Remote Sensing Applications: Society and Environment 36 (2024) 101272

Contents lists available at ScienceDirect

Remote Sensing Applications: Society and Environment

journal homepage: www.elsevier.com/locate/rsase

The influence of temporal resolution on crop yield estimation with Earth Observation data assimilation

Biniam Sisheber [a, b, *], Michael Marshall [a], Daniel Mengistu [b], Andrew Nelson [a]

[a] Department of Natural Resources, Faculty of Geo-information Science and Earth Observation (ITC), University of Twente, Hallenweg 8, 7522 NH, Enschede, the Netherlands

[b] Geospatial Data and Technology Center (GDTC) and Department of Geography and Environmental Studies, Bahir Dar University, Bahir Dar, Ethiopia

ARTICLE INFO

Keywords:
Agriculture
Food security
Remote sensing
Data fusion
Smallholder systems

ABSTRACT

Crop growth simulation models are often used to estimate crop yield. For most models, this requires crop, water, and soil management information, though this information is often lacking in many regions of the world. Assimilation of Earth observation (EO) data in crop growth models can generate field-level yield estimates over large areas. The use of EO for assimilation often requires a trade-off between spatial and temporal resolution. Spatiotemporal data fusion can provide higher spatial (≤30m) and temporal resolution data to avoid this trade-off. In this study, we evaluated the timing and frequency of EO data assimilation in the Simple Algorithm for Yield Estimation (SAFY) in a persistently cloudy and fragmented agroecosystem of Ethiopia for 2019 and 2020 growing seasons. We used Landsat and MODIS data fusion to obtain frequent and spatially detailed LAI estimates and assimilated at each main maize growth stage to evaluate the effect of timing and frequency of LAI assimilation. The jointing to grain filling stage observations were more important (RMSE = 117 g/m^2, rRMSE = 16%) than other growth stages to improve yield estimation. Using LAI estimates at key crop growth stages was more influential than the frequency of LAI estimates. Reasonably accurate yield estimation (rRMSE = 20%) was obtained using the pre-peak growth stage LAI observations, suggesting that the method is suitable for in-season yield forecasting. LAI retrieval errors from EO data, particularly at the early and late growth stages, were the source of yield estimation uncertainty. Therefore, assimilation of other EO-derived biophysical variables and improving LAI retrieval accuracy from EO data could further improve crop growth model performance in smallholder agricultural systems.

1. Introduction

Timely and accurate crop growth and yield estimates contribute to national food security assessments, production estimates, and early warning systems (Hunt et al., 2019; Lobell et al., 2015). Such estimates are particularly important in regions where food insecurity is pervasive, such as sub-Saharan Africa (Elders et al., 2022). Fragmented agricultural landscapes, data uncertainty and poor model parameterization are the main challenges to estimate crop yield in many African countries (Lobell 2013). Crop growth models can provide timely and reliable yield estimation, but most models are designed for plant and field-level simulations and have many input data requirements that limit their application over larger areas (Curnel et al., 2011). EO data can be integrated into crop growth models to obtain frequent field-level information for large-area crop yield estimation (Marshall et al., 2018). However, often involve a

* Corresponding author. Department of Natural Resources, Faculty of Geo-information Science and Earth Observation (ITC), University of Twente, Hallenweg 8, 7522 NH, Enschede, the Netherlands.

E-mail addresses: b.s.tilahun@utwente.nl, biniamsisheber@gmail.com (B. Sisheber).

https://doi.org/10.1016/j.rsase.2024.101272
Received 22 October 2023; Received in revised form 29 February 2024; Accepted 11 June 2024
Available online 26 June 2024
2352-9385/© 2024 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

trade-off between temporal and spatial resolution. We outline the state of the art in crop modelling with EO data and identify knowledge gaps related to the optimal assimilation of EO time series data for yield estimation and crop yield forecasting in smallholder agriculture landscapes.

Complex crop growth models have been developed to simulate crop yield by reproducing the interaction of crop physiology and management activities (Jin et al., 2018b). These models require many inputs and parameters related to crop phenology, management, and soil characteristics to estimate crop yield, which are often unavailable for smallholder farming systems (Battude et al., 2016; Lobell et al., 2015; Meshesha and Abeje 2018). Alternatively, semi-empirical models combine the theory of light use efficiency (LUE) (Monteith 1972) and crop biophysical processes to estimate biomass and yield with few inputs and parameters (Claverie et al., 2012). The LUE-based models can also translate EO-derived biophysical parameters into total biomass based on the assumption that biomass productivity is proportional to photosynthetically active radiation to fill the large ground data requirements (Lobell 2013). The Simple Algorithm for Yield Estimates (SAFY) (Duchemin et al., 2008), for example, simplifies the process of crop growth modelling by describing the biophysical crop processes (e.g., leaf senescence, biomass accumulation) with empirical parameters to overcome limitations arising from complicated parameterization (Claverie et al., 2012).

EO data have been integrated with crop models - through data assimilation - to provide model parameters that require frequent observation and then used to estimate yield over a large area (Huang et al., 2015). Assimilation of remotely sensed biophysical variables into a crop growth model can reduce the uncertainties of varying soil, crop type and crop management and improve the accuracy of crop growth and yield estimations (Huang et al., 2019a; Jégo et al., 2012). Assimilation of EO data for large scale yield estimation in smallholder farming systems requires both high spatial and high temporal resolution EO data. Assimilation of higher spatial resolution EO data such as Landsat can offer high spatial detail but often fails to provide sufficient observations through different crop growth stages due to satellite revisit and cloud cover (Huang et al., 2019b; Whitcraft et al., 2015). MODIS data are available at daily or 8- day frequencies, but the coarse spatial resolution (≥250m) contains mixed pixel information in smallholder farming systems (Waldner et al., 2019). Assimilation using both high spatial and high temporal resolution cloud-free time series data is an ideal but often unobtainable situation (Huang et al., 2015).

Spatiotemporal data fusion that blends low spatial/high temporal with high spatial/low temporal resolution data (Zhu et al., 2010) can be used to obtain high spatial and high temporal resolution EO data for crop yield estimation. Landsat and MODIS data have been fused to capture the spatial variation of crop yield due to the spectral similarities of the sensors (Gao et al., 2018; Liao et al., 2019). Dong et al. (2016) showed that data fusion improves crop growth model simulations when the number of available high-resolution observations is limited. Previous studies in smallholder farming systems demonstrated that Landsat–MODIS fusion can detect the spatial and temporal variation of crop phenology and LAI in fragmented agricultural landscapes (Sisheber et al. 2022, 2023, 2024). While these studies have addressed the trade-off between temporal and spatial resolution, there is limited research on the effect of timing and frequency of EO data in the calibration of crop growth models to obtain optimal yield prediction in data-scarce regions (Tewes et al., 2020b).

Temporal gaps in EO data are the source of model parameterization uncertainties and yield estimation errors (Huang et al. 2015, 2019b; Jégo et al., 2012). The frequency of EO data can increase the accuracy of LAI retrieval and phenology detection from EO data, which are essential in capturing the crop growth and development (Tewes et al., 2020a). Data assimilation studies also suggest that acquiring observations at key crop growth stages is more important for optimal crop model performance (Li et al., 2017; Waldner et al., 2019; Wu et al., 2021). For instance, Wu et al. (2021) showed that observations between jointing and booting had the highest impact on WOrld FOod STudies (WOFOST) model estimation in Hebeir Province, China. Tewes et al. (2020b) stated that the assimilation of one LAI estimate at the tillering or booting to the flowering stage was sufficient for accurate yield estimation.

The effect of frequency and timing of observations also depends on the data assimilation methods (Casa et al., 2012). EO data can be assimilated using forcing, updating and calibration approaches (Huang et al., 2019a). In the forcing and updating approaches, EO-derived state variables such as LAI replace model-simulated LAI based on the assumption that a better simulation at the date of data assimilation will also improve the model simulation for the subsequent date's yield simulation (Jin et al., 2018b; Li et al., 2017). Although the calibration approach is preferred because it minimizes model simulated and EO estimation LAI errors, the optimal frequency and timing are less clear since the EO-derived LAI is used to optimize model parameters rather than directly replacing model simulated values (Curnel et al., 2011). Therefore, evaluating the optimal timing and frequency of observations to calibrate crop growth models needs investigation in regions where persistent cloud cover obscures the surface from remote sensing.

The objective of this study was to explore the optimal timing and frequency of optical EO to improve maize yield estimation in a semi-empirical crop growth model (SAFY) using Landsat and MODIS fused time series observations in Ethiopia. The study analysed the uncertainty of crop model simulation due to the temporal resolution of the input data used for assimilation. We optimised SAFY using Landsat–MODIS fused LAI at 8-, 16-, 24- and 32-day intervals to explore the effect of frequency on yield simulation. The timing and distribution of EO data for assimilation were subsequently analysed by assimilating LAI at different maize phenological stages.

Understanding the optimal frequency and timing of EO data required by the crop growth model can be used for pre-and post-harvest yield estimation over growing seasons and over large areas, which is essential for food security related applications in data-scarce smallholder agricultural systems.

2. Materials and methods

2.1. Study area

The study covers four districts surrounding the Lake Tana sub-basin in Ethiopia (Fig. 1). The elevation ranges from 1600 to 3500 m above mean sea level. The mean annual average temperature and cumulative precipitation range from 13 °C to 22 °C and

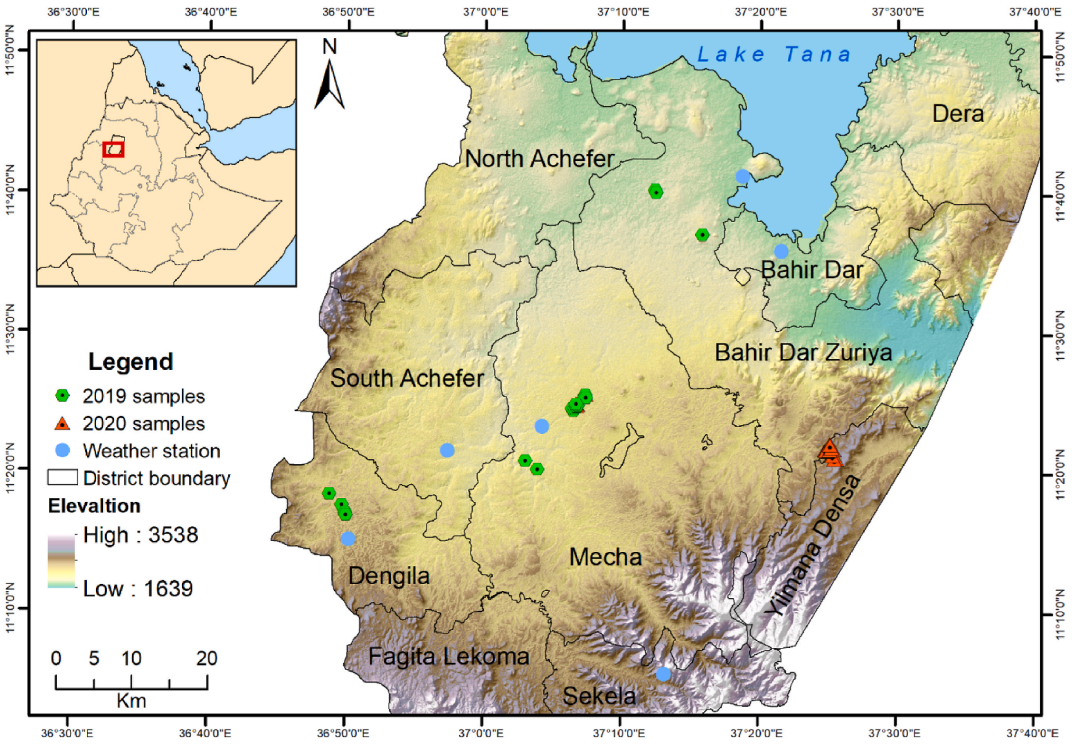

Fig. 1. Map of the study area showing the distribution of sample points.

970 mm–1900 mm, respectively. The main rainy season (locally called *kiremt,* or summer) occurs between June and September. It supports the main (*meher)* cropping season. The northward movement of the inter-tropical convergent zone (ITCZ) drives *kiremt* rainfall. A short and highly variable rainfall period (*belg* or spring) also occurs between March and May, while October to January (*bega* or winter) is the dry season that also influences *meher* crop production (Gummadi et al., 2018). Maize is the primary crop in these districts and is sown between June and early August and harvested between November and mid-December (see Fig. 3).

2.2. Description of the crop growth model

Complex crop models such as WOFOST and the Decision Support System for Agrotechnology Transfer (DSSAT) require large inputs and parameters, which are often difficult to obtain (Dong et al., 2020; Huang et al., 2019a). The simple models calculate biomass as an empirical sum of vegetation indices derived from EOs and are uncomplicated to parameterize, but they are less suited for crop monitoring because they do not account for crop type and management. However, semi-empirical models that combine the LUE theory with a simulation of the successive plant phenological stages gather the descriptions of the main biophysical processes and empirical parameterizations (Claverie et al., 2012; Lobell 2013). These models are becoming more popular due to their ability to use remotely sensed biophysical variables (Marshall et al., 2018; Silvestro et al., 2017).

In this study, we used the SAFY model because it has few input requirements and because it can be coupled with EO to estimate yield over a large area (Claverie et al., 2012; Dong et al., 2016; Silvestro et al., 2017). SAFY is a semi-empirical model designed to estimate the dynamics of Leaf area index (LAI), dry aboveground biomass (DAM), and yield at daily time steps from the day of crop emergence (D_0) to the day of senescence. The SAFY model simplifies crop nutrition and water stress into one effective light-use efficiency (ELUE) parameter. The daily incoming shortwave radiation (Rg), climate efficiency (εC), the fraction of absorbed photosynthetically active radiation (fAPAR), daily average air temperature (Ta) and ELUE are then used as the input to simulate DAM (eq. (1)). The final grain yield is determined as a function of accumulated DAM production and the grain partitioning (Py) parameters, similar to the harvest index (eq. (2)) (Duchemin et al., 2008).

$$DAM = Rg * \varepsilon C * (1 - \exp(-k * LAI)) * ELUE * Ft \quad (1)$$

$$Y = DAMmax \times Py \quad (2)$$

In SAFY, LAI is the only state variable that determines the light interception capability of crops and the daily change in the ΔDAM. During the vegetative stage, LAI tends to increase (eq. (4)) as the fraction of daily ΔDAM partitioned to leaf growth and the leaf partitioning function driven by the specific leaf area of crops (SLA) (Pl: Eq (3)). LAI decreases (ΔLAI-) (eq. (5)) during the senescence stages when the daily sum of temperature (SMT) reaches a certain threshold (S_{TT}: sum of temperature for senescence) controlled by the rate of senescence (Rs) (Claverie et al., 2012; Duchemin et al., 2008).

The small number of parameters and easy coupling with EO data makes SAFY an attractive option for large-area crop yield estimation. LAI derived from high spatial and temporal resolution EO images has been successfully assimilated with SAFY to estimate crop yield in various crop-growing environments (Battude et al., 2016; Claverie et al., 2012; Liao et al., 2019). We used the original SAFY model with 15 parameters to estimate maize LAI and yield to keep the model's simplicity (Table 1). Fig. 2 shows the overall workflow of the methodology used to evaluate the influence of LAI input timing and frequency in SAFY model accuracy.

Table 1
Initial and calibrated values of SAFY parameters.

Parameters	Unit	Maize	Source
		Range or value	
Inputs			
Daily incoming global variation (Rg)	MJ m^{-2} d^{-1}	–	METRA2 and station
Daily mean air temperature (Ta)	°C	–	Weather stations
Fixed parameters			
Initial LAI	$m^2.m^{-2}$	0.1	LAI estimated
Initial dry aboveground mass (DAM_0)	kg ha^{-1}	4.0	Claverie et al. (2012)
Climatic efficiency (εC)	–	0.5	In situ measurement
Light-interception coefficient (K)		0.5	In situ measurement
Temperature for crop growth (T_{min}, T_{opt}, Tmax)	°C	10,30,45	Srivastava et al. (2019)
Polynomial degree (β)	–	2	Claverie et al. (2012)
Specific leaf area (SLA)	m^2 g^{-1}	0.024	Claverie et al. (2012)
Harvest index (HI)	–	0.5	Srivastava et al. (2019)
Input parameters			
Emergence date (D_0)	DOY	[170,190]	Phenology detection
Sum of temperature for senescence (S_{TT})	°C	[1365,1920]	Ta and phenology detection
End of simulation (EOS)	DOY		Phenology detection
Calibrated parameters			
Partition-to-leaf function: parm a (Pla)	–	[0.1,0.5]	Calibrated
Partition-to-leaf function: parm b (Plb)	–	[0.0007,0.002]	Calibrated
Rate of senescence (Rs)	°C d^{-1}	[3000, 10000]	Calibrated
Effective light use efficiency (ELUE)	g MJ^{-1}	[1.5, 4]	Calibrated

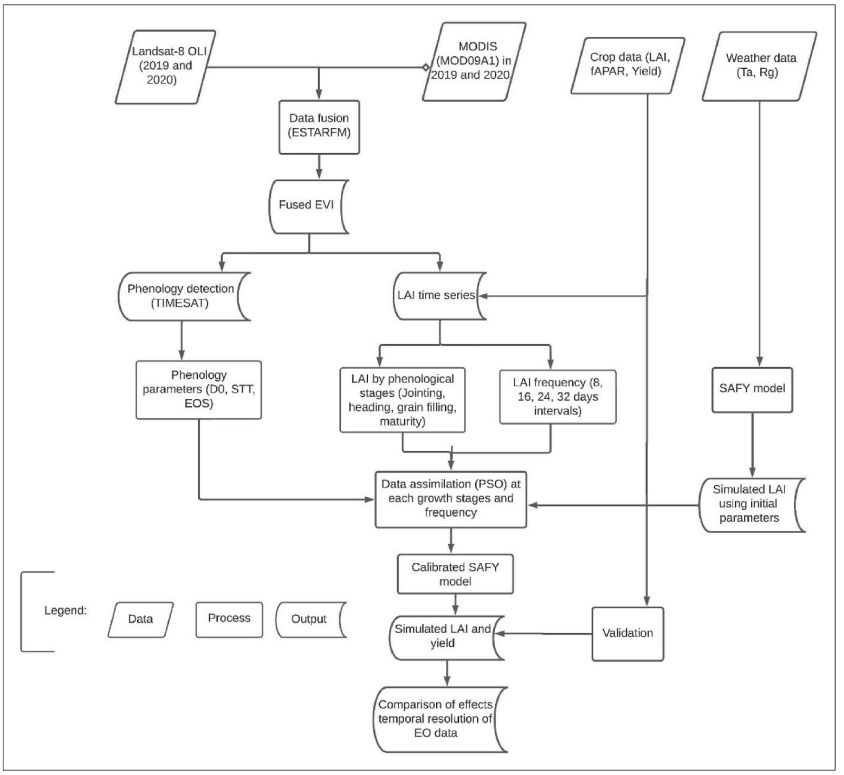

Fig. 2. Flow chart of the methodology.

$$Pl\left(\sum Ta\right) = 1 - Pla \times e^{plb \times STM} \tag{3}$$

$$\Delta LAI+ = \Delta DAM \times Pl\left(\sum Ta\right) \times SLA, Pl > 0 \tag{4}$$

$$\Delta LAI- = LAI \times (SMT - STT)/Rs), SMT > STT \tag{5}$$

2.3. Field data acquisition

In order to validate the SAFY model, field measurement were conducted in four districts in the study area during the 2019 and 2020 growing seasons between July and December. Fig. 1 shows the spatial distribution of the sample fields. LAI, fAPAR, and crop yield samples were collected from 35 maize fields. LAI and fAPAR were collected two to four times during the two growing seasons (Table 2) using Decagon's AccuPAR ceptometer®. A total of 64 maize LAI samples were measured at the vegetative, reproductive, and maturity growth stages. LAI and fAPAR were sampled over 60 × 60 m^2 elementary sampling units (ESUs) to account for possible geo-location errors between the field measurements and Landsat pixels. Within each ESU, 10 LAI readings (averaged from multiple positions within the canopy) were recorded in 1 × 1 m^2 quadrats. The results were averaged to produce one LAI reading for each ESU. We also collected 46 maize fAPAR samples during the 2019 growing season. The fAPAR value was calculated from PAR readings transmitted through the canopy, absorbed or reflected at the soil surface, and reflected by the canopy as specified in AccuPAR device manual.

Crop yield was collected using destructive and non-destructive methods. During the 2019 growing season, we used the yield component method to harvest maize ears from 17 farm fields and counted the number of ears per 5 m^2 area to account for field-level spatial variation, the number of kernel rows per ear, and the number of kernels per row. We dried the maize ear samples at a temperature of 80^0C for 72 h and weighed using a sensitive balance to obtain grams yield per square meter. In 2020, we collected 20 maize yields from farmers' reports due to COVID-19 restrictions to conduct field campaigns.

Daily incoming shortwave radiation (Rg) and mean daily temperature (Ta) are the two climate-forcing variables in SAFY. A weather station was installed for the purpose of this study in the Mecha district (Fig. 1) to obtain daily Rg and temperature for the nearby samples. To account for spatial variability, Rg was extracted from NASA's Modern-Era Retrospective Analysis for Research and Applications (MERRA-2) (http://disc.sci.gsfc.nasa.gov), which is a reanalysis product with a spatial resolution of 0.5° latitude by 0.67° longitude. Temperature records were also obtained from the National Metrological Agency of Ethiopia (NMA).

A crop-type map classified in 2019 from Landsat-8 (Sisheber et al., 2022) was used to mask maize fields for field-level crop yield simulation. The classified map discriminated maize from other land cover units with an overall accuracy of 82.9%, a kappa coefficient of 0.80, a user's accuracy of 79% and a producer's accuracy of 78%. We used crop calendar information and EVI empirical thresholds similar to Bolton and Friedl (2013) to minimize missclassification and exclude other crop types from analysis. On cropland pixels if the green up EVI value in late June and early July is greater than 0.25 and reached peak value before the end of September (EVI > 0.6) was classified as maize following ground sowing date and LAI observation. In this way we identified 2120 km^2 of maize growing area for the pixel-level yield simulation.

2.4. Satellite imagery

Level-2 terrain corrected Landsat-8 data were acquired from the United States Geological Survey (USGS) Earth Explorer (https://earthexplorer.usgs.gov) for path/row: 70/52 in 2019 and 2020. The images were atmospherically corrected using the Landsat-8 Surface Reflectance Code (LaSRC), to improve aerosol determination (Roy et al., 2016). We obtained low cloud cover (≤30%) Landsat images for data fusion and for validation (30–70%), distributed across the major phenologial stages of maize from land preparation (May) to harvest (December) to capture the temporal variation in reflectance during data fusion (Fig. 3).

MODIS surface reflectance data (MOD09A1:V006.1) was acquired from NASA's LP DAAC website (https://lpdaac.usgs.gov) for image tile: 21/7. The product consists of atmospheric and aerosol-corrected surface reflectance spectral bands at 500m spatial resolution composited over 8-day intervals. We resampled the MODIS data to WGS84 UTM zone 37 and 30m with nearest neighbour to match the Landsat projection and resolution.

We used six of the Landsat and MODIS bands (blue, green, red, NIR, SWIR1, and SWIR2) for data fusion since using many bands improves feature separability during fusion (Gao et al., 2015). The Landsat pixel quality and MODIS state flag bands were used to flag low-quality pixels.

Table 2
LAI, fAPAR, and yield measurements in maize fields during the 2019 and 2020 growing season.

Observation date	Sampling site	Growth stage	LAI samples	Mean LAI (m^2/m^2)	fAPAR samples	Mean fAPAR	Yield (g/m^2)
July 16 to25, 2019	Bahir Dar Zuria,	Green up	16	0.58 ± 0.3	16	0.23 ± 0.10	781 ± 130
Aug 28–30, 2019	Mecha, Dangla	Jointing	17	2.75 ± 1.0	17	0.67 ± 0.11	
Sep 16–25, 2019		Heading	17	3.72 ± 1.0	17	0.77 ± 0.07	
Nov 11, 2019		Maturity	6	2.26 ± 1.0	–	–	
Aug 12–19, 2020	Yilmana Densa	Jointing	4	1.75 ± 0.4			454 ± 135
Oct 07, 2020		Heading	4	3.25 ± 0.9	–	–	
Nov 09–13, 2020	Mecha	Maturity	2	2.51 ± 10	–	–	

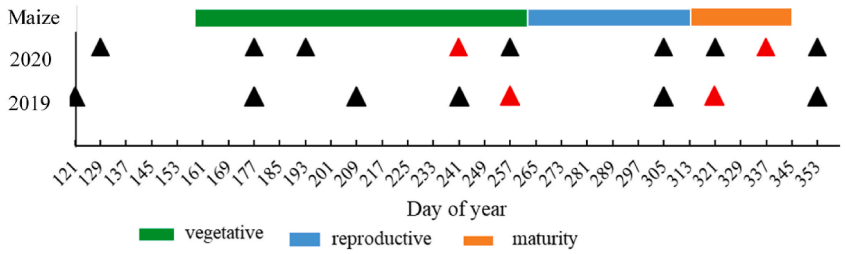

Fig. 3. Distribution of relatively cloud-free Landsat images during the main maize growth stages. Black squares are images used for data fusion, and the red triangles are images used for validation.

2.5. Analysis

2.5.1. Landsat-MODIS fusion

We used ESTARFM (Zhu et al., 2010), modified for fragmented agricultural landscapes (Sisheber et al., 2022) to fuse Landsat-8 OLI and MODIS during the 2019 and 2020 growing seasons to evaluate the importance of the timing and frequency of EO data assimilation in crop models. ESTARFM uses pairs of Landsat and MODIS images at base dates (t_0) before (t_1) and after (t_2) the prediction dates (t_p) and MODIS images at t_p to predict Landsat–MODIS fused data. In the modified ESTARFM we distributed the Landsat inputs based on the maize crop calendar to minimize the effect of reflectance change between t_0 and t_p due to the lack of frequent cloud-free Landsat inputs. When frequent cloud-free Landsat inputs were available, we selected Landsat and MODIS inputs with similar acquisition dates. When the available cloud-free Landsat inputs were at different phenological stages, we used the MODIS composite that was best correlated with MODIS input at t_p and with Landsat at the base dates (MODIS t_0-t_p correlation) to compensate for phenology change. Then, we searched for similar pixels based on the intersection of Landsat reflectance at t_1 and t_2 within a moving window (w) of 50 × 50 Landsat pixels. We used a land cover map of the study area (Sisheber et al., 2022) as input to select similar pixels from the homogeneous land cover class (c). Finally, the spatial weighting function (w_i) and conversion coefficient (v_i) were calculated based on the original ESTARFM. A detailed description of the model is found in Sisheber et al. (2022) and Zhu et al. (2010).

The enhanced vegetation index (EVI) (Huete et al., 2002) was calculated from the Landsat–MODIS fused data to estimate LAI and determine crop phenology. We used EVI rather than NDVI because it does not saturate in dense canopies (Yang et al., 2020). Validation of data fusion comparing EVI calculated from data fusion and the original Landsat image is presented in the supplement material (Fig. S1).

2.5.2. Phenology detection

Phenological parameters constitute an important part of the SAFY model. We determined phenological parameters from the fused data and used them as input variables in the model to the capture spatial variability of crop growth in smallholder systems. Maize phenology was determined from the EVI time series using a Savtizky–Golay (S-G) smoothing and seasonal amplitude (dynamic threshold) method in the TIMESAT software package (Eklundh and Jönsson 2017). TIMESAT was used to extract key phenological dates from the EVI time-series as required by SAFY: start of season (SOS), representing the emergence date; the peak of season (POS), used to determine the beginning of the day of senescence (DOS); and the end of season (EOS), which is the termination of the model simulation. The phenological dates were used to determine the field-level input parameters of SAFY (D_0, S_{TT} and EOS). SOS replaced the $D_{0,}$ and the mean daily temperature aggregates between D_0 and DOS were used to determine the S_{TT}. EOS was adjusted to correspond to the end of maturity date and used as the termination date for the simulation. An example of a maize phenological cycle determined from data fusion for one of the sample fields (Fig. S2) and phenology detection accuracy (Table S1) are presented in the supplemental material.

2.5.3. LAI and fAPAR estimation

LAI and fAPAR were estimated from the S-G smoothed EVI time series to minimize distortion from cloud, data fusion uncertainty and mixed pixel problems. LAI is the most widely used crop model state variable as it is a good indicator of crop growth status and is closely linked to crop biomass and yield (Casa et al., 2012; Huang et al., 2015). An empirical relationship between EVI and in-situ LAI - measured within 2–4 days of the data fusion or MODIS acquisition dates was formed to estimate pixel-level LAI. Since empirical models lack transferability in space and time, we fitted different models for the 2019 and 2020 growing seasons (Fig. 4). To reduce the LAI estimation error, LAI values higher than 0.1 (emergence date value) and lower than 7 (peak value) were selected for data assimilation. The temporal LAI progression from the green up to maturity was obtained. fAPAR values were estimated from field-measured fAPAR and LAI following the Beer–Lambert law to determine the light interception efficiency (k). The coefficient of determination (R^2), the root mean square error (RMSE), and the relative root mean square error (rRMSE) were used to fit the models and assess the robustness of the LAI and fAPAR estimations.

2.5.4. Data assimilation in the SAFY model

We assimilated LAI derived from Landsat–MODIS fused data in SAFY at different phenological stages and with different temporal frequencies to assess the impact of timing and temporal resolution of EO data. The temporal resolution-dependent phenological parameters (D_0, S_{TT}, and EOS) determined from crop phenology detection served as input parameters to reduce the model bias caused by generic crop calendar information. The fixed model parameters were determined from in situ measurements, the literature, and cali-

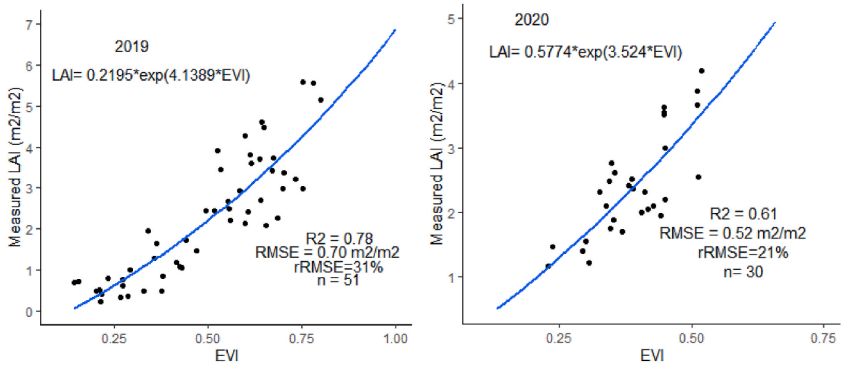

Fig. 4. Maize LAI estimation model for the 2019 and 2020 growing seasons.

brated from data fusion retrieved LAI (Table 1). The remaining sensitive parameters of SAFY were assimilated using the calibration approach because calibration minimizes the error of EO data in crop growth models (Jin et al., 2018b). In order to select the proper time scale and input frequency for optimal model accuracy, we tested the calibration of the sensitive parameters (ELUE, Pla, Plb, and R_S) identified in previous studies Sisheber et al. (2024) with LAI obtained in different maize phenological stages. The environment and crop-specific parameters (Pla, Plb, and R_S) determining the process of leaf growth and the ELUE were calibrated to evaluate the effect of LAI frequency and timing on the model performance.

To explore the effect of the timing of LAI on the model calibration, we grouped the input LAI by phenological stages based on the phenometrics derived from EVI time series and crop calendars of the dominant local maize variety (Firezer 2019). We calibrated the model with the full LAI time series from green up to maturity and by excluding one growth stage at a time, including jointing, heading, grain filling, and maturity, to assess the contribution of the growth stages to the final yield estimation. LAI frequency at 8-day, 16-day, 24-day, and 32-day intervals between D_0 and EOS was also assimilated to analyse the model response with different temporal resolutions. These temporal frequencies were tested in the four districts (Fig. 1) during the 2019 and 2020 growing seasons.

The particle swarm optimization (PSO) algorithm was used for calibration for its simplicity, precision, and computational efficiency at pixel-level (Silvestro et al., 2017). PSO assumes a group of m particles ($m = 30$ based on the previous literature) with certain speeds without quality and size in n d-dimensional space. The particles modify their position and velocity based on the best point in the current generation ("personal best") and the best point of all particles in the swarm ("global best"). To optimize the calibrated model parameters, we applied a cost function (eq. (6)) representing the minimum RMSE between data fusion retrieved LAI (F_{LAI}) and SAFY simulated LAI (S_{LAI}). The optimization was deemed successful when the cost function approximated the least error. After optimization, the crop model was run to simulate maize LAI and yield.

$$C = \sqrt{\sum_{i=1}^{N}\left(\frac{\left(\frac{S_{LAI}-F_{LAI}}{F_{LAI}}\right)^2}{n}\right)} \tag{6}$$

2.6. Accuracy evaluation

The effect of frequency and timing of LAI observation was evaluated by comparing simulated LAI and yield with measured LAI and yield data. We used bias, which calculates the difference between observed and predicted value, RMSE and relative RMSE (rRMSE) measuring the normalized difference between observed and predicted yield and the coefficient of determination (R^2), which represents the proportion of variation between simulated LAI and yield compared to measured LAI and yield. The change in the value of calibrated parameters and the simulated LAI and yield after optimization were examined to evaluate the contributions of different phenological stages assimilations on SAFY model simulations. Since weather conditions and crop management variations strongly influence crop yield, we compared the model accuracy across the 2019 and 2020 growing seasons.

3. Results

3.1. Maize LAI retrieval from data fusion

Maize LAI estimated from Landsat–MODIS data fusion for the 2019 and 2020 growing seasons is presented in Fig. 4. An exponential model estimated maize LAI with moderate to high accuracy over the two growing seasons. The correlation coefficient was higher in 2019 ($R^2 = 0.78$) than in 2020 ($R^2 = 0.61$), but the prediction error was higher in 2019 (RMSE = 0.70 m^2/m^2, rRMSE = 31%) than in 2020 (RMSE = 0.52 m^2/m^2, rRMSE = 21%). The model estimated the annual dynamics of maize LAI from Landsat–MODIS data fusion (supplement, Fig. S2) for assimilation in the SAFY crop growth model.

3.2. SAFY model calibration and LAI simulation performance

Table 3 presents the calibrated model parameters in different phenological stages and the calibration accuracy comparing simulated LAI with data fusion retrieved LAI. Calibration using the full LAI time series from green up to maturity (DOY 185 to 313) provided the best model LAI simulation accuracy (RMSE = 0.47 m^2/m^2, rRMSE = 17% and R^2 = 0.91). When comparing the effects of phenological stages, the jointing to grain filling stage calibration (DOY 217 to DOY 289) slightly reduced the value of ELUE and increased R_S, which decreased the model accuracy compared to calibration using the full LAI time series.

Among the phenological stages, the jointing to grain filling stages were the best estimators of LAI (RMSE = 0.57 m^2/m^2, rRMSE = 23% and R^2 = 0.87) compared to the data fusion retrieved LAI. In the green up to peak growth stage calibration, the senescence stage parameter (R_S) increased, and the agronomic parameter (ELUE) decreased to reduce model accuracy for LAI simulation (RMSE = 0.73 m^2/m^2, rRMSE = 25% and R^2 = 0.83) compared to the jointing to grain filling calibration. Excluding the vegetative stage observation in the peak to maturity stage calibration (DOY 265 to DOY 313) resulted in the highest model prediction error (RMSE = 0.91 m^2/m^2, rRMSE = 32% and R^2 = 0.78), showing the importance of acquiring vegetative stage observations for model calibration. Using only pre- or post-peak LAI resulted in weaker model accuracy since the model could not capture the annual dynamics of LAI during the growing season. The peak-to-maturity stage calibration estimated a higher ELUE and the lowest Rs parameter and vice-versa for green up-to-peak stage calibration. A single observation at each growth stage provided relatively reasonable model accuracy (RMSE = 0.61 m^2/m^2, rRMSE = 26%, R^2 = 0.86), showing the importance of a distribution of observations across the growing season.

Fig. 5 shows the validation of SAFY-simulated LAI using field-measured maize LAI. Overall, the model performed well in terms of representing the temporal variability of simulated LAI during the growing season. The deviation between simulated and field-observed LAI varied with crop growth stage in all assimilation cases. The error of simulated LAI compared with field-measured LAI was, however, higher than that between simulated and data fusion estimated LAI (Table 3) in all assimilation cases since the model was calibrated using data fusion estimated LAI rather than field-measured values. Overall, LAI simulated using the full LAI time series (a) was more consistent with field-measured LAI than the assimilation of specific phenological stage observations (b-e). The full growing season calibration minimized the bias between the field and simulated LAI in all the phenological stages.

SAFY LAI estimation performance was weaker when the model was calibrated using reproductive (R^2 = 0.56, RMSE = 1.19 m^2/m^2 and rRMSE = 46%) and vegetative (R^2 = 0.61, RMSE = 0.91 m^2/m^2 and rRMSE = 36%) stage LAI observation alone. Calibration that included both vegetative and reproductive stages (a, b, and d) minimized LAI estimation error compared to field observed values because the phenological parameters (Pla, Plb and R_S) and the ELUE depend on crop growth dynamics throughout the growing season. Jointing to grain filling stage LAI observations (b) (R^2 = 0.72, RMSE = 0.83 m^2/m^2, rRMSE = 32%) provided the best model performance among phenological stages. The early vegetative and maturity stage LAI was not well simulated when the model was calibrated with only vegetative or reproductive stage LAI input (c-d). The bias in the maturity stage LAI was minimized when calibrated using the peak-to-maturity date LAI assimilation, while peak LAI was best simulated in all phenological stage calibrations.

3.3. The timing of LAI data assimilation on maize yield estimation

Table 4 and Fig. 6 show the validation of yield estimated using the five phenological stage assimilations. Consistent with the LAI simulation performance (Fig. 5), the full LAI time series during the growing season provided the best estimate of maize yield (R^2 = 0.77, RMSE = 105 g/m^2 and rRMSE = 14.8%). Among the phenological stages, jointing to grain filling observations provided the highest correlation coefficient (R^2 = 0.68) between observed and predicted yield and a smaller yield prediction error (RMSE = 117 g/m^2 and rRMSE = 16.4%). The small difference between the full-time series and jointing to grain filling period calibrations indicates that early vegetative and maturity stage observations are less critical for yield estimation. Calibration using peak to maturity (RMSE = 214 g/m^2, rRMSE = 30.2%) and green up to peak growth stages LAI (RMSE = 140 g/m^2, rRMSE = 19.7%) increased yield prediction error compared to the jointing stage calibration. Using a single observation at each growth stage provided reasonably good yield prediction (R^2 = 0.57, RMSE = 139 g/m^2, rRMSE = 20%) with few observations (n = 5), suggesting that the distribution of inputs was more important than the frequency of inputs to capture crop growth dynamics during the growing season.

As shown in Fig. 6, the model overestimated yield when calibrated using peak-to-maturity stage LAI and underestimated yield when the calibration inputs were only from the vegetative stage. As also shown in Table 3, calibration using only the post-peak period increased ELUE and decreased the R_S parameters that coincided with the yield overestimation and vice-versa for pre-peak stage cali-

Table 3
Effect of calibration of SAFY parameters at different maize phenological stages for the 35 maize sample fields during the 2019 and 2020 growing season.

Calibration periods	#LAI	Parameters				LAI error		R^2
		Pla	Plb	RS	ELUE	RMSE	rRMSE (%)	
Green up to maturity	17	0.375	0.0015	7500	2.796	0.47	17	0.91
Jointing to grain filling	10	0.370	0.0016	7650	2.729	0.57	23	0.87
Green up to heading	9	0.375	0.0016	8000	2.655	0.73	25	0.83
Peak to maturity	10	0.370	0.0016	7300	2.988	0.91	32	0.78
One observation at each stage	5	0.380	0.00155	7420	2.787	0.61	26	0.86

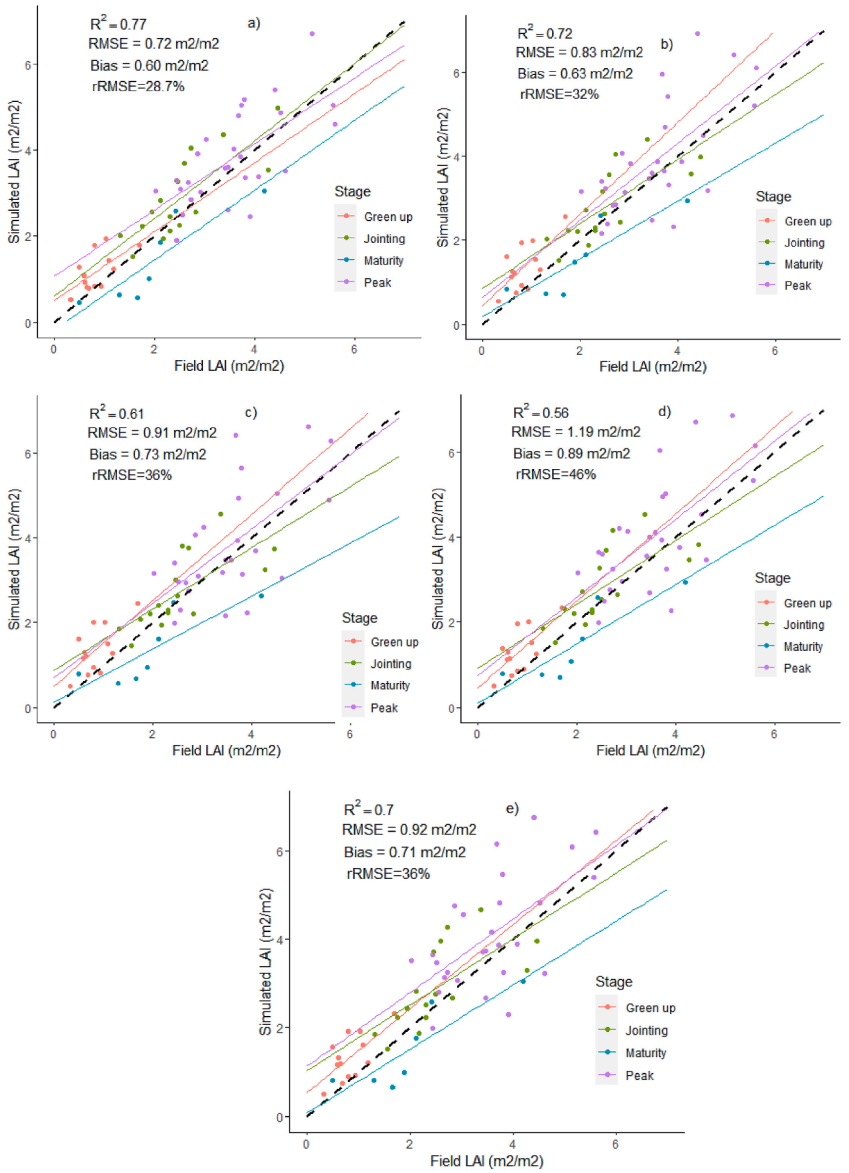

Fig. 5. Comparison of field measured LAI and LAI simulated under different phenological stage assimilation: a) green up to maturity, b) jointing to grain filling, c) green up to peak, d) peak to maturity, e) one observation at each growth stage.

Table 4
Validation of maize yield simulations based on different phenological stage LAI assimilation for 35 maize sample fields in 2019 and 2020.

Calibration periods	#LAI	Yield estimation error			R^2
		RMSE (g/m2)	rRMSE (%)	Bias (g/m2)	
Green up to maturity	19	105	14.8	87	0.77
Jointing to grain filling	13	117	16.4	88	0.68
Green up to peak	10	140	19.7	109	0.53
Peak to maturity	9	214	30.2	172	0.35
One observation at each stage	5	139	20.0	108	0.57

bration. In comparison, the pre-peak LAI assimilation (green up to peak) better estimated crop yield than post-peak (peak to maturity) stage assimilations. Reasonably accurate yield estimation using the pre-peak period LAI observations and the smaller significance of maturity stage observations suggest that the model could be used for pre-harvest yield forecasting with about 19.7% error (rRMSE) from the actual yield at the end of the season.

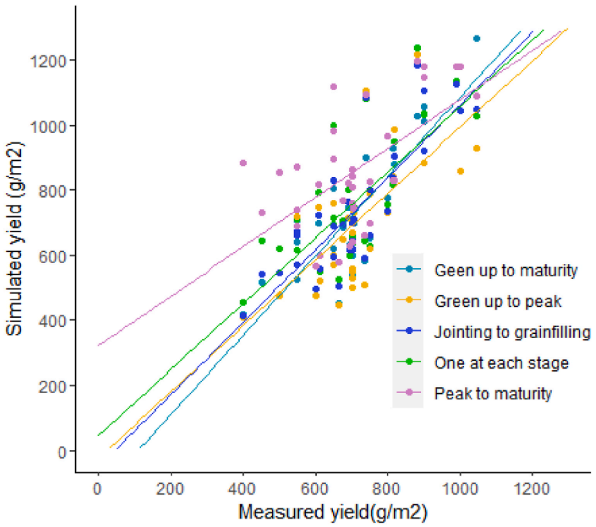

Fig. 6. Comparison of measured and simulated maize yield based on assimilation of different phenological stage LAI observations in 2019 and 2020 growing seasons.

3.4. Effect of LAI frequency on yield estimation

Fig. 7 shows that SAFY simulated LAI was sensitive to the frequency of the LAI input used for calibration. The increase in LAI interval from 8 days to 16 days increased the mean LAI prediction error (from rRMSE = 20% to rRMSE = 24%). The relative mean LAI error of the sample fields also increased (from rRMSE = 24–27%) when the interval of LAI observation increased from 24 days to 32 days. However, Table 5 shows that the effect of LAI input frequency on the final yield prediction was smaller than on LAI simulation since the optimization procedure minimizes the difference between observed and predicted LAI. Maize yield was best estimated using 8-days LAI frequency, beyond which the correlation coefficient gradually decreased and the error between observed and predicted yield increased. Yield estimation errors of the 16-day, 24-day, and 32-day intervals were comparable, suggesting that the distribution and quality of LAI inputs were more important than the number of observations at particular crop growth stages consistent with Fig. 5e.

3.5. Annual and spatial variation of yield estimation

We analysed the annual variation of SAFY crop yield estimates using the full LAI time-series calibration (Fig. 8) and the model parameters (Table 6) in the 2019 and 2020 growing seasons. SAFY was able to capture the annual variation of maize yield in the study

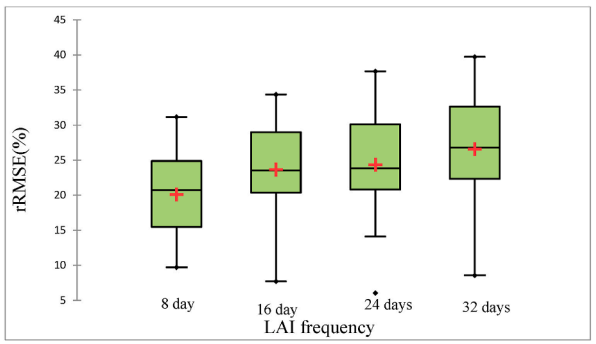

Fig. 7. LAI calibration error (rRMSE) of 35 maize sample fields in 2019 and 2020 at 8-day, 16-day, 24-day and 32-day frequency.

Table 5
Accuracy of maize LAI and yield estimation using different phenological stages LAI assimilations compared with field-measured yield.

LAI frequency	#LAI	Mean yield (g/m^2)	Yield estimation error			R^2
			RMSE (g/m^2)	Bias (g/m^2)	rRMSE (%)	
8 days	19	732 ± 206	105	87	14.8	0.77
16 days	9	738 ± 191	124	100	17.6	0.60
24 days	6	735 ± 197	127	103	18.1	0.59
32 days	5	745 ± 209	123	105	19.0	0.57

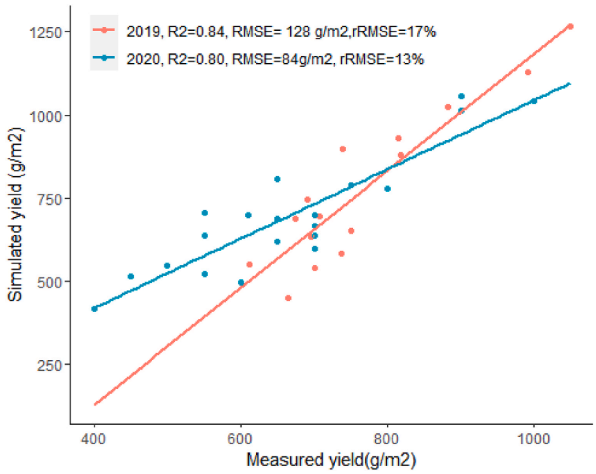

Fig. 8. Comparison between measured and simulated yield in maize using the field for the 2019 and 2020 growing seasons.

Table 6
Annual variation of SAFY parameters between 2019 and 2020 growing seasons.

Parameters	2019			2020		
	Max	Min	Mean	Max	Min	Mean
D_0	185	169	175	189	160	174
S_{TT}	1091	562	681	762	531	656
Pla	0.375	0.375	0.375	0.375	0.35	0.37
Plb	0.0015	0.0015	0.0015	0.0015	0.0015	0.0015
R_S	6650	6650	6650	7500	6500	7000
ELUE	3.6	2.2	2.8	3.75	2.1	2.8
Ta	22.5	15.1	18.2	20.1	15.5	17.5
Rg	28.5	6.5	18.2	23.3	5.3	17.4

area. The mean annual maize yield was higher in 2019 (778 ± 236 g/m^2) than in 2020 (627 ± 160 g/m^2). SAFY underestimated maize yield in the 2019 prediction, and the underestimation decreased for the 2020 growing season. The 2019 prediction had a slightly higher R^2 (0.84) than the 2020 (R^2 = 0.80) growing season, but the RMSE and relative RMSE (rRMSE) were lower for 2020 than 2019, consistent with the LAI retrieval accuracy in Fig. 4.

Table 6 shows that the crop-specific parameters (Pla, Plb and R_S) were invariant between the 2019 and 2020 growing seasons. Thus, in addition to the fixed parameters of SAFY, the crop-specific parameters can be transferable once calibrated for the study area and crop type. However, the time-dependent parameters (D_0 and S_{TT}) and the ELUE showed annual variation. These parameters require calibration every growing season and can be calibrated successfully using EO-derived time series LAI between the jointing and grain filling stages. Therefore, inter-annual yield estimation depends on the availability of climate input data (Ta and Rg) and EO estimated LAI.

Fig. 9 shows the spatial distribution of yield estimation based on the 2019 maize crop mask map. The mean maize yield based on the assimilation of the full-LAI time series was 683 ± 111 g/m^2. Maize yields varied spatially. For instance, the maize dominant location in the mechanized farming site in Koga had the highest yield value (≥700 g/m^2) whereas the highlands such as south Mecha and Yielmana Densa, and the mixed crop growing locations of Bahir Dar Zuriya and Dera districts, where fields are small had lower yields (≤600 g/m^2). The zoom windows show higher pixel-level variability in the major maize-growing site.

4. Discussion

EO data assimilation in crop growth model is increasingly used to estimate crop yield over large areas, which is important for early warning, production distribution, and food security assessments in such regions. However, persistent cloud cover, satellite revisit periods and mixed pixels are challenges to meet the spatial and temporal resolution input requirements of crop models from a single open-source sensor in smallholder agricultural systems (Whitcraft et al., 2015). This study evaluated the effects of the timing and frequency of EO data obtained from Landsat and MODIS fusion on maize yield estimation in Ethiopia. The main findings of the study include: (1) the full LAI time series during the growing season provided the best yield estimations when compared to calibration using LAI at a particular phenological stage; (2) reasonably accurate yield estimation using only the pre-peak stage LAI observations can be useful for in-season yield foresting; (3) the temporal distribution and accuracy of the EO retrieved LAI was more critical than the frequency of LAI in improving yield estimation and; (4) calibration of SAFY using frequent LAI observations contributed to capturing an-

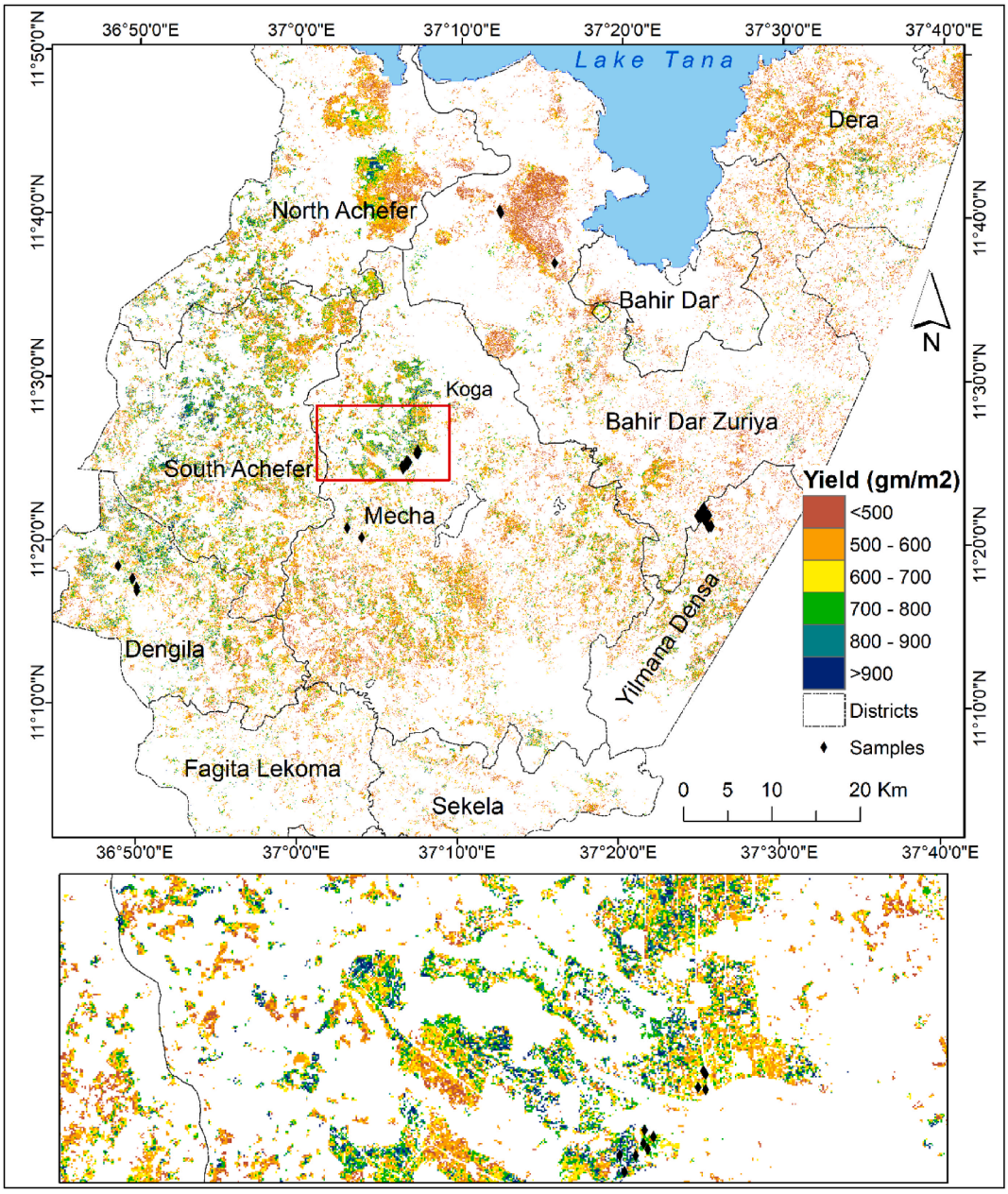

Fig. 9. Maize yield map using the full LAI time series assimilation in 2019 growing season.

nual variability of yield estimates. Our result revealed that high temporal resolution data obtained through Landsat–MODIS fusion improves yield estimation accuracy in persistently cloud-growing seasons in smallholder agricultural systems.

4.1. Timing of LAI assimilation on crop yield estimation

The best yield estimates were attained when SAFY was calibrated using LAI time series covering the full growing season (R^2 = 0.77, rRMSE = 15%) rather than calibrating at a particular phenological stage. Our result agrees with previous studies, which indicated the importance of continuous time series of EO data for improved yield estimation (Dong et al., 2016; Huang et al., 2015; Hunt et al., 2019). Temporal gaps in EO data are the source of uncertainties in model parameterization resulting in larger errors in crop yield estimates (Huang et al., 2019b; Jégo et al., 2012). A time series of LAI observations during the growing season minimized uncertainty in parameterization of the phenological parameters (D_0, Pla, Plb and R_S) and ELUE, which depend on crop growth dynamics over the growing season. The accuracy of these parameter extractions depends on the timing and frequency of time series EO data, which cannot be obtained from a single open-source optical sensor in the study area. In this regard, Jégo et al. (2012) and Dong et al. (2016) indicated that crop growth models were not well reinitialized when the Landsat-derived LAI was not well distributed across

the growing season, leading to large errors in crop yield estimation. Our study showed that multi-sensor data fusion offers frequent high spatial resolution observations to determine time-dependent model parameters and improved yield estimation. Recent studies have shown the potential application of synthetic aperture radar (SAR) in crop yield estimation because of its capability to provide all-day and all-weather observations (Wu et al., 2021). Future studies integrating optical and radar data can provide valuable input for yield estimation in smallholder agricultural systems.

LAI observations from the jointing to the grain filling stage increased the robustness of yield estimation (R^2 = 0.68, rRMSE = 16%) compared to the contributions from assimilation in other growth stages. LAI observations from the jointing and grain-filling stage were important to detect the various factors influencing crop growth, such as management practices and weather conditions. The result is consistent with the fact that yield is related to the duration of green biomass formation during the head-to-grain filling stage (Huang et al., 2015; Tewes et al., 2020b). For instance, Li et al. (2017) found acceptable wheat yield estimation accuracy with jointing to anthesis stage LAI assimilations in the CERES-Wheat model. Thus, the assimilation of LAI during these key crop growth stages can be used for yield estimation when data are limited.

Peak LAI is the best yield estimator in many empirical models that form relationships between EO-derived biophysical parameters and yield (Xie et al., 2017). Nevertheless, weather events and other stress factors that occur after the peak period can have a significant influence on crop yield. This indicates that the timing of available observations could also depend on the crop model architecture and data assimilation approach (Casa et al., 2012). In the calibration of SAFY, we found that LAI and yield estimation accuracy were improved when observations were available in both the vegetative and reproductive stages to capture pre-heading events, such as nutrient stress, and post-heading events, such as water moisture and temperature stress, which are common in the study area.

Yield prediction error was the highest when the model was calibrated with only post-peak growth stage observations (R^2 = 0.35, rRMSE = 30%), which shows the small contribution of LAI observations in the maturity and senescence stage. Calibration with post-peak period increased ELUE and decreased R_S, leading to overestimation and underestimated when calibrating with pre-peak stage LAI observation. However, our result also indicated that the pre-peak stage LAI observation estimated maize yield with a lower prediction error (rRMSE = 20%), which can be useful for in-season yield forecasting by combining with seasonal weather forecasts (Chen and Tao 2022). In-season prediction with EO can be used to complement the laborious and untimely pre-harvest yield forecasting system in the study area. Therefore, developing a robust technique for in-season yield forecasting when EO data become available within the growing season requires future investigation to complement the food security analysis in smallholder agricultural systems in Africa.

The temporal distribution of the LAI estimation error from data fusion was the source of uncertainty in the different phenological stages of model calibration. Despite the valuable role of assimilation in obtaining the input requirement of crop growth models, the bias in the EO retrieved LAI could be the source of model uncertainty (Chen and Tao 2022; Li et al., 2017). We found that the deviation of SAFY simulated LAI from field-measured LAI was higher than that between simulated and data fusion retrieved LAI, which could be associated with LAI retrieval error from data fusion. The LAI retrieval error was higher at the early and late crop growth stages. Thus, calibrating the SAFY model with peak to maturity stage LAI overestimated yield estimation. Underestimation of yield occurred when the calibration inputs were only from the vegetative stage. The temporal distribution of LAI retrieval error from EO data could be due to weather effects, the LAI retrieval method and the quality of ground input data. A lower signal-to-noise ratio at the early crop growth stage increased LAI retrieval error. For high-canopy crops like maize, saturation during the peak growth stage could influence the LAI retrieved from EO and hence yield estimation accuracy (Casa et al., 2012; Waldner et al., 2019). A lower correlation between maturity stage LAI and the final crop yield could also reduce the contribution of maturity stage LAI assimilation (Xie et al., 2017). Joint assimilation of LAI and other crop information derived from multi-source EO data (e.g., soil moisture, land surface temperature) could better account for water stress in rainfed agroecosystems. Moreover, improving the LAI retrieval method could reduce the effect of LAI input uncertainty in the model calibration.

4.2. Effect of frequency of LAI on crop model simulation

LAI and simulated yield were sensitive to the frequency of the calibration input LAI. SAFY yield estimation error tends to increase when the number of LAI inputs for assimilation decreases. Increasing the temporal resolution improves the dynamical simulation accuracy of crop model state variables (e.g., LAI) and the final crop yield estimation (Jin et al., 2018a). LAI simulation error also progressively increased as the input LAI interval increased from 8 to 16, 24 and 32 days. However, the effect of input frequency on crop yield estimation was not as high as the impact on LAI simulation. Our finding agrees with Li et al. (2017), who found that the frequency of assimilated LAI modestly affected CERES-Wheat model estimation performance, but the increased computing time becomes problematic for large-area applications. We found that the 8-day interval was the best estimator of crop yield, and accuracy steadily decreased when the frequency of LAI decreased to 16-day, 24-day and 32-day intervals. Previous studies also showed that the availability of good quality LAI from 7 to 10 days interval adequately characterized crop growth and provided optimal yield simulation accuracy (Waldner et al., 2019; Wang et al., 2019). Multi-sensor data fusion offers sufficiently frequent high spatial resolution observations for crop model assimilation to improve yield estimation in persistently cloudy environments.

A good temporal distribution of LAI throughout the key crop growth stages was more important than the frequency of LAI observations with respect to improve yield estimation accuracy because the temporal distribution influences the optimization of some model parameters. Our findings showed that one LAI observation at each maize growth stage provided more accurate yield estimation than having frequent observations at a particular growth stage. Similarly, Silvestro et al. (2017) and Casa et al. (2012) indicated that the primary source of error in data assimilation stems from the distribution of LAI inputs across the crop growth stages, and optimal estimation can be found using few observations as long as the inputs are well distributed across the main phenological stages. Our result

highlighted that the calibration of ELUE, Pla, Plb and R_S requires well-distributed LAI during the late vegetative and reproductive crop growth stages.

The effect of input frequency and timing could depends on crop model types and the data assimilation approach. Based on our calibration approach, this study only evaluated the effect of EO-derived LAI assimilation into one semi-empirical model. Previous studies showed that the frequency and timing of EO data influence the WOFOST (Huang et al., 2015; Wu et al., 2021) model using forcing and updating data assimilation approaches (Waldner et al., 2019; Xie et al., 2017) and empirical models (Casa et al., 2012). Model inter-comparisons could provide further insight into the best model/EO data assimilation approaches for yield estimation.

5. Conclusion

The study explored the effect of LAI input frequency provided from Landsat- MODIS fusion in the parameterization of the SAFY model and resulting maize yield estimation accuracy. It also analysed the relative importance of LAI at different maize crop phenological stages. We used Landsat and MODIS fused data to obtain frequent LAI inputs to calibrate SAFY. The simulated outputs were evaluated using field LAI and yield data collected during the 2019 and 2020 crop-growing seasons. In general, SAFY model calibration was sensitive to the timing, frequency, and quality of LAI observations during the crop-growing season. We can draw the following conclusions and recommendation.

(1) Calibration of SAFY using LAI time series throughout the growing season provided the best-estimate of maize yield. Model calibration covering the period from days of full canopy development to the maturity is important for improved yield prediction.
(2) The temporal distribution of LAI was more important than the frequency of LAI observation. Therefore, when LAI inputs are limited because of cloud cover, obtaining at least one observation at each key growth stage (jointing to grain filling) can provide a sufficiently accurate yield estimation.
(3) The quality of the LAI observation that is retrieved from EO influences the model calibration. The frequency of field observation, field instrument uncertainties, atmospheric effect, and the difference between EO acquisition date and field measurement all influenced LAI retrieval accuracy. Ideally ground LAI data should be collected within 2–3 days of the satellite overpass period and the time series should be smoothed.
(4) Our validation over two growing seasons showed that the model can capture annual and spatial yield variation. However, experiments over longer periods are required.

CRediT authorship contribution statement

Biniam Sisheber: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Writing – original draft. **Michael Marshall:** Resources, Conceptualization, Investigation, Methodology, Supervision, Writing – review & editing. **Daniel Mengistu:** Resources, Conceptualization, Supervision, Writing – review & editing. **Andrew Nelson:** Resources, Conceptualization, Investigation, Methodology, Supervision, Writing – review & editing, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgment

The authors would like to thank the developers of SAFY (Dr.Benoît Duchemin) and ESTARFM (Dr. Xiaolin Zhu) for providing their codes to be used this study. This research was funded by the Dutch organization for internationalization in education (Nuffic), University of Twente, Faculty of Geo-information Science and Earth Observation (ITC), The Netherlands and Ministry of Education of Ethiopia under the Ethiopia Education Network to Support Agricultural Transformation (EENSAT) project (Grant No: CF13198,2016). We also thank farmers in the study area for providing crop growth data.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.rsase.2024.101272.

References

Battude, M., Al Bitar, A., Morin, D., Cros, J., Huc, M., Sicre, C.M., Le Dantec, V., Demarez, V., 2016. Estimating maize biomass and yield over large areas using high spatial and temporal resolution Sentinel-2 like remote sensing data. Rem. Sens. Environ. 184, 668–681.
Bolton, D.K., Friedl, M.A., 2013. Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics. Agric. For. Meteorol. 173, 74–84.
Casa, R., Varella, H., Buis, S., Guerif, M., De Solan, B., Baret, F., 2012. Forcing a wheat crop model with LAI data to access agronomic variables: Evaluation of the impact of model and LAI uncertainties and comparison with an empirical approach. Eur. J. Agron. 37, 1–10.
Chen, Y., Tao, F., 2022. Potential of remote sensing data-crop model assimilation and seasonal weather forecasts for early-season crop yield forecasting over a large

area. Field Crops Res. 276, 108398.
Claverie, M., Demarez, V., Duchemin, B., Hagolle, O., Ducrot, D., Marais-Sicre, C., Dejoux, J.F., Huc, M., Keravec, P., Beziat, P., Fieuzal, R., Ceschia, E., Dedieu, G., 2012. Maize and sunflower biomass estimation in southwest France using high spatial and temporal resolution remote sensing data. Rem. Sens. Environ. 124, 844–857.
Curnel, Y., de Wit, A.J.W., Duveiller, G., Defourny, P., 2011. Potential performances of remotely sensed LAI assimilation in WOFOST model based on an OSS Experiment. Agric. For. Meteorol. 151, 1843–1855.
Dong, T.F., Liu, J.G., Qian, B.D., He, L.M., Liu, J., Wang, R., Jing, Q., Champagne, C., McNairn, H., Powers, J., Shi, Y.C., Chen, J.M., Shang, J.L., 2020. Estimating crop biomass using leaf area index derived from Landsat 8 and Sentinel-2 data. ISPRS J. Photogrammetry Remote Sens. 168, 236–250.
Dong, T.F., Liu, J.G., Qian, B.D., Zhao, T., Jing, Q., Geng, X.Y., Wang, J.F., Huffman, T., Shang, J.L., 2016. Estimating winter wheat biomass by assimilating leaf area index derived from fusion of Landsat-8 and MODIS data. Int. J. Appl. Earth Obs. Geoinf. 49, 63–74.
Duchemin, B., Maisongrande, P., Boulet, G., Benhadj, I., 2008. A simple algorithm for yield estimates: evaluation for semi-arid irrigated winter wheat monitored with green leaf area index. Environ. Model. Software 23, 876–892.
Eklundh, L., Jönsson, P., 2017. TIMESAT 3.3 with Seasonal Trend Decomposition and Parallel Processing Software Manual, 92. Lund University. http://www.nateko.lu.se/TIMESAT/.
Elders, A., Carroll, M.L., Neigh, C.S.R., D'Agostino, A.L., Ksoll, C., Wooten, M.R., Brown, M.E., 2022. Estimating crop type and yield of small holder fields in Burkina Faso using multi-day Sentinel-2. Remote Sens. Appl.: Society and Environment 27, 100820.
Firezer, G.K., 2019. Association of traits and adaptability of hybrid maize (Zea mays L.) varieties in western part of Ethiopia. Greener Journal of Agricultural Sciences 9, 7–13.
Gao, F., Anderson, M., Daughtry, C., Johnson, D., 2018. Assessing the variability of corn and soybean yields in central Iowa using high spatiotemporal resolution multi-satellite imagery. Rem. Sens. 10.
Gao, F., Hilker, T., Zhu, X.L., Anderson, M.C., Masek, J.G., Wang, P.J., Yang, Y., 2015. Fusing Landsat and MODIS data for vegetation monitoring. IEEE Geoscience and Remote Sensing Magazine 3, 47–60.
Gummadi, S., Rao, K.P.C., Seid, J., Legesse, G., Kadiyala, M.D.M., Takele, R., Amede, T., Whitbread, A., 2018. Spatio-temporal variability and trends of precipitation and extreme rainfall events in Ethiopia in 1980-2010. Theor. Appl. Climatol. 134, 1315–1328.
Huang, J., Gómez-Dans, J.L., Huang, H., Ma, H., Wu, Q., Lewis, P.E., Liang, S., Chen, Z., Xue, J.-H., Wu, Y., Zhao, F., Wang, J., Xie, X., 2019a. Assimilation of remote sensing into crop growth models: current status and perspectives. Agric. For. Meteorol. 276–277, 107609.
Huang, J.X., Ma, H.Y., Sedano, F., Lewis, P., Liang, S., Wu, Q.L., Su, W., Zhang, X.D., Zhu, D.H., 2019b. Evaluation of regional estimates of winter wheat yield by assimilating three remotely sensed reflectance datasets into the coupled WOFOST-PROSAIL model. Eur. J. Agron. 102, 1–13.
Huang, J.X., Tian, L.Y., Liang, S.L., Ma, H.Y., Becker-Reshef, I., Huang, Y.B., Su, W., Zhang, X.D., Zhu, D.H., Wu, W.B., 2015. Improving winter wheat yield estimation by assimilation of the leaf area index from Landsat TM and MODIS data into the WOFOST model. Agric. For. Meteorol. 204, 106–121.
Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., Ferreira, L.G., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Rem. Sens. Environ. 83, 195–213.
Hunt, M.L., Blackburn, G.A., Carrasco, L., Redhead, J.W., Rowland, C.S., 2019. High resolution wheat yield mapping using Sentinel-2. Rem. Sens. Environ. 233, 111410.
Jégo, G., Pattey, E., Liu, J., 2012. Using Leaf Area Index, retrieved from optical imagery, in the STICS crop model for predicting yield and biomass of field crops. Field Crops Res. 131, 63–74.
Jin, X., Kumar, L., Li, Z., Feng, H., Xu, X., Yang, G., Wang, J., 2018a. A review of data assimilation of remote sensing and crop models. Eur. J. Agron. 92, 141–152.
Jin, X.L., Kumar, L., Li, Z.H., Feng, H.K., Xu, X.G., Yang, G.J., Wang, J.H., 2018b. A review of data assimilation of remote sensing and crop models. Eur. J. Agron. 92, 141–152.
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. Rem. Sens. 9, 190.
Liao, C., Wang, J., Dong, T., Shang, J., Liu, J., Song, Y., 2019. Using spatio-temporal fusion of Landsat-8 and MODIS data to derive phenology, biomass and yield estimates for corn and soybean. Sci. Total Environ. 650, 1707–1721.
Lobell, D.B., 2013. The use of satellite data for crop yield gap analysis. Field Crops Res. 143, 56–64.
Lobell, D.B., Thau, D., Seifert, C., Engle, E., Little, B., 2015. A scalable satellite-based crop yield mapper. Rem. Sens. Environ. 164, 324–333.
Marshall, M., Tu, K., Brown, J., 2018. Optimizing a remote sensing production efficiency model for macro-scale GPP and yield estimation in agroecosystems. Rem. Sens. Environ. 217, 258–271.
Meshesha, D.T., Abeje, M., 2018. Developing crop yield forecasting models for four major Ethiopian agricultural commodities. Remote Sens. Appl.: Society and Environment 11, 83–93.
Monteith, J.L., 1972. Solar-radiation and productivity in tropical ecosystems. J. Appl. Ecol. 9, 747–766.
Roy, D.P., Kovalskyy, V., Zhang, H.K., Vermote, E.F., Yan, L., Kumar, S.S., Egorov, A., 2016. Characterization of Landsat-7 to Landsat-8 reflective wavelength and normalized difference vegetation index continuity. Rem. Sens. Environ. 185, 57–70.
Silvestro, P.C., Pignatti, S., Pascucci, S., Yang, H., Li, Z.H., Yang, G.J., Huang, W.J., Casa, R., 2017. Estimating wheat yield in China at the field and district scale from the assimilation of satellite data into the aquacrop and simple algorithm for yield (SAFY) models. Rem. Sens. 9, 24.
Sisheber, B., Marshall, M., Ayalew, D., Nelson, A., 2022. Tracking crop phenology in a highly dynamic landscape with knowledge-based Landsat–MODIS data fusion. Int. J. Appl. Earth Obs. Geoinf. 106, 102670.
Sisheber, B., Marshall, M., Mengistu, D., Nelson, A., 2023. Detecting the long-term spatiotemporal crop phenology changes in a highly fragmented agricultural landscape. Agric. For. Meteorol. 340, 109601.
Sisheber, B., Marshall, M., Mengistu, D., Nelson, A., 2024. Assimilation of Earth observation data for crop yield estimation in smallholder agricultural systems. IEEE J. Sel. Top. Appl. Earth Obs. Rem. Sens. 17, 557–572.
Srivastava, A.K., Mboh, C.M., Faye, B., Gaiser, T., Kuhn, A., Ermias, E., Ewert, F., 2019. Options for sustainable intensification of maize production in Ethiopia. Sustainability 11.
Tewes, A., Hoffmann, H., Krauss, G., Schäfer, F., Kerkhoff, C., Gaiser, T., 2020a. New approaches for the assimilation of LAI measurements into a crop model ensemble to improve wheat biomass estimations. Agronomy 10, 446.
Tewes, A., Montzka, C., Nolte, M., Krauss, G., Hoffmann, H., Gaiser, T., 2020b. Assimilation of sentinel-2 estimated LAI into a crop model: influence of timing and frequency of acquisitions on simulation of water stress and biomass production of winter wheat. Agronomy-Basel 10.
Waldner, F., Horan, H., Chen, Y., Hochman, Z., 2019. High temporal resolution of leaf area data improves empirical estimation of grain yield. Sci. Rep. 9, 15714.
Wang, Kun Yu, Lu, B., Tian, M., 2019. Regional rice yield estimation based on assimilation of remote sensing data and crop growth model with Ensemble Kalman method. The 40th Asian Conference on Remote Sensing (ACRS 2019).
Whitcraft, A.K., Becker-Reshef, I., Killough, B.D., Justice, C.O., 2015. Meeting Earth observation requirements for global agricultural monitoring: an evaluation of the revisit capabilities of current and planned moderate resolution optical Earth observing missions. Rem. Sens. 7, 1482–1503.
Wu, S., Ren, J., Chen, Z., Yang, P., Li, H., Liu, J., 2021. Evaluation of winter wheat yield simulation based on assimilating LAI retrieved from networked optical and SAR remotely sensed images into the WOFOST model. IEEE Trans. Geosci. Rem. Sens. 59, 9071–9085.
Xie, Y., Wang, P., Bai, X., Khan, J., Zhang, S., Li, L., Wang, L., 2017. Assimilation of the leaf area index and vegetation temperature condition index for winter wheat yield estimation using Landsat imagery and the CERES-Wheat model. Agric. For. Meteorol. 246, 194–206.
Yang, Y.J., Ren, W., Tao, B., Ji, L., Liang, L., Ruane, A.C., Fisher, J.B., Liu, J.G., Sama, M., Li, Z., Tian, Q.J., 2020. Characterizing spatiotemporal patterns of crop phenology across North America during 2000-2016 using satellite imagery and agricultural survey data. ISPRS J. Photogrammetry Remote Sens. 170, 156–173.
Zhu, X.L., Chen, J., Gao, F., Chen, X.H., Masek, J.G., 2010. An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions. Rem. Sens. Environ. 114, 2610–2623.