In-season Crop Yield Forecasting in Africa by Coupling Remote Sensing and Crop Modeling: A Systematic Literature Review

Wei Xiong, International Maize and Wheat Improvement Center (CIMMYT), w.xiong@cgiar.org

Abstract Timely and accurate estimation of crop yield before harvest is crucial for national food policy and security assessments. Crop models and remote sensing techniques have been combined and applied in crop yield estimation on a regional scale. Previous studies have proposed models for estimating canopy state variables and soil properties based on remote sensing data and assimilating these estimated canopy state variables into crop models. This paper presents an overview of the comparative introduction, latest developments and applications of crop models, remote sensing techniques, and data assimilation methods in the growth status monitoring and yield estimation of crops, facilitating the improvement of crop models and RS coupling approach in Africa.

CIMMYT MR
International Maize and Wheat Improvement Center

Table of Contents

In-season Crop Yield Forecasting in Africa by Coupling Remote Sensing and Crop Modeling: A Systematic Literature Review

Abstract

Timely and accurate estimation of crop yield before harvest is crucial for national food policy and security assessments. Crop models and remote sensing techniques have been combined and applied in crop yield estimation on a regional scale. Previous studies have proposed models for estimating canopy state variables and soil properties based on remote sensing data and assimilating these estimated canopy state variables into crop models. This paper presents an overview of the comparative introduction, latest developments and applications of crop models, remote sensing techniques, and data assimilation methods in the growth status monitoring and yield estimation of crops, facilitating the improvement of crop models and remote sensing coupling approach in Africa.

Introduction

Agriculture is the major land use in Africa and is the primary income source for smallholder farmers. However, in many forms, African agriculture remains highly sensitive to both climate extremes and variations in climate and trends over a range of time series, particularly in regions where rainfed agriculture supports the majority of the population and plays crucial roles in national economies like East Africa (Ogutu et al., 2018). Improving the resilience of the agriculture sector by preparing vulnerable populations for extremes weather variability and developing reliable crop production (Matthew et al., 2015) can not only have a positive effect on socioeconomic development but also enhance food security through better agricultural management and policy formulation that proactively accounts for variable climatic conditions (Bahaga et al., 2015).

Forecasting in-season crop yield, namely estimating several weeks in advance how much there will be on the field at harvest, has become increasingly tangible and important. Farmers, commodity markets, insurance, seed traders, and logistics companies, as well as regional authorities and food aid programs, need outlooks on expected harvests to adapt their management of fields, firms, or food balances (Schauberger et al., 2020). Examples of the use of forecasts are crop insurance or early warning systems, which were ranked as the top adaptation measure with the highest economic return on investment (Global Commission on Adaptation, 2019).

There are a number of systems and techniques to forecast crop yield, both in the scientific literature and in practical application. Global and transnational yield forecasting systems exist based on models or as exchange platforms that combine national forecasts to provide a global outlook (Fritz et al., 2019). In Africa, efforts towards improved resilience to extreme climate variability are ongoing through the

issuance of pre-season climate forecasts generated by both statistical and dynamic methods. For example, in east Africa, the Greater Horn of Africa Climate Outlook Forum (GHACOFs) (Martines et al., 2010) brings together scientists to develop a consensus on rainfall and temperature forecasts for the coming seasons plus likely impacts on climate-sensitive sectors, including agriculture (Hansen et al., 2011). The scientists further downscale the consensus seasonal climate outlooks for national impacts and other purposes. However, these seasonal climate impact outlooks are generally based on subjective expert judgment rather than explicit quantitative methods. In addition, gaps exist between the African applications and global exercises due to data scarcity, the slow adoption of new technologies, and small research teams. Here we present a systematic review of existing approaches in Africa from the scientific literature to forecast in-season crop yield. The review covers almost the whole continent and major food crops, with a focus on methods to link crop models (CMs) and remote sensing (RS), due to their increasing importance in forecasting and the objectives of the project.

This systematic review complements previous reviews on successful applications of yield forecast with crop models and remote sensing in Africa. We also compared the applications with studies conducted in developed countries and developing countries on other continents, reviewed by a number of papers. This review provides an overview of the usage data (weather, remote sensing data), method, and application examples of in-season yield forecast in Africa, highlighting their virtues and deficiencies compared to current development across the world. The report is structured as follows. First, we describe the review method used to collect the paper and define the questions. Then, we summarize the papers regarding research crops, geographical boundaries, and the type of models or data used. In the following sections, we discuss the method used in Africa and their advantages

and drawbacks, including seasonal climate forecasting methods, crop modeling, remote sensing, and linking. In the last section, we draw a conclusion.

Research Method

Review methodology

This systematic literature review helps us to understand the application of remote sensing and crop modeling in in-season crop yield prediction. This systematic literature review is carried out to highlight the existing research gaps in Africa and guide us in utilizing the remote sensing indices and crop models to enable in-season crop yield forecasting. For the systematic literature review, not only are all research studies from journals, conferences, and other electronic databases assessed but they are also integrated and presented in correspondence to the research questions mentioned in our study.

A systematic literature review is an exceptional way to evaluate a theory or evidence in a specific area or to study the accuracy or validity of a specific theory (Muruganantham et al., 2022). The review guidelines given by Kitchenham and Charters (2007) are appropriate for our systematic literature review as they provide objectivity and transparency. Based on the review guidelines, initially, the research questions are formulated. The review is undertaken in accordance with the Preferred Reporting Items for Systematic reviews and Meta-Analysis (PRISMA) statement (Page et al., 2021). Several databases, such as IEEE Explorer, ScienceDirect, Scopus, Google Scholar, MDPI, and Web of Science, are used for selecting relevant research articles. These research articles are assessed and

filtered based on the quality criteria. A complete checklist of PRISMA[1] is used for conducting and reporting the results of the review.

Research questions

The following research questions are developed to guide the systematic review:

1. "What data products such as weather are used for in-season crop yield forecasts in Africa?"

   This question helps us to analyze the weather, remote sensing, field observation data availability, advantages, and limitations of different datasets.

2. "What remote sensing technologies are used for crop yield prediction in Africa?"

   With various remote sensing technologies in existence, this question helps us to understand the suitable remote sensing technology based on the data acquisition requirements for the study of crop yield prediction, such as land area and crop type.

3. "What crop models are used for crop yield prediction in Africa?"
   Answering this allows us to understand the advantage and disadvantages of crop models when utilizing them in the data scarcity continent - Africa.

4. "What are the vegetation indices and environmental parameters used in coupling the RS and CMs?"
   This question enables us to learn about the various features that are influencing the coupling approach and the forecasting accuracy in crop yield prediction.

---

[1] See https://prisma-statement.org

5. "What are the challenges in using RS and CM for in-season crop yield forecasts in Africa?"

   This question helps us to understand the limitations and challenges in the existing approaches.

Procedure for article search

The approach to searching the articles is designed based on the framed research questions and the aim of the systematic literature review. Narrowing down the focus from a major concept to the central idea of the review helps in creating an effective search strategy. Using "Crop model" or "Remote Sensing" alone as a search string will generate a lot of published articles from various application fields that are not likely related to the aim of the review and cause the search to be complicated. Redefining the search strategy as "crop yield prediction" AND "remote sensing" AND "crop model" AND "Africa" can reduce the probability of deviating from the scope of the review. Initially, by using these search strings, the articles were retrieved from five databases, including IEEE Explorer, Science Direct, Scopus, Google Scholar, and MDPI. Further, to include any other relevant studies, the following keywords, namely "crop yield prediction" OR "crop yield estimation" AND "seasonal" AND "remote sensing" OR "yield forecast" AND "Africa," were used to retrieve the articles from the databases. The articles from the last ten years (2012–2022) were used for the study.

Article selection criteria

The retrieved articles are initially selected based on aspects such as the quality of a journal, any type of remote sensing technology and crop models used for the study, and the type of coupling approaches adopted. Analyzing the abstracts of articles

helps in understanding the keywords and the selection of articles. The exclusion of irrelevant articles was carried out based on the following criteria:

- Articles that belong to the agricultural sector but that do not fall under crop yield prediction;
- Articles that do not cover Africa;
- Publications that have no open access;
- Literature search for articles that are published before 2012;
- Articles in different languages other than English.

Literatures Overview

After applying all the exclusion criteria, a total of 41 articles are selected. Further removing the repeated articles across the selected databases, 36 articles are selected for the review. In Figure 1, we explain the process for article selection and rejection from databases for the review based on PRISMA. Figure 1 shows the number of articles retrieved after selection criteria are applied and the number of articles obtained after excluding the repeated articles from the selected databases. The research questions are addressed after all the data from retrieved articles are summarized and synthesized.
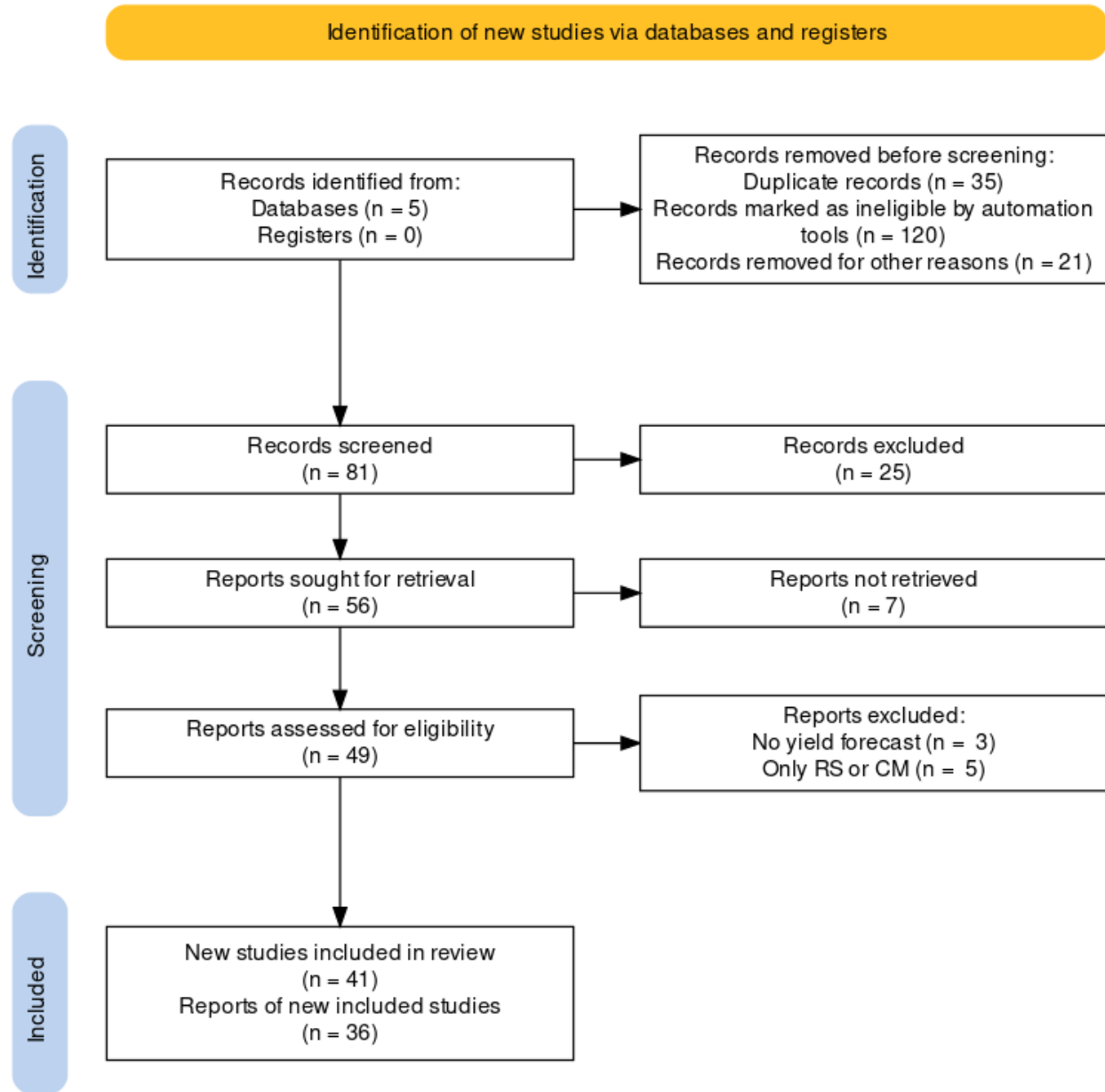
Identification

Records identified from:
Databases (n = 5)
Registers (n = 0)

Records removed before screening:
Duplicate records (n = 35)
Records marked as ineligible by automation
tools (n = 120)
Records removed for other reasons (n = 21)

Screening

Records screened
(n = 81)

Records excluded
(n = 25)

Reports sought for retrieval
(n = 56)

Reports not retrieved
(n = 7)

Reports assessed for eligibility
(n = 49)

Reports excluded:
No yield forecast (n = 3)
Only RS or CM (n = 5)

Included

New studies included in review
(n = 41)
Reports of new included studies
(n = 36)

Figure 1 PRISMA flow chart showing the results of searches.

Figure 2 shows the number of articles published between 2012 and 2022. It is evident that the study of crop yield forecasts using crop models and remote sensing has increased in recent years, particularly in 2022. This is because of the availability of free satellite data, such as Landsat, MODIS, Sentinel, and the open-sourcing campaign of various crop models through global modeling programs, such

as the Agricultural Models Intercomparison Program (AgMIP). Studies have been conducted in most countries in Africa, most frequently in Ethiopia (9), Burkina Faso (5), and South Africa (4). The selected articles on yield forecasting show a wide range of crops, including non-staple or horticultural crops. Maize is the most crop that is studied in the literature, followed by wheat, sugarcane, millet, and sorghum.
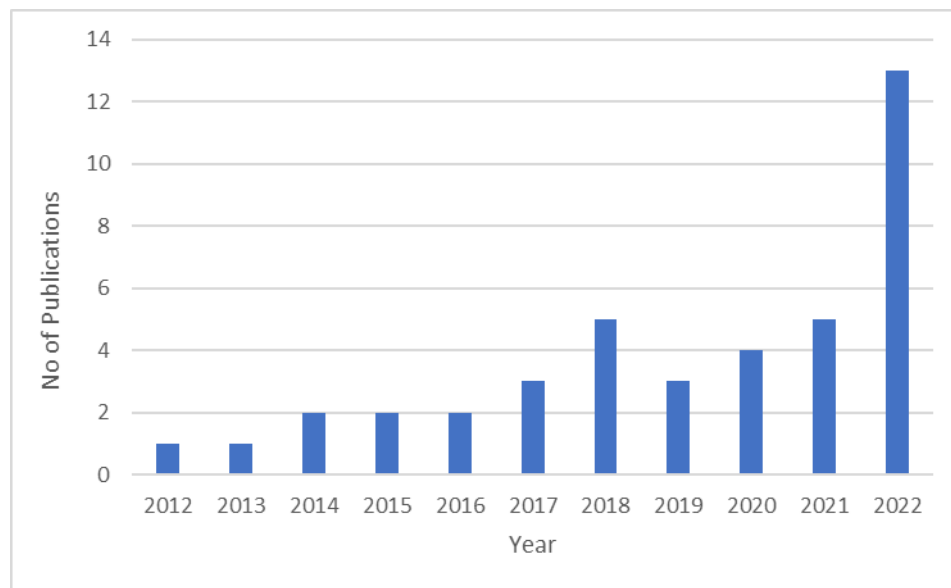

Figure 2 Distribution of articles between 2012 and 2022.

In terms of crop models, although a wide range of model types and specific models are used for forecasting, more than half of the studies employ machine learning or statistical models because of their straightforward nature and less requirement for inputs. Processed-based crop models are also frequently used in Africa to forecast crop yields, including the widely used crop models APSIM (n=2), AquaCrop (n=1), CROPWAT (n=1), DSSAT (n=2), SARRA (n=3), and WOFOST (n=2). All these models are developed in developed countries, with training data from environments that are

different from Africa. A small number of studies (<10) tested or validated the models in African environments with real field observations.
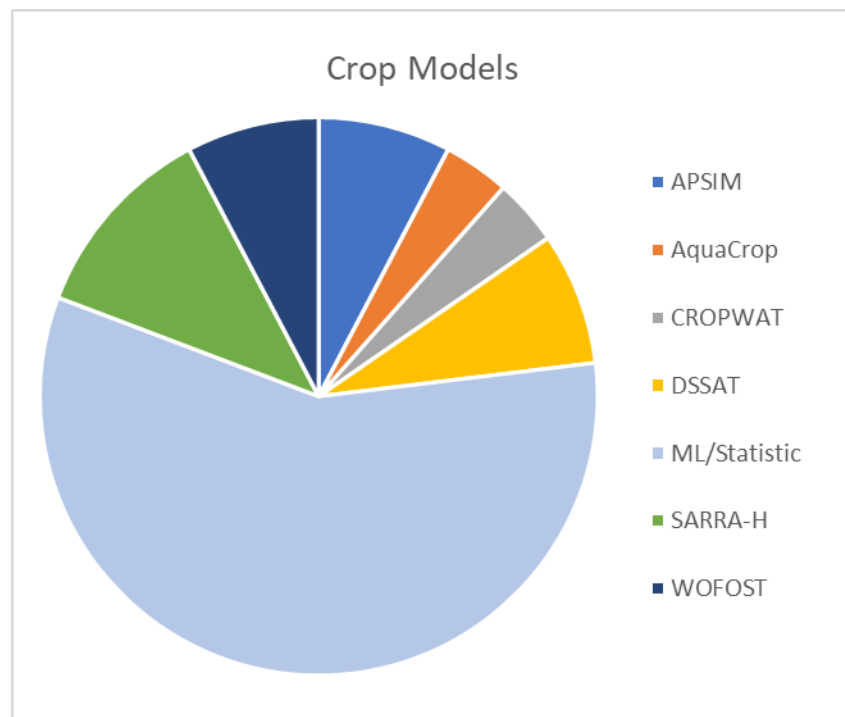

Figure 3 Types of crop models have been used in the publications.

Seasonal Climate Forecasts for Africa

Seasonal climate forecasts are currently routinely issued up to 12 months before the start of seasons (lead-time) by numerous operational global forecast centers. With sufficient lead time before the start of a growing season, different adaptation options are possible (e.g., choosing different crops or varieties, heavy or low investment in farm inputs) as opposed to forecasts issued after crops are planted. The exactness of lead times indicated in the studies varies, with six studies mentioned only that a forecast before the harvest was performed, but the actual

lead time is not stated. The most frequent lead times are several months to one month.

Table 1 Overview of weather products useful for yield forecasting.

Name	Time period	Latency	Temporal resolution	Spatial resolution, coverage	Variables[2]	If used in Africa
Global near real-time products						
NASA POWER	1981-present	1-7 days	Daily	0.5° global	T, P, H, SW, LW, W	Yes
NASA POWER – GEWEX SRB 3.0	1983-2007, 2008-present	7-8 days	3 - hly	1° global	SW, LW	No
NASA POWER – MERRA 2	1981 to present	Few months	hourly	0.5° x 0. 625° global	T, P, SW, LW	Yes
NASA POWER – GEOS FP	End of MERRA 2 to near-real time	4 h	hourly	0.25° x 0. 3125° global	T, P, SW, LW	No
ERA5 (ECMWF ReAnalysis)	1980 to present	3 months	hourly	0.25° global	T, P, SW, LW, W	Yes
NCEP/NCAR	1948 to present	1 days	hourly	0.25° global	T, P, H, W	Yes
NASA GLDAS	1948 to present	A month	3 - hly	0.25° global	T, P, SW, LW	No
CHIRPS	1981 to present	2 days	daily	0.05° global	P	Yes
CHIRTS	1983 to 2016		daily	0.05° global	Tmax, Tmin	Yes
Global short-term to season forecasts						
NCEP GFS	8 days	6 h	3 -hly	0.5° global	T, P, H, W	No
TIGGE	1 to 15 days	48 h	6 -hly		T, P, SW	
ECMWF - HRES	10 and 15 days		Twice daily	9 and 19 km	T, P, H, SW, W	No
ECMWF	46 days and 7 months		Twice weekly and monthly	36 km	T, P, H, SW, W	Yes
S2S	Up to 60 days	Updated daily with a 21-day delay	Weekly anomalies	Global	T, P	No
NCEP – CFSv4	45 days to 6 months	6 h	6 -hly	0.5° global	T, P, SW, LW	No
JMA/MR1-CPS2	Up to 7 months	few days	daily	0.5° global	T, P, W, H	No

---

[2] Variables codes are 2 m air temperature (T), precipitation (P), humidity (H), shortwave (SW), longwave surface radiation (LW), and wind (W). A detailed description of each weather data product, see Schauberger et al. (2020).

Forecasts are provided at various growth stages, with the longest lead times observed for sugarcane.

The short-term and seasonal weather forecasts that have been used in the publications were mainly based on products produced by developed countries like Europe. Schauberger et al. (2020) provide an overview of near-real-time observation-based weather products. Many have been used in Africa to conduct crop yield forecasts (Table 1). NASA POWER (Prediction Of Worldwide Energy Resources) is a prominent near-real-time product with a latency of a few days for most variables, combining solar radiation data based on radiative transfer models from satellite observations with meteorological data from MERRA2 (Modern-Era Retrospective analysis for Research and Application). The time gap (latency of a few months) until MERRA 2 becomes available is bridged with GEO FP (Global Earth Observing System Forward-Processing) to provide near-real data. ERA5 is the new ECMWF (European Centre for Medium-Range Weather Forecasts) near-real-time reanalysis product, replacing ERA-Interim reanalysis. CHIRPS (Climate Hazards Infrared Precipitation with Stations) and CHIRTS (Climate Hazards Infrared Temperature with Stations) are new high-resolution products at 3-arcmin global coverage, combining satellite observations with station data. These products provide an advantage, especially in Africa, with scarce station data.

Several options are available to continue yield simulations after the forecasting day throughout the growing season. These include historical weather data, employing a weather generator, or using weather predictions. Table 1 is the overview of the most common weather forecast products currently available. General short-term weather forecasts (up to 2 weeks) can be continued with sub-seasonal (up to 3 months) and seasonal outlooks, but forecast skill beyond ten days is as yet generally marginal. Prominent short-term products used in Africa include ECEP's

GFS and the ECMWF ensemble. Several studies in our literature data use historical climate as a proxy for unobserved weather in the future. This can be achieved via trend extrapolation, historical averages, or more sophisticated choices where the observed weather during the season of interest until the forecasting day is used to identify historical weather analogues to prescribe a weather trajectory until the end of the season.

Global Climate Model (GCM) based seasonal climate forecasts have been used in agricultural impacts modeling globally and in Africa with varied results, suggesting variations in skill due to factors like spatiotemporal scales used, level of surface heterogeneity, crop management practices, and model initialization, amongst others. Driving crop models with skillful seasonal climate forecasts may not guarantee good yield forecasts (Shin et al., 2010), but the reverse, i.e., better skill in crop forecast than in the meteorological forecast, has also been reported (McIntosh et al., 2005). In addition, whether a crop in a certain region experiences temperature or moisture limitations affects yield predictability differently. For example, since temperature influences crop phenological development and its predictability is generally higher than for precipitation (Ogutu et al., 2016), its predictability influences yield predictability differently. Finally, the time of the year in which a forecast is useful depends on the crop and region (McIntosh et al., 2007), depending on the local cropping calendars.

Crop Yield Forecast

Several tools have been developed over the years to assess the production and distribution of food resources across Africa as part of the food security assessments. To satisfy the long-term requirement for crop yield forecasting, many methods have been developed to provide information in advance about potential

crop production. These methods can be grouped into (i) physical field and survey-based assessment, (ii) time trend analysis, (iii) crop growth simulation models, and (iv) remote sensing-based methods.

Physical field and survey-based assessments

Field-based or survey-based methods are the traditional way of conducting crop yield estimates where trained and experienced field or survey staff sample and qualitatively score sampled crop fields to estimate crop yield and quality. This is the commonly used approach in many countries in sub-Saharan Africa, where ministries use their dense network of field extension officers to report on the quantity of the season based on these field reports (Chitsiko et al., 2022). It is, therefore, widely utilized, well-accepted, and an official approach to estimating crop yields in many counties. This method is based on field-observed data and relies on verified actual crop fields for prediction. In addition, with experience, the quality of the forecasts increases as the field officers become more accurate, resulting in better decision-making. Field-based methods based on estimating crop yield rely on currently established networks to provide crop yield estimates, and therefore there are little to no establishment and operational costs. However, these methods also come with challenges.

Among the greatest setbacks of field-based forecasting is the reliance on subjective judgments by the individual, which gives different scores for the same condition (Kuri et al., 2014). They also require extensive fieldwork to produce representative results, and this makes it not only grueling and time-consuming but also very expensive. In addition, given the different planting dates in different regions, it is very difficult to harmonize the crop yield estimates based on different crop stages. Related to that is that the results from these estimates are not instantaneous as it

can take a long time to complete the assessments across large areas. It is also a paradox that the estimate relies on field extension officers employed and expected to increase crop productivity in their areas of jurisdiction. There is, therefore, a general temptation for field officers to overstate the crop yield estimates in order to be considered effective, thereby compromising the results.

Time trend analysis

Time trend analysis is another method used in estimating crop yields. The method estimates yields using statistical analysis of historical trends and adjusts for other variables such as weather, soils, and markets. The model is parameterized at different spatial and temporal scales, and when the relationship between the variables and yield is established, then the yield estimate can be predicted for a season or for many years. The method is widely used as the FAO's Early Warming System (GIEWS) for yield estimation. In time trend analysis, the field reports on crop yield data area regressed against known influential meteorological parameters such as the start of the rainfall season, total rainfall, and mean monthly temperature of the agricultural season to generate a functional model. An example of time trend analysis was done by Monatsa (2011), where they used long-term rainfall data as input into a crop water balance model to calculate the water requirement satisfaction index (WRSI) for maize in Zimbabwe.

Yield forecast using agrometeorological inputs and time series yield data into a statistical regression is another common method and is used in many yield forecast research and programs. In general, a simple statistical model is built using a matrix with historical yield and several agrometeorological parameters (e.g., temperature and rainfall). Then, a regression equation is derived between yields as a function of one or several agrometeorological parameters. For example, Davenport (2019)

utilized the long-term yield data and environmental data in developing a statistical yield forecast model and forecast for 56 regions located in two severely food insecure countries: Kenya and Somalia.

Time trend analysis or a statistical model has many advantages in crop yield prediction, especially when compared to field-based methods. This approach does not necessarily require extensive fieldwork, and it is thus cheaper. The calculation is easy, less time is required to run the model, and the data required are limited. In addition, the approach can be easily extended and extrapolated beyond the areas where data was available or not available and produce robust results. Since the approach is statistical in nature, there is more confidence in its application as there are established standardized measures of model performance that are used to evaluate the accuracy of the model before use. Given the recent improvements in computational powers and Machine Learning approach, these methods are also faster to implement for decision making.

Time trend analysis or statistical model, however, has challenges. They are limited in the information they can provide outside the range of values for which the model is parameterized. Also, the output of such models might not have any agronomic meaning while they are still statistically legitimate. In addition, they do not take into consideration the soil-plant-atmosphere continuum, which is important when dealing with regions having different soil types. For example, the response of a crop to a given amount of rainfall on sandy soil is different than a crop on clay soil. The timing of the water stress occurring during the growing season is also important and often ignored. For example, heat stress occurring at flowering will reduce yields more than heat stress happening during the vegetative phase. This is important for correctly forecasting yield and for giving farmers important agronomic advice.

In Africa, one of the big downsides of the time trend analysis and statistical model is mainly related to the availability and quality of input data used in the time trend analysis. In many African countries, there is a paucity of data. Most of the available data on yield is at a localized scale, and in a few cases where this data is available, the quality is often poor. In addition, the approach is dependent on the availability of a dense network of meteorological stations, which are non-existent in many areas, particularly those facing severe food shortages. This is so because many meteorological stations are located in urban areas, and their representativeness of communal areas where much questionable. In addition, with climate change and variability, the validity of relying on long-term weather data trends to make yield predictions is becoming less reliable. For example, mid-season droughts can significantly reduce yields in a season when the rainfall totals received in a season remain relatively unchanged. Thus using the total rainfall for yield prediction becomes less accurate under these conditions.

Crop growth simulation models (CSM)

Crop growth simulation is another family of yield estimation methods. Agroecosystems are complex entities where crop yield is the result of many interactions, such as soil, atmosphere, water, and socioeconomic factors. CSMs are built with the aim of considering the continuum soil-plant-atmosphere and its daily changes on the daily accumulation of biomass and nitrogen. These models estimate crop yields based on the known characteristics of crop growth, the biophysical environment, management practices, and other factors. There are many CSM around the world, such as CERES, APSIM, and AquaCrop. Crop growth simulation models such as CERES, WOFOST, AquaCrop, and SWAP capture the most likely production potential of an area by weighing the constraints against the factors promoting production to obtain the most likely production potential. Unlike

the time trend analysis approach that relies mostly on historical data, crop growth simulation models are able to provide in-season forecasts by relating crop conditions at particular physiological stages to the yield of the crop for each land use system, management regime, and other related production factors. There are a number of crop growth simulation models that have been utilized over the years for use in yield modeling. Examples of such models include those that relate to probabilistic maize yield prediction in Kenya, Ethiopia, and Tanzania with dynamic ensemble seasonal climate forecasts (Ogutu et al., 2018), yield reduction, and water management in Ethiopia (Eze et al., 2020). Others link soil moisture condition to the potential yield assuming that soil moisture is the most limiting condition for yield in certain land use systems (Luciani et al., 2019).

Crop growth simulation models tend to be very accurate for localized applications when compared to other methods when properly parameterized. They also have fewer data needs and are, therefore, less complex, meaning that they satisfy the parsimony requirement for models. They are also based on field observations making them more empirical and data-driven. In addition, they are also able to adjust according to different significant factors that affect crop yields, such as crop varieties, soil conditions, water supply, management commitment, and other factors. However, crop growth simulation models are based on experimental data, which significantly differs from actual field conditions for crop production. They also produce an indication of the production potential of a particular land use system but not necessarily the actual production.

The difficulties of adopting CSM have usually been associated with the intensive data for models' parameterization. The need for calibration can be quite data-extensive and not applicable to some developing countries. In fact, it has been argued that several variables are needed to calibrate/evaluate the CSMs,

concluding the usefulness of CSMs in some "real" situations because of the impossibility of gathering inputs and calibration datasets. However, a close look at the literature and the work done by other researchers points out that the CSM can be run using "Minimum Data Set" inputs. Models, like the SARRA example used in Mali, have shown to be easy to use by anyone but still maintain their robustness in yield predictions. It has been pointed out that another limitation of the CSM is that they are "point-based" and inadequate to run at a regional/national scale. However, gridded crop models have developed and can be run at regional and national scales with fewer demands on inputs and calibration datasets.

Remote sensing-based methods

Remote sensing-based methods are gaining momentum and gain acceptance in crop yield estimations. Remote sensing systems capture radiation in different wavelengths reflected/emitted by the earth's surface features, which are recorded by sensors to generate images. Over time, the biophysical understanding, algorithms for data handling, data storage capacity, and sensor technology have grown, resulting in many applications of remote sensing methods in crop yield estimations. Remote sensing-based methods have thus been used to predict crop condition and yields in agriculture through directly assessing crop growth and vigor and indirectly through estimation of leaf nutrient status, weed pressure, disease severity, insect attach, and other useful biophysical crop properties related to yields.

Prospects for yield modeling using remote sensing are high, considering the fast research developments in this area. There are many methods used to forecast crop growth and yield from RS, such as the empirical regression model, the biomass production model as a function of absorbed or intercepted solar radiation, and the

stress-degree-day model. The meteorological variables used for forecasting yield are mainly based on two variables, temperature and precipitation, since they are related to crop yields and can be easily obtained from meteorological stations or from satellite measurements. For example, Kuri et al. (2014) successfully developed an approach that uses SPOT VGT-derived dry dekads to predict crop yields at the national level for drought early warming and yield estimation. Dutta et al. (2015) applied a normalized NDVI to get the vegetation condition index that indicates changes in maize crop conditions related to drought. Abebe et al. (2022) calculated vegetation indices from the Landsat 8 and Sentinel 2A observations to estimate sugarcane yield in Ethiopia. Table 3 summarizes the remote sensing data and method used in the reviewed studies.

Remote sensing-based methods have many advantages compared to other crop yield estimation methods. Remote sensing can instantaneously provide estimates from large areas covering countries and entire regions, significantly reducing the costs of doing such exercises. Results from remote sensing are also timely as indications can be obtained in advance, enabling planners and policymakers to efficiently make decisions in advance. In addition, when appropriately analyzed, satellite data provides not just estimates of the quantity but of the quality of the yields (Davis et al., 2016). This is because it is able to integrate the effect of soil type, relief, climate, varieties, and other socioeconomic factors that influence crop performance at different locations, making results more representative and accurate. However, the learning curve and establishment costs of remote sensing applications are high, making their uptake limited in developing countries. Remote sensing estimations are confounded by clouds and non-crop areas, and thus, their success may be limited in many subtropical areas.

Table 2 Summary of strengths and weakness of crop yield prediction methods.

Method	Strengths for application in crop yield estimation	Weaknesses for application	Potential for use with other methods
Physical Field and Survey assessments	Well-accepted as the standard There is already experience with this method as it has been used for a long time Based on field observation and survey data and can be verified Rely on currently established networks, and therefore there are little to no establishment and operational costs.	Relies on subjective assessment by individuals, which gives different scores for the same condition Require extensive fieldwork to produce representative results Tedious, time-consuming, and expensive Results affected by planting dates	Used mainly as a parameter for evaluating or calibrating other assessment methods Little potential for integration with other methods.
Time trend analysis	Does not require extensive fieldwork and is, therefore, cheaper and quicker The method can be easily extended and extrapolated beyond the areas where data was available or not available and produce robust results The statistical approach is good for ensuring confidence It is faster to implement for decision-making.	No quality data is available at the required disaggregated level The approach is dependent on the availability of a dense network of meteorological stations which are not there in many countries the statistical relationships are changing with climate change	Can be integrated with a remote sensing approach where remote sensing can provide long-term data on the condition, yield, or other parameters Can be used with field-based methods where the field data is used to determine the required statistical relationships.
Crop Growth Simulation Models	Can be very accurate for localized application They have fewer data needs and are, therefore, less complex Based on field observations making them more empirical and actual data-driven Can adjust according to different significant factors that affect maize yields.	Based on experimental data, which significantly differs from actual field conditions for crop production Conditions for their development have since changed from now which affects their use More useful for production potential than the actual production estimation.	Can be integrated with remote sensing methods as sources of meteorological data required in running the models
Remote Sensing-Based Methods	Instantaneously provide estimates from large areas covering countries and entire regions, significantly	The learning curve and establishment costs of remote sensing applications are high	Can be integrated with both statistical yield forecasting and crop simulation models.

reducing the costs of doing such exercises Results are timely as indication can be obtained in advance Provide both the quantity and quality of the crop yields.	Remote sensing estimations are confounded by clouds and non-crop areas.

Seasonal crop yield forecasts have been derived from either historical statistical relationships with rainfall or large-scale climate indices such as the El Nino Southern Oscillation (ENSO) Index (Iizumi et al., 2014; Hansen et al., 2009) and its influence on seasonal rainfall in some parts of the world such as eastern and southern Africa. These statistical methods are successful at broader spatial extents like national boundaries or regions and may not suffice for smaller spatial scales where heterogeneities exist. For example, normal rainfall season may result in low yields related to nutrient leaching depending on soil types. High rainfall variability exists in small regional extents even in an otherwise "good rainfall season," and statistical relationships do not capture rainfall characteristics (such as distribution during a season and frequency) that are important for crop yields. In addition, poor records of historical yields on which the statistical models are calibrated also influence prediction skills.

Confronted with the current climate change and variability together with climate teleconnections between a region of interest and other parts of the globe, any past statistical relationships between yields and climate indices may no longer hold true because the future will be under climate regimes not observed before. It is not clear if the relationships between phenological observations and satellite-derived vegetation indices will hold true since observation will also vary under different climate regimes (for example, higher temperatures than in the historical period),

and since crop response to climate is not linear; hence historical observations may not suffice.

Crop Yield Forecast by Coupling CM and RS

The integration of RS and CSM for crop yield forecasting has been researched for almost three decades. The RS can quantify crop status at any given time during the growing season, while CSM can describe crop growth every day throughout the season. RS can indirectly provide a measure for canopy variables which can then be used to adjust the model simulation. Since the first satellite information became variable to scientists, they have developed algorithms to estimate canopy state variables, such as LAI, vegetation fraction, and a fraction of APRR. One of the main integration procedures between RS and CSM focuses on adjusting the LAI simulated with the crop models against the one estimated through R. LAI is an important agronomic parameter between leaves where water and CO_2 are exchanged between the plant and atmosphere; in addition, the LAI is used to model crop evapotranspiration, biomass accumulation, and final yield. Researchers working on such integrations generally adopted three steps: (1) estimate canopy variables with RS; (2) run the CSM; (3) use a proper integration method to adjust model runs. The first step can affect the subsequent results of the integration because if the crop variable is not properly estimated, then adjusting the model with a biased variable will lead to a wrong model evaluation. There are two ways of using RS for the estimation of canopy variables: through the use of statistical/empirical relationships; and the Physical Reflectance Models that simulate the interactions between the solar beam and the various canopy components through the use of physical laws in which the LAI can be entered as the input obtained from the crop model.

The important step in the integration of the crop model and RS is the technique used to combine them, both in space and time, canopy state variables with various information using remote sensing methods for optimizing crop parameters in the crop model. In the integration, one has to first distinguish observed variables (from remote sensing data resources), state variables (from a complete crop model system), and model parameters (described relationships between the observed variables and state variables). Several methods have been described in detail and applied for combining remote sensing data and crop models in different papers. Jin et al. (2018) summarized the methods into three types, namely crop model calibration, forcing methods, and updating methods. Some of the methods have been used in Africa.

1. Calibration method. The initial parameters of crop models are adjusted to optimal consistency between the remote sensing data and the simulated state variables (the simulation data of the crop model). Crop models are manually or automatically run using a realistic scope of different integration parameter values to calibrate the sensitivity and uncertainty of crop model parameters. Many studies have been carried out using the data assimilation of crop models and remote sensing data using calibration methods. There are various specific algorithms, including the simplex search algorithm (Ma et al., 2013), Maximum Likelihood Solution (MLS), Least Squares Methods (LSM) (Zhao et al., 2013), Very Fast Annealing Algorithm (VFSA) (Dong et al., 2013), Powell's conjugate direction method (PCDM) (Huang et al., 2015), and Particle Swarm Optimization Algorithm (VDSA) (Liu et al., 2015). The calibration method was used to minimize differences between the remote sensing data acquisition date and the date of crop model simulations using an optimization algorithm. In our selected literature, Jin et al. (2017) successfully

used this method to predict maize yield in Eastern Africa. This method has also been used to derive new values of LAI and used to calibrate the model for forecast crop yield in Ethiopia (Abebe et al., 2022).

2. Forcing methods. Forcing methods use remote sensing data to replace the crop model simulation data. The remote sensing data is directly used to prescribe the simulation data that required the feasibility of the remote sensing data at each crop time step, which is daily, weekly or monthly in most crop models. Under normal circumstances, satellite transit frequency is less than the time step of the crop model. To drive remote sensing data at the time step in the simulation data of crop models, linear interpolation, fast Fourier transform, and wavelet methods are used to fill the gaps between remote sensing data observations. The estimated LAI from remote sensing data was mainly used as a state variable and was input into crop models. Yao et al. (2015) have estimated LAI using different remote sensing data, and the simulated results of crop models were directly replaced by the estimated LAI to improve the simulated LAI, aboveground biomass, yield, or crop transpiration of crop models. Based on the forcing method, the data assimilation of crop models and remote sensing data is easy to operate, but strictly speaking, it does not involve data assimilation methods. The simulated state variables or initial input data of crop models were only replaced by the estimated state variables or initial input data of remote sensing data. Remote sensing provided the state variables with high precision and enabled good estimated results to be obtained. Because the forcing method needs comprehensive knowledge of crop models and processing the model code, there is a limited number of studies using this method in Africa, at least in the selected literature.

3. Updating methods. When obtaining remote sensing data is feasible, the updating method includes continuously updating crop model simulation data. This is based on the assumption that better simulation data on day t will improve the accuracy of the simulation data on succeeding days. Updating method is usually called data assimilation, and a number of algorithms have been applied to the assimilation of remote sensing data and crop models. With in-depth research on data assimilation and the development of computer technology, data assimilation has been widely applied to combine remote sensing data and crop models. The EnKF, 4DVar, PF, the proper orthogonal decomposition technique into 4DVar, and ensemble square root filter methods were used to combine the state variables of remote sensing data and crop models and estimated soil moistures, AGB, LAI, and yield (Eze et al., 2020).

In the case of the forcing method, crop models do not use their own information but follow the observed state variables, which include some errors. Remote sensing observation data have errors, and these errors will be introduced into crop models when the assimilation is completed by the forcing method. The calibration and updating methods have greater flexibility, and their minimization errors are brought into the crop model when remote sensing data is used in the assimilation process. The calibration method is hoped to get more representative input crop parameters into crop models and improve its prediction accuracy. Remote sensing observation data are used to calibrate crop models if there are sufficient observations, and the observation error is small. To a certain extent, this method could be used to reduce the accumulation and spread of remote-sensing data errors during the process of assimilation. Compared with the forcing and updating method, theoretically, the calibration method is better than the forcing and

updating methods, but the main drawback of this method is that it requires a lot of optimization iterations, resulting in more computing time. Compared to the calibration methods, the updating method significantly reduces the computation time because only the crop model is run. However, this method can be also flawed because it requires the most expensive calculation and measurement uncertainty. In addition, the updating method requires adjusting the crop parameters variables when running the crop model. The date of selected remote sensing images is an important factor affecting estimation accuracy using the updating method.

Conclusions

The review consolidates lessons from previous research on the tools, approaches, and applications by researchers who couple remote sensing with crop models to forecast in-season crop yields as an approach to enhance climate variability management in smallholder farming systems in Africa. The review shows that the most recent work on the integration of remote sensing with crop models has been predominantly over West and Eastern Africa, with limited work being done in other regions of Africa. Specifically, there is a need for more research in central and north and south-western Africa, where no studies have been recorded in recent times. Cereals crop, particularly maize, dominates research on the integration of remote sensing and crop models, but this provides a foundation to focus research on other crops of economic interest. This also includes an increased focus on drought-tolerant crops as well as widening cropping systems, with emphasis on planting dates and fertilizer application. The study realized limited research related to integrating remote sensing into crop models for policy development. The widening of the interdisciplinary nature and scope of the studies through the involvement of social scientists, as well as agricultural economists, will improve the scope and aim

of such studies in policymaking. The application of research towards policymaking is critical for governments to steer human and financial resources toward the application of integrated crop and seasonal forecast information. This can be beneficial in smallholder farming systems, which are the most vulnerable to climate fluctuations.

For in-season yield forecasts, statistical models are simple in their usage and less parameter-intensive, but they are limited in the information they can provide outside the range of values for which the model is parameterized. They do not take into consideration the soil-plant-atmosphere continuum and the timing of the stresses occurring during the growing season and do not give farmers any important agronomic advice (e.g., timing and amount of fertilizer, time of sowing, irrigation, and so on). Crop simulation models vary greatly between them. Some of them are rather hard to use and parameterize. The need for calibration can be quite data-extensive and not applicable to some developing countries. RS techniques have been extensively used in research for yield forecast but might not be suitable in developing countries because of their stratified agricultural systems and very small farm sizes. However, this problem will be hard to overcome in the near future because of the inability of RS to estimate yield in mixed agriculture. But, the increased availability of high-spatial-resolution RS at a reasonable cost makes this technique a possible interesting alternative for yield forecast. In fact, RS is often used in Early Warning Systems in Africa. The integration of RS and CSM represents an interesting alternative in crop yield forecasting. RS can quantify crop status at any given time during the growing season in a spatial context, while CSM can describe crop growth every day throughout the season (Maas, 1988). RS indirectly can provide a measure for canopy state variables used by the CSM as well as both spatial and temporal information about those variables, which can then be used to

adjust the model simulation. However, the integration of RS and crop models for in-season yield forecasts are still rare in Africa compared to other continents. Significant investment should be aligned to this area and target in Africa.

References

Ogutu GEO, Franssen WHP, Supit I, Omondi P, Hutjes RWA. 2018. Probabilistic maize yield prediction over East Africa using dynamic ensemble seasonal climate forecasts. Agricultural and Forest Meteorology 250-251: 243-261.

Matthew OJ, Abiodun BJ, Salami AT. 2015. Modeling the impacts of climate variability on crop yields in Nigeria: performance evaluation of RegCM3-GLAM system. Meteorol. Appl. 22, 198-212.

Bahaga TK, Kucharski F, Tsidu GM, Yang H. 2015. Assessment of prediction and predictability of short rains over equatorial East Africa using a multi-model ensemble. Theor. Appl. Climatol. 637.

Schauberger B, Jagermeyr J, Gornott C. 2022. A systematic review of local to regional yield forecasting approaches and frequently used data resources. Euro J Agron. 120, 126153.

Global Commission on Adaptation, 2019. https://cdn.gca.org/assets/2019-09.

Martine R, Garanganga BJ, Kamga A, Luo Y, Mason S, Pahala J, Runmmukainen M. 2010. Regional climate information for risk management: capabilities, Procedia Environ. Sci. 1, 354-368.

Hansen JW, Mason SJ, Sun L, Tall A. 2011. Review of seasonal climate forecasting for agriculture in sub-saharan Africa. Exp. Agric. 47, 205-240.

Fritz S et al, 2019. A comparison of global agricultural monitoring systems and current gaps. Agric. Syst. 168, 258-272.

Muruganantham P, Wibowo S, Grandhi S, Samrat NH, Islam N, 2022. A systematic literature review on crop yield prediction with deep learning and remote sensing. Remote Sensing, 14, 1990.

Kitchenham BA, Charters S. Guidelines for Performing Systematic Literature Reviews in Software Engineering (EBSE 2008-001); Keele University: Keele, UK; Durham University: Durham, UK, 2007.

Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE. 2021. The PRISMA 2020 statements: An updated guideline for reporting systematic reviews. Int. J. Surg. 88, 105906.

Beyene AN, Zeng H, Wu B, Zhu L, Gebremicael TG, Zhang M. Coupling remote sensing and crop growth model to estimate national wheat yield in Ethiopia. Big Earth Dat, 6,

Ververs MT. 2012. The East African food crisis: did regional early warming sysem functions? J. Nutr. 142, 131-133.

Shin dwm Baigorria GA, Lim YK, Cocke S, LaRow TE, O'Brien JJ, Jones JW. 2010. Assessing maize and peanut yield simulations with various seasonal climate data in the southeastern United States. J. Appl. Meteorol. Climatol. 49, 592-603.

McIntosh PC, Ash AJ, Smith MS. 2005. From oceans to farms: the value of a novel statistical climate forecast for agricultural management. J. Clim. 18, 4287-4302.

McIntosh PC, Pook MJ, Risbey JS, Lisson SN, Rebbeck M. 2007. Seasonal climate forecasts for agriculture: towards better understanding and value. F. Crop. Res. 104, 130-138.

Ogutu G, Franssen W, Supit I, Omondi P, Hutjes RT, Ogutu GEO, Franssen WHP, Supits I, Omondi P, Hutjes RWA. 2017. Skill of ECMWF system-4 ensemble seasonal climate forecasts for East Africa. Int. J. Climatol. 37, 3734-2756.

Dutta D, Kundu A, Patel NR, Saha SK, Siddiqui AR, 2015. Assessment of agricultural drought in Rajasthan using remote sensing derived vegetation condition index

(VCI) and standardized precipitation index (SPI). Egyptian J. Rem. Sens. Space Sci. 18(1), 53-63.

Abebe G, Tadesse T, Gessesse B. 2022. Combined use of Landsat 8 and Sentinel 2A imagery for improved sugarcane yield estimation in Wonjishoa, Ethiopia. J. Indian Soc Remote Sensing, 50, 143-157.

Eze E, Girma A, Zenebe A, Kourouma JM, Zenebe G. 2020. Exploring the possibilities of remote yield estimation using crop water requirements for area yield index insurance in a data-scarce dryland. J. Arid Environ. 183, 104261.

Iizumi T, Luo JJ, Challinor AJ, Sakurai G, Yokozawa M, Sakuma H, Brown ME, Yamagata T. 2014. Impacts of El Nino southern oscillation on the global yields of major crops. Nat. Commun. 5, 3712.

Hansen JW, Mishra a, Rao KPC, Indeje M, Ngugi RK. 2009. Potential value of GCM-based seasonal rainfall forecasts for maize management in semi-arid Kenya. Agric. Syst. 101, 80-90.

Chitsiko RJ, Mutanga O, Dube T, Kutywayo D. 2022. Review of current models and approaches used for maize crop yield forecasting in sub-Saharan Africa and their potential use in early warming systems. Physics Chemistry of the Earth Parts A/B/C, 10399. Using crop model to forecast yield is still rare in Africa.

Kuri F, Murwira A, Murwira KS, Masocha M. 2014. Predicting maize yield in Zimbabwe using dry dekads drived from remoted sened Vegetation Condition Index. Int. J. Appl. Earth Obs. Geoinf. 33, 39-46.

Monatsa D, Nyakudya I, Mukwada G, Matsikwa H, 2011. Maize yield forecasting for Zimbabwe farming sectors using satellite rainfall estimates. Nat. Hazards 59(1), 447-463.

Davis KF, Gephart JA, Emery KA, Leach AM, Galloway JN, D'Odorico P, 2016. Meeting future food demand with current agricultural resources. Global Environ. Change 39, 125-132.

Chitsiko RJ, Mutanga O, Dube T, Kutywayo D. 2022. Review of current models and approaches used for maize crop yield forecasting in sub-Saharan Africa and their potential use in early warning system. Phy. Chem. Earth 127, 103199.

Davenport FM, Harrison L, Shukla S, Husak G, Funk C, McNally A. 2019. Using out-of-sample yield forecast experiments to evaluate which earth observation products best indicate end of season maize yields. Environ. Res. Lett. 14, 124095.

Eze E, Girma A, Zenebe A, Kourouma JM, Zenebe G. 2020. Exploring the possibilities of remote yield estimation using crop water requirements for area yield index insurance in a data-scarce dryland. J. Arid. Environ. 183, 104261.

Luciani R, Laneve G, Jahjah M. 2019. Agricultural monitoring, an automatic procedure for crop mapping and yield estimation: the great rift valley of Kenya case. IEEE J, 12, 2196-2208.

Abebe G, Tadesse T, Gessesse B. 2022. Assimilation of leaf area index from multisource earth observation data into the WOFOST model for sugarcane yield estimation. Int. J. Remote Sensing, 43, 698-720.

Ma, G.N., Huang, J.X., Wu, W.B., Fan, J.L., Zou, J.Q., Wu, S.J., 2013. Assimilation of MODIS-LAI into the WOFOST model for forecasting regional winter wheat yield. Math. Comput. Model. 58, 634-643.

Zhao, Y., Chen, S., Shen, S., 2013. Assimilating remote sensing information with crop model using Ensemble Kalman Filter for improving LAI monitoring and yield estimation. Ecol. Model. 270, 30-42.

Dong, Y., Zhao, C., Yang, G., Chen, L., Wang, J., Feng, H., 2013. Integrating a very fast simulated annealing optimization algorithm for crop leaf area index variational assimilation. Math. Comput. Model. 58, 877–885.

Huang, J., Ma, H., Su, W., Zhang, X., Huang, Y., Fan, J., Wu, W., 2015. Jointly assimilating MODIS LAI and ET products into the SWAP model for winter wheat yield estimation. IEEE JSTAR 8, 4060–4071.

Liu, F., Liu, X., Zhao, L., Ding, C., Jiang, J., Wu, L., 2015. The dynamic assessment model for monitoring cadmium stress levels in rice based on the assimilation of remote sensing and the WOFOST model. IEEE JSTAR 8, 1330–1338.

Jin Z, Azzari G, Burke M, Aston S, Lobell DB. 2017. Mapping smallholder yield heterogeneity at multiple scales in Eastern Africa. Remote Sens. 9, 931.

Schwalbert R, Amado T, Nieto L, Corassa G, Rice C, Peralta N, Schauberger B, Gornott C, Ciampitti I. 2020. Mid-season county-level corn yield forecast for US corn belt integrating satellite imagery and weather variables. Crop Sci 60: 739-750.