹ in the northern and eastern to more than 2000 mm year⁻¹ in western and southwestern parts of the country (Kim, Kaluarachchi, & Smakhtin, 2009; Tesemma, Mohamed, & Steenhuis, 2010). The annual mean temperature of the country varies between 12°C and 15.1°C (Edwards, Berhan, Egziabher, & Araya, 2010). The soil types across the study region consist of mainly hypereutrophic Nitisols, hypereutric Vertisols, hypereutric Luvisols, and vertic Fluvisols.

The agricultural sector of the country contributes to 42% of the gross domestic product (GDP), 85% to the total employment and 80% to the export earnings of the country. As
one of the major cereal crops in Ethiopia, wheat is planted in more than 4.7 million smallholder farmers and occupied about 1.6 million hectares of land. Ethiopian wheat crop is generally planted in early July and harvested in late October. The average yearly wheat yield in the country is about 2.71 tons ha\(^{-1}\), significantly lower than the world average wheat yield (3.51 tons ha\(^{-1}\)).

2.2. Data collection and pre-processing

2.2.1. Wheat yield and phenological data

Ground measurement on the yield of the wheat crop was collected from each location to validate the estimated yield from the remote sensing and crop growth model. The ground crop yield was collected from a survey conducted from October 20 to November 10, 2018 (Figure 2). Yield from 119 sample plots was collected during the field campaign. The samples were taken from Tigray (59 samples) and Oromia region (60 samples) where are known for their main wheat-producing regions in Ethiopia. The sample plots were designed based on the different wheat-growing conditions of the country. These two regions, which are characterized by different agro-ecological conditions are the most wheat-producing areas, and the number of samples collected from both regions can represent the different wheat-growing areas of the country. Samples obtained from the Tigray region are representative of the wheat-growing belts in the Northern Ethiopia, while samples from the Oromia region are enough to represent wheat-growing areas in the central parts of the country. Overall, the data collection was carried out based on the following criteria:

(1) The selected sample plots were located at a minimum of 100 m away from roads, residential areas or trees to minimize their influence on the measurements;

(2) Each sample plot was taken from a farm size of more than 300 m × 300 m, with three 1 m × 1 m areas measured in each plot. This sampling size was taken considering the

Figure 1. Location and distribution of ground sample plots and meteorological stations of the study area (Tigray is northern while Oromia is the central part in the maps shown above).
Figure 2. Field observations and measurement photos. Crop cutting yield measurements using quadrant method (a), weighting samples after threshing at Kulumsa Agriculture Research Center (b), wheat sample location confirmation (c) and wheat crop phenology and wet weight measurements (d).
farmer’s landholding size which is not beyond 300 m × 300 m. To make sure the selected MODIS pixels are within the specified area and to avoid taking samples from mixed crop areas, we selected sampling plots from adjacent land who are only wheat-growing farmers.

2.2.2. Soil, weather, and crop data
The saturated soil moisture content, bulk density, wilting point of soil, water content at field capacity are critical parameters for the WOFOST model. These datasets were obtained from the Africa Soil Information Service (http://africasoils.net/services/data/soil-databases/) with a spatial resolution of 250 m. The features and sources of the soil parameters are summarized in the supplementary file (Table S1). Weather data including, daily maximum and minimum temperature, humidity, vapour pressure, precipitation, wind speed and sunshine hours were obtained from eight meteorological stations, provided by the National Meteorological Agency of Ethiopia (NMAE).

Crop information such as sowing dates, plant spacing, density, and fertilizer rate and application time was collected during the field campaign from October 20, 2018, to November 12, 2018.

In the study areas, the crop-growing duration ranges from 110 to 130 days. This variation is attributed mainly to the variations in wheat cultivars, rainfall distribution, and altitude. As a result, wheat crop phenology, including the flowering date is varied across the different wheat-growing regions of Ethiopia. For this study, the assumption was that different flowering dates existed in the two study regions. The detailed crop information and associated parameter values used in the model are summarized in the supplementary material (Table S2). The fraction of the above-ground dry matter of organs, stem and leaves were analyzed in Kulumsa Agricultural Research Centre (MARC) and Ethiopian Agricultural Research Institute.

2.2.3. Leaf area index data
LAI has been extensively used in many crop estimation and forecasting, because it is an essential vegetation biophysical parameter and it can be forecasted by many crop models. In this study, LAI of the MOD15A2H collection 6 data was accessed from LPDAAC NASA products (http://doi.10.5067/MODIS/MOD15A2H.006). The abnormal MODIS-LAI pixels caused by clouds, vapour and aerosol were eliminated according to FparLAI_QC (Xiao, Liang, Wang, Jiang, & Li, 2011). Next, atmospheric correction, geometric correction, and mosaicking re-projection of the datasets were done using the Modis Reprojection (MRT) Tools. Finally, Savitzky–Golay (SG) filter method (Chen et al., 2004) was employed to fill the missed pixels and confirm the key phenological stage. The values of the points of window and the polynomial order of SG filter were set to 11 and 2 in this research work.

3. Methods
To estimate the regional wheat yield over Ethiopia, a combination of three methods was employed. First, pre-processing and extraction of critical crop phenology stages were done from the MODIS-LAI time-series and assessed by linear regression model between remotely sensed LAI and observed wheat yield. Next, the WOFOST model was calibrated using crop parameters, soil, and weather data. After the calibration of the WOFOST model,
the LAI, crop phenology, and grain yield were predicted. Finally, the EnKF assimilation algorithm was applied to run at the date when a new observation value of LAI becomes available from the calibrated model. The assimilated remotely sensed LAI and WOFOST were applied to forecast regional wheat yield. The general workflow of wheat yield estimation is summarized in Figure 3.

**3.1. Assessing sensitive stage of wheat phenology using remotely sensed data and field observed data**

The entire growing period, including emergence, anthesis, flowering, and maturity stage of wheat extends from July to October in both study regions. This study attempted to assess the most sensitive phenology stage of wheat crop using linear regression models. The performance of the model was evaluated using statistical indices such as determination ($R^2$) and Root Mean Square Error (RMSE). The model with the highest coefficient of determination ($R^2$) and the lowest RMSE is considered as the best fitting model (Aptula, Jeliazkova, Schultz, & Cronin, 2005).

The first step in this approach is to extract dates of key phenological stages, including emergence, anthesis, flowering, and maturity. Based on the process of the LAI time profile of SG filter, the dates of four key phenological stages were identified using the phenology extraction method demonstrated in Wang et al. (2017). Basic principles of data confirmation are (1) emergence data: the minimal point along with the LAI profile; (2) flowering date: LAI reach to peak during the growth period; (3) maturity date: LAI curve begins to decrease; (4) harvesting date: the maturity stage reaches to leaf senescence.

Next, a linear model was developed to identify the relationship between LAI in a specific stage and wheat yield from the collected samples and the most sensitive growth stages. To
calibrate the WOFOST model and verify the extracted crop phenology from LAI, 20 field experimental sites were selected during the field survey at the period of wheat-growing months (from June to October 2018). These 20 field experimental sites were collected from areas that are representative across the wheat-growing regions. The dates of emergence, anthesis, and maturity were measured from each experimental site. The model with optimal cultivar parameters was quantitatively evaluated by comparing the simulated crop phenology and yield with the corresponding observations. The WOFOST simulated emergence, anthesis, and maturity dates agreed with the measured dates.

### 3.2. WOFOST calibration and yield estimation

The WOFOST model was developed by Wageningen Environmental Research (De Wit & Van Diepen, 2007) for simulating all crop growth process, yield, and their quantitative analysis. The model has been used globally to assess potential crop productivity, and explore the influences of climate change, potential water availability, and fertilizer management on grain yields (Mann & Warner, 2017; McCarty, Neigh, Carroll, & Wooten, 2017). WOFOST model simulates the whole process of crop growth by considering the interaction of crops with a light interception, taking CO₂ assimilation as growth-driving processes and phenological development (Wang, Li, et al., 2017). The WOFOST model provides three scenarios to simulate crop growth: (1) potential mode: crop growth driven without any limiting factors, (2) water-limited mode: considering the scarce of water availability and low soil physical prosperities, (3) nutrition-limited mode: which depends on the supply of nutrients to the crop.

As the study area is commonly known for its water scarcity and low moisture availability, the WOFOST water-limited was selected to simulate wheat yield. The most sensitive model parameters of WOFOST, such as DTSMTB, TBASEM, TSUMEM, TSUM1, AMAXTB, SPAN, CVO, SLATB, and TDWI were selected for this study (Table S2). The parameters of the WOFOST model were calibrated using field-measured data of crop management, weather, and soil obtained from the field campaign in 2018 (Table S2) and yield data from Mekelle and Kulumsa Agricultural Research Centers. Both Mekelle and Kulumsa Agricultural Research Centers are center of excellence for wheat in Ethiopia. We collected these data during the growing season of 2018 from their field experiments in different areas in Tigray and Oromia, respectively.

During the calibration of the crop model, we used forcing data (i.e. meteorological data and soil data) and observations data (i.e. flowering date, maturity date, and yield) which were applied during the wheat-growing season. This step is important to improve the model simulation accuracy. Model input parameters (Table S2) were calibrated until the simulated and observed yields reached the best agreement. This procedure continued until the best-performing parameter values (such as date of flowering, dates of maturity, biomass, and yield) were generated. The calibrated WOFOST model was then validated using wheat grain yield collected from 119 experimental sites. Detailed descriptions of the most important parameters for model calibration and the input crop parameters of the WOFOST model are summarized in Table S2.
3.3. EnKF-based assimilation and yield estimation

The EnKF-based assimilation algorithm was coded and coupled with the WOFOST model using the PCSE/Python programming language. EnKF algorithm was proposed by Evensen (2003) and its principle can be expressed as Eq. (1)-(3):

\[ Y_t = HX_t + \epsilon_t \]  
\[ X_t^f = MX_{t-1}^a + W_t \]  
\[ X_t^a = X_t^f + K_t(Y_t - HX_t^f) \]

Where, \( Y \) is the observation ensemble, \( H \) is the observation operator matrix, \( X \) is state variable, \( \epsilon \) is the random noise of the observation, \( X_t^f \) is an ensemble of forecast, \( X_t^a \) is the optimal estimate ensemble, the mean of \( X_t^a \) is the best estimate at time \( t \). \( M \) is the equation that transfers the state from time \( t \) to time \( t-1 \), and \( W_t \) is the error that is produced when the state transfers. \( K_t \) is the Kalman gain matrix and it is the key variable of EnKF data assimilation and can be estimated using Eqs. (4)-(9) (Burgers, Jan van Leeuwen, & Evensen, 1998).

\[ K_t = P_t^f H^T (HP_t^f H^T + R_t)^{-1} \]  
\[ P_t^f = \frac{1}{N-1} \sum_{i=1}^{N} (X_{i,t}^f - \overline{X_t^f})(X_{i,t}^f - \overline{X_t^f})^T \]  
\[ \overline{X_t^f} = \frac{1}{N} \sum_{i=1}^{N} X_{i,t}^f \]  
\[ H P_t^f H^T = \frac{1}{N-1} \sum_{i=1}^{N} (H(X_{i,t}^f) - H(\overline{X_t^f}))(H(X_{i,t}^f) - H(\overline{X_t^f}))^T \]  
\[ R_t = \frac{1}{N-1} \sum_{i=1}^{N} (Y_{i,t} - \overline{Y_t})(Y_{i,t} - \overline{Y_t})^T \]  
\[ Y_t = \frac{1}{N} \sum_{i=1}^{N} Y_{i,t} \]

Where \( N \) is the number of ensemble members, \( H^T \) is the inverse matrix of \( H \), \( \overline{X_t^f} \) is the mean of forecast ensemble, \( \overline{Y_t} \) is the mean of the observation ensemble, \( X_{i,t}^f \) is the error covariance of the forecast ensemble, \( R_t \) is the error covariance of the observation ensemble. Detailed information about EnKF above can be obtained from Ma et al. (2013).

In this study, the Gaussian perturbation was introduced into the WOFOST model with 10% of uncertainty level. During the first day of the simulation, the white Gaussian noise was added to shift the simulated LAI and generate the first ensemble of the forecasted LAI.
(f_{LAI}^1, f_{LAI}^2, \ldots, f_{LAI}^n). If there is observation LAI at time t, a Gaussian perturbation was added into the observed MODIS LAI to generate an observed LAI ensemble (OLAI^1, OLAI^2, \ldots, OLAI^n). The forecasted and the observation LAI ensembles were then assimilated with EnKF to obtain the optimal LAI estimate ensemble (LAI^1, LAI^2, \ldots, LAI^n), which was then added to the WOFOST crop growth model to obtain the forecast ensemble at time t+1. If an observation LAI did not exist for time t, the forecasted LAI ensemble was directly put into the WOFOST crop growth model. This process was repeated until the simulated wheat by WOFOST reaches the maturity stage. The mean of the optimal estimated ensemble provides the best estimate of LAI at time t, which is finally input into the WOFOST crop growth model to estimate the wheat yield.

### 3.4. Model performance evaluation

Data assimilation and its performance on yield estimation were evaluated by comparing the measured values against the predicted values. Two indicators, the $R^2$ (Eq.10) and Root Mean Square Error (RMSE, Eq. 11), are employed to measure the difference between estimated and ground measured yield.

\[
R^2 = \frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}
\]

\[
RMSE = \sqrt{\frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{n}}
\]

Where $\hat{y}_i$ and $y_i$ are predicted and actual wheat yield respectively and $\bar{y}$ is the average actual wheat yield.

### 4. Results

#### 4.1. Identification of sensitive crop growth stages from remote sensing data and field observation data

The emergence, anthesis, flowering and maturity dates were observed on July 10, August 12, September 5 and October 11, respectively. Based on this information, linear regression models between LAI and wheat yield were constructed at anthesis, flowering, and maturity stages. The results summarized in Table 1 indicate that LAI values at flowering stages were the most closely linked to wheat yield.

Compared to anthesis and maturity periods, a better agreement with LAI was found at the flowering stage ($R^2 = 0.48$ and RMSE = 661.53 kg ha$^{-1}$). LAI at the flowering stage has a stronger correlation with wheat yield compared to the other two stages that clearly shows this stage is the most sensitive and important to predict wheat yield.

The WOFOST model was calibrated using observed phenological stages collected from the field survey. The calibrated model was applied to predict the crop growth stages and validated with an independent observed crop phenological stage (Figure 4(a)). The performance of the model indicated by the RMSE and RE during simulation of emergency,
anthesis, and maturity were 5 days and 2.6%, 4 days and 2%, and 11 days and 3.38%, respectively. The estimated yield from the 20 field plots was also in consistency of the predicted yield from the WOFOST model (Figure 4(b)). This implies the extracted crop phenology from MODIS LAI time-series is reliable to predict crop phenology stages and grain yield.

4.2. Wheat grain yield estimation based on the WOFOST model

After calibrating at water-limited mode, the WOFOST model was applied to predict wheat yield at 119 sites. Compared to the observed values at the experimental sites, the model was able to predict the wheat grain yield with satisfactory performance ($R^2 = 0.58$, RMSE = 594.5 kg ha$^{-1}$) (Figure 5(a)). Besides, grain yield was also simulated at the 119 sites and regional level through the assimilating of smoothed LAI into the WOFOST model using EnKF assimilation approach. The performance of the model in predicting grain yield was improved after the data assimilation with $R^2$ increasing from 0.58 to 0.8 and RMSE decreasing from 594.5 to 414.9 kg ha$^{-1}$ (Figure 5(b)).

Furthermore, this study employed the assimilation method to generate a spatial distribution of wheat during the growing season of 2018 (Figure 6). Based on the comparison results at the national scale, the average wheat yield is 3616.3 kg ha$^{-1}$, and RMSE equals to 586.5 kg ha$^{-1}$. The predicted wheat yield showed a higher variability that ranges from

### Table 1. Analysis results of the observed wheat yield and LAI relations at anthesis, flowering, and maturity during the 2018 field observation.

| Critical growing periods | Regression Model | Adj.$R^2$ | RMSE (kg ha$^{-1}$) |
|-------------------------|------------------|-----------|---------------------|
| Anthesis (August 12)    | $Y = 496.96x + 3258.79$ | 0.36      | 732.87              |
| Flowering (September 5) | $Y = 787.38x + 3035.00$ | 0.48      | 661.53              |
| Maturity (October 10)   | $Y = 1255.52x + 2978.26$ | 0.40      | 764.28              |

$Y$ and $X$ represent yield and LAI values, respectively.
The regional yield distributions showed a decreasing pattern from the central to the northern parts of Ethiopia (Figure 6). The estimated yield showed that the maximum potential of wheat production in the central region. Particularly, in Oromia the estimated yield was larger (5000 kg ha\(^{-1}\)) than that of the northern region (3120 kg ha\(^{-1}\)) (Figure 6). The findings of this study are in agreement with previous studies in the country that have concluded wheat productivity in the central part of the country is higher than that of the northern part (McCarty et al., 2017;
Minot et al., 2015). The observed decreasing pattern in wheat yield could be mainly attributed to the decreasing pattern in moisture availability toward the northern part. In a comparison of moisture availability, the northern part of the country is commonly known for their high moisture stress and there receives lower annual rainfall than central part (Mann & Warner, 2017; White, Tanner, & Corbett, 2001). Another reason could be due to poor farm management practices, lower soil fertility, and the potential of the areas to grow wheat (White et al., 2001).

5. Discussion

In this study, the WOFOST crop model with assimilating remotely sensed LAI using the EnKF method was applied to estimate wheat yield at the specific sites and at a national level. The results showed good insight into grain yield estimation by integrating remote sensing information with crop models. A total of 119 sites observed wheat yield were used to validate the predicted result before and after assimilation. The finding of this study shows that the performance of yield estimation has improved after assimilating remote sensing LAI into the calibrated WOFOS model. A significant increase in the performance indicators was observed after assimilation. The flowering stage appeared to be the most sensitive stage for wheat yield estimation.

Our findings are in agreement with previous studies (e.g. Pan et al., 2019; Li, Chen, et al., 2017; Ines, Das, Hansen, & Njoku, 2013). Pan et al. (2019) assimilated the Soil Moisture (SM) and LAI from sentinel-1 and 2 into the WOFOST and the result showed that the RMSE indices decreased by 69, 39 and 169 kg ha\(^{-1}\) after assimilating LAI, SM and the combination of them, respectively. Similar studies in other countries such as China and the United States have reported an increase in the estimation accuracy of yield by assimilating MODIS LAI into the Decision Support System for the Agrotechnology Transfer-cropping System (DSSAT-CSM)-Maize model and improved the accuracy from 0.47 to 0.51 (Ines et al., 2013). Li, Jiang, et al. (2017) applied the same method and improved estimation accuracy from 0.81 to 0.84. In contrast to this finding, a study by Huang et al. (2019) in China reported that assimilation of LAI with 4DVar showed a lower performance \(R^2 = 0.44\) and RMSE = 598 kg ha\(^{-1}\)). This implies that our results are reasonably acceptable in terms of its effects on the assimilating of biophysical variables using EnKF instead of estimating by crop models. The data assimilation methods are suitable for operation in a large area, such as EnKF algorithm (De Wit & Van Diepen, 2007). This demonstrates that the application of combined WOFOST model and remote sensed LAI is a promising approach to forecast wheat yield in Ethiopia.

According to the report from IndexMundi (https://www.indexmundi.com/agriculture/?country=et&commodity=wheat&graph=yield), the average wheat yield of Ethiopia in 2018 was 2609 kg ha\(^{-1}\). Our result at the national level is higher than the result of IndexMundi. The actual average wheat yield at the 119 sites reached 4108.22 kg ha\(^{-1}\), which is much higher than the average wheat yield of IndexMundi. This is mainly due to the observed yield that might be limited by important nutrients (e.g. nitrogen) and mostly affected by the yield-limiting factors (e.g. weeds, diseases, pests). However, the estimated grain yield by the model has an assumption of the yield-limiting factors were under control. The reason could be due to the differences in capturing crop characteristics and management practices considered. There is a strong spatial heterogeneity of LAI and
intensive field measurement that is required to take enough samples over larger areas. This explicitly shows that further study is needed on retrieving LAI using remotely sensed data with higher temporal and spatial resolutions.

6. Conclusions

In this study, we proposed an assimilation approach to integrating LAI into the WOFOST crop model to predict wheat yield of Ethiopia. Remote sensing LAI was assimilated at the key growth stages into WOFOST to estimate the wheat grain yield in Ethiopia. The estimated wheat yield with and without assimilating MODIS-LAI were compared in order to estimate at the regional and a national level, respectively. We found that the LAI at the flowering stage is more sensitive to wheat yield. The assimilated values showed a significant improvement in wheat yield simulation accuracy with an increase of $R^2$ from 0.58 to 0.8 and a decrease of RMSE from 594.5 kg ha$^{-1}$ to 414.9 kg ha$^{-1}$. The result of this study explained that assimilating remote sensing LAI into the WOFOST crop growth model is an efficient strategy for large-scale wheat yield estimation in Ethiopia. The findings can help policy and decision-makers and other similar sectors to implement appropriate and timely measurable yield estimation strategies. Assimilation of different combined input parameters including soil moisture, LAI, evapotranspiration into crop growth models is a further research agenda to investigate which combinations can best improve the crop yield estimation.

Acknowledgments

We are very grateful to thank Dr. Taifeng Dong from the Science and Technology Branch, Agriculture and Agri-Food Canada for improving this manuscript. We also thank Kulumsa Agricultural Research Center and Tigray Agricultural Research Institute (TARI) for their support and overall contributions to successful cooperation. One of the authors, Awetahegn Niguse Beyene, acknowledges the Chinese Academy of Science (CAS) and the World Academy of Science (TWAS) for awarding the CAS-TWAS President’s Fellowship Program to carry out the research. Finally, we express our deep gratitude to the editors for their hard work.

Data availability statement

The data that support the findings of this study are available from the corresponding author, [Hongwei Zeng], or the first author [Awetahegn Niguse Beyene] upon reasonable request.

Disclosure statement

The authors declare no conflict of interest.

Funding

This work was supported by the National Key Research and Development Project of China [2019YFE0126900], and Strategic Priority Research Program of Chinese Academy of Sciences [XDA19030200].
Notes on contributors

Dr. Hongwei Zeng is an associate professor at the State Key Laboratory of Remote Sensing Science, Aerospace Information Research Institute of Chinese Academy of Sciences. He received his Ph.D. in Natural Resources from the Graduate School of the Chinese Academy of Sciences in 2012. He is a master’s tutor at the University of Chinese Academy of Sciences, a key member of the global agricultural remote sensing monitoring team (CropWatch) of China and a member of the Executive Committee of the Global Earth Observation Global Agricultural Monitoring (GEOGLAM) Flagship. His research interests include remote sensing identification of crop types, remote sensing estimation methods for crop yields, and comprehensive evaluation of water resources availability at the basin scale driven by machine learning. He has published more than 40 scientific papers.

Dr. Wu Bingfang is a professor at the State Key Laboratory of Remote Sensing Science, Aerospace Information Research Institute of Chinese Academy of Sciences. He is the founder and leader of the global agricultural remote sensing monitoring team (CropWatch) of China, and the co-chair of the Global Earth Observation Global Agricultural Monitoring (GEOGLAM) Flagship. He has made outstanding achievements in agricultural remote sensing monitoring, river basin water resources management and other fields. He has published more than 400 scientific research papers.

Dr. Miao Zhang is an associate professor at the State Key Laboratory of Remote Sensing Science, Aerospace Information Research Institute, Chinese Academy of Sciences. He received his Ph.D. in Cartography and Geographical Information System from University of Chinese Academy of Sciences in 2014. He worked as a key member of CropWatch team which serves as remote sensing-based global agricultural monitoring Cloud Platform. He was nominated as Executive Committee member of GEOGLAM Flagship and one of the co-leads of the Joint Experiments for Crop Assessment and Monitoring (JECAM) Initiative which strengthened the internationally collaborative relationships with researchers from more than 20 countries and other international organizations under GEOGLAM context. His research focuses on remote sensing-based agriculture monitoring, mainly includes crop classification, crop area estimation, crop condition assessment and yield prediction. He got more than 60 scientific papers since 2010 and his research outputs span across various disciplines, including remote sensing methodology, water resource, and agronomy.

Dr. Liang Zhu is an assistant professor at the State Key Laboratory of Remote Sensing Science, Aerospace Information Research Institute, Chinese Academy of Sciences. He received his Ph.D. in Cartography and Geographical Information System from Institute of Remote Sensing Application, Chinese Academy of Sciences in 2009. He worked as a key member of CropWatch team which serves as remote sensing-based global agricultural monitoring Cloud Platform. His research focuses on LUCC and remote sensing-based agriculture monitoring. He has published more than 20 scientific papers.
**Awetahegn Niguse Beyene** is a Ph.D. student at the University of Chinese Academy of Sciences, Aerospace Information Research Institute, Beijing, China. He was born in Ethiopia in 1988. He received B.Sc. degree in Meteorology Science from Arbamich University (Ethiopia) in 2010 and MSc. degree in Climate Science from Mekelle University in 2015 (Ethiopia). His research focuses on yield forecasting and estimation based on remote sensing technologies and crop growth models. He is also interested in understanding climate variability and anthropogenic effects on crop growth and grain yield.

**Dr. Tesfay Gebretsadkan** is an agricultural engineer by training and specialized in hydrology and water resources. He is presently a senior researcher in water resources and hydrology at Tigray Agricultural Research Institute, Ethiopia. He has wide experience in catchment hydrology, water resources and hydrological modelling, sedimentation and river systems modelling, irrigation water management, Water resources management, remote sensing, climate and spatial analysis. Moreover, he has vast skills in hydrological modelling and programming tools application for water management. He has supervised several MSc and PhD students in Netherlands and Ethiopia. He has published widely in international top journals and conferences and served as a reviewer for several international peer-reviewed journals.

**Temesgen Bezabh Hagos** is a Ph.D. student in Climate change and Rural development at Mekelle University, Ethiopia. He was born in Ethiopia in 1988. He received B.Sc. degree in Animal Science from Mekelle University in 2009 and MSc. degree in Climate and Society from Mekelle University in 2014. His research focuses on climate change and variability impacts on crop yield. He is also interested in agroclimatology, studying climate change and variability impacts in crop and livestock productivity and associated climate policy issues.

**References**

Adimassu, Z., Mekonnen, K., Yirga, C., & Kessler, A. (2014). Effect of soil bunds on runoff, soil and nutrient losses, and crop yield in the central highlands of Ethiopia. *Land Degradation & Development, 25*(6), 554–564. https://doi.org/10.1002/ldr.2182.

Aptula, A. O., Jeliazkova, N. G., Schultz, T. W., & Cronin, M. T. D. (2005). The better predictive model: High q2 for the training set or low root mean square error of prediction for the test set? *QSAR & Combinatorial Science, 24*(3), 385–396.

Beyene, F. (2010). Determinants of food security among rural households of central Ethiopia: An empirical analysis. *Quarterly Journal of International Agriculture, 49*(4), 299–318. doi:10.22004/ag.econ.155555.

Burgers, G., Van Leeuwen, P. J., & Evensen, G. (1998). Analysis scheme in the ensemble Kalman filter. *Monthly Weather Review, 126*(6), 1719–1724.

Chen, J., Jönsson, P., Tamura, M., Gu, Z., Matsushita, B., & Eklundh, L. (2004). A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky–Golay filter. *Remote Sensing of Environment, 91*(3), 332–344. doi:10.1016/j.rse.2004.03.014.

Demeke, A. B., Keil, A., & Zeller, M. (2011). Using panel data to estimate the effect of rainfall shocks on smallholders food security and vulnerability in rural Ethiopia. *Climatic Change, 108*(1), 185–206. doi:10.1007/s10584-010-9994-3.
de Wit, A. J. W., & Van Diepen, C. A. (2007). Crop model data assimilation with the Ensemble Kalman filter for improving regional crop yield forecasts. *Agricultural and Forest Meteorology*, 146(1–2), 38–56. https://doi.org/10.1016/j.agrformet.2007.05.004.

de Wit, A., Boogaard, H., Fumaqallib, D., Janssena, S., Knapena, R., van Kraalingen, D., … Van D., K. (2018). 25 years of the WOFOST cropping systems model. *Agricultural Systems*, 0–1. doi:10.1016/j.agsy.2018.06.018.

Edwards, S., Egziabher, T. B. G., & Araya, H. (2010). Sucesses and challenges in ecological agriculture: Experiences from Tigray, Ethiopia. Food and Agriculture Organization of the United Nations, Rome, Italy, (pp. 1–58).

Evensen, G. (2003). The Ensemble Kalman Filter: Theoretical formulation and practical implementation. *Ocean Dynamics*, 53(4), 343–367.

Gebremedhin, Y., & Berhe, A. A. (2015). Impact of climate change on potato yield (Solanum tuberosum L.) At Mekelle Areas, in Northern Ethiopia. *World Journal of Agricultural Sciences*, 11, 62–69.

Han, E., Crow, W. T., Holmes, T., & Bolten, J. (2014). Benchmarking a soil moisture data assimilation system for agricultural drought monitoring. *Journal of Hydrometeorology*, 15(3), 1117–1134.

Hansen, J. W., Challinor, A., Ines, A., Wheeler, T., & Moron, V. (2006). Translating climate forecasts into agricultural terms: Advances and challenges. *Climate Research*, 33, 27–41.

Huang, J., Gómez-Dans, J. L., Huang, H., Ma, H., Wu, Q., Lewis, P. E., … Xie, X. (2019). Assimilation of remote sensing into crop growth models: Current status and perspectives. *Agricultural and Forest Meteorology*, 276–277, 107609.

Huang, J., Tian, L., Liang, S., Ma, H., Becker-Reshef, I., Huang, Y., … Wu, W. (2015). Improving winter wheat yield estimation by assimilation of the leaf area index from Landsat TM and MODIS data into the WOFOST model. *Agricultural and Forest Meteorology*, 204, 106–121. https://doi.org/10.1016/j.agrformet.2015.02.001.

Ines, A. V. M., Das, N. N., Hansen, J. W., & Njoku, E. G. (2013). Assimilation of remotely sensed soil moisture and vegetation with a crop simulation model for maize yield prediction. *Remote Sensing of Environment*, 138, 149–164.

Jones, J. W., Tsuji, G. Y., Hoogenboom, G., Hunt, L. A., Thornton, P. K., Wilkens, P. W., … Singh, U. (1998). Decision support system for agrotechnology transfer. *DSSAT*, v3, 157–177.

Kassie, B. T., RöTter, R. P., Hengsdijk, H., Asseng, S., Van Ittersum, M. K., Kahluoto, H., & Van Keulen, H. (2013). Climate variability and change in the Central Rift Valley of Ethiopia: Challenges for rainfed crop production. *The Journal of Agricultural Science*, 152(1), 58–74. doi:10.1017/S0021859612000986.

Kim, U., Kaluarachchi, J. J., & Smakhtin, V. U. (2009). Generation of monthly precipitation under climate change for the upper blue nile river basin. *Ethiopia*, 1(44), 1231–1247.

Li, H., Chen, Z., Liu, G., Jiang, Z., & Huang, C. (2017). Improving winter wheat yield estimation from the CERES-Wheat model to assimilate leaf area index with different assimilation methods and spatio-temporal scales. *Remote Sensors*, 9, doi:10.3390/rs9030190.

Li, H., Jiang, Z., Chen, W., Ren, Z. X., Liu, J. Q., & Hasituya, B. (2017). Assimilation of temporal-spatial leaf area index into the CERES-Wheat model with ensemble Kalman filter and uncertainty assessment for improving winter wheat yield estimation. *Journal of Integrative Agriculture*, 16(10), 2283–2299.

Li, Y., Zhou, Q., Zhou, J., Zhang, G., Chen, C., & Wang, J. (2014). Assimilating remote sensing information into a coupled hydrology-crop growth model to estimate regional maize yield in arid regions. *Ecological Modelling*, 291, 15–27.

Ma, H., Huang, J., Zhu, D., Liu, J., Su, W., Zhang, C., & Fan, J. (2013). Estimating regional winter wheat yield by assimilation of time series of HJ-1 CCD NDVI into WOFOST–ACRM model with Ensemble Kalman Filter. *Mathematical and Computer Modelling*, 58(3), 759–770. doi:https://doi.org/10.1016/j.mcm.2012.12.028.

Mann, M. L., & Warner, J. M. (2017). Ethiopian wheat yield and yield gap estimation: A spatially explicit small area integrated data approach. *Field Crops Research*, 201, 60–74.
McCarty, J. L., Neigh, C. S. R., Carroll, M. L., & Wooten, M. R. (2017). Extracting smallholder cropped area in Tigray, Ethiopia with wall-to-wall sub-meter WorldView and moderate resolution Landsat 8 imagery. *Remote Sensing of Environment*, 202, 142–151.

McCown, R. L., Hammer, G. L., Hargreaves, J. N. G., Holzworth, D., & Huth, N. I. (1995). APSIM: An agricultural production system simulation model for operational research. *Mathematics and Computers in Simulation*, 39(3–4), 225–231.

Minot, N., Warner, J., Lemma, S., Kasa, L., Tadesse, G., & Rashid, S. (2015). The wheat supply chain in Ethiopia: Patterns, trends, and policy options. Washington, DC: International Food Policy Research Institute (IFPRI). http://ebrary.ifpri.org/cdm/ref/collection/p15738coll2/id/130011

Muluneh, A., & Biazin, B. (2015). Impact of predicted changes in rainfall and atmospheric carbon dioxide on maize and wheat yields in the Central Rift Valley of Ethiopia. *Regional Environmental Change*, 15(6), 1105–1119.

Pan, H., Chen, Z., de Wit, A., & Ren, J. (2019). Joint assimilation of leaf area index and soil moisture from sentinel-1 and sentinel-2 data into the WOFOST model for winter wheat yield estimation. *Sensors*, 19(4). doi:10.3390/s19143161.

Prasad, A. K., Chai, L., Singh, R. P., & Kafatos, M. (2006). Crop yield estimation model for Iowa using remote sensing and surface parameters. *International Journal of Applied Earth Observation and Geoinformation*, 8(1), 26–33.

Raes, D., Steduto, P., Hsiao, T. C., & Fereres, E. (2009). AquaCrop- The FAO crop model to simulate yield response to water: II. Main algorithms and software description. *Agronomy Journal*, 101(3), 438–447.

Steduto, P., Hsiao, T. C., Raes, D., & Fereres, E. (2009). Aquacrop-the FAO crop model to simulate yield response to water: I. Concepts and underlying principles. *Agronomy Journal*, 101(3), 426–437.

Tadesse, T., Senay, G. B., Berhan, G., & Regassa, T. (2015). International Journal of applied earth observation and geoinformation evaluating a satellite-based seasonal evapotranspiration product and identifying its relationship with other satellite-derived products and crop yield: A case study for Ethiopia. *International Journal of Applied Earth Observation and Geoinformation*, 40, 39–54.

Tesemma, Z. K., Mohamed, Y. A., & Steenhuis, T. S. (2010). Trends in rainfall and runoff in the Blue Nile Basin: 1964–2003. *Hydrological Processes*, 24(25), 3747–3758. doi:10.1002/hyp.7893.

Van Diepen, C. V., Wolf, J., & Van Keulen, H. R. (1989). WOFOST: A simulation model of crop production. *Soil Use and Management*, 5(1), 16–24.

Wang, C., Li, J., Liu, Q., Zhong, B., Wu, S., & Xia, C. (2017). Analysis of differences in phenology extracted from the enhanced vegetation index and the leaf area index. *Sensors*, 17(9). doi:10.3390/s17091982.

Wang, L., Huang, J., Gao, P., & Wu, H. (2017). Estimating winter wheat yield by assimilation of MODIS LAI into WOFOST model with Ensemble Kalman Filter. Paper presented at the 2017 6th International Conference on Agro-Geoinformatics. Fairfax, VA, USA. doi:10.1109/Agro-Geoinformatics.2017.8047023.

White, J. W., Tanner, D. G., & Corbett, J. D. (2001). An agro-climatical characterization of bread wheat production areas in Ethiopia (NRG-GIS Series 01-01). Mexico, DF: CIMMYT.

Xiao, Z., Liang, S., Wang, J., Jiang, B., & Li, X. (2011). Real-time retrieval of Leaf Area Index from MODIS time series data. *Remote Sensing of Environment*, 115(1), 97–106.