ISRS

RESEARCH ARTICLE

Check for updates

Assimilation of Remote Sensing Data into Crop Growth Model for Yield Estimation: A Case Study from India

Murali Krishna Gumma[1] · M. D. M. Kadiyala[2] · Pranay Panjala[1] · Shibendu S. Ray[3] ·
Venkata Radha Akuraju[1] · Sunil Dubey[3] · Andrew P. Smith[1] · Rajesh Das[4] · Anthony M. Whitbread[5]

Abstract
Crop yield estimation is important to inform logistics management such as the prescription of nutrient inputs, financing, storage and transport, marketing as well as to inform for crop insurance appraisals due to loss incurred by abiotic and biotic stresses. In this study, we used a suite of methods to assess yields at the village level ($< 5 \text{ km}^2$) using remote sensing technology and crop modeling in Indian states of Telangana, Andhra Pradesh and Odisha. Remote sensing products were generated using Sentinel-2 and Landsat 8 time series data and calibrated with data collected from farmers' fields. We derived maps showing spatial variation in crop extent, crop growth stages and leaf area index (LAI), which are crucial in yield assessment. Crop classification was performed on Sentinel-2 time series data using spectral matching techniques (SMTs) and crop management information collected from field surveys along with ground data. The locations of crop cutting experiments (CCEs) was identified based on crop extent maps. LAI was derived based on the SAVI (soil-adjusted vegetation index) equation were using Landsat 8-time series data. We used the technique of re-parametrization of crop simulation models based on the several iterations using remote sensing leaf area index (LAI). The data assimilation approach helps in fine-tuning the initial parameters of the crop growth model and improving simulation with the help of remotely sensed observations. Results clearly show a good correlation between observed and simulated crop yields (R^2 is greater than 0.7) for all the crops studied. Our study showed that by assimilation of remotely sensed data in to crop models, crop yields at harvest could be successfully predicted.

Keywords Yield assessment · Crop classification · Leaf area index (LAI) · Remote sensing · Crop cutting experiments (CCEs) · Sentinel-2 · Landsat 8 and India

✉ Murali Krishna Gumma
m.gumma@cgiar.org; gummamk@gmail.com

1 RS/GIS Lab, Innovation Systems for the Drylands,
   International Crops Research Institute for the Semi-Arid
   Tropics (ICRISAT), Patancheru, Hyderabad 502 324, India

2 Acharya N. G. Ranga Agricultural University (ANGRAU),
   Lam, Andhra Pradesh, India

3 Mahalanobis National Crop Forecast Centre (MNCFC),
   Ministry of Agriculture, Delhi, New Delhi, India

4 Directorate of Agriculture & Food Production, Govt of
   Odisha, Bhubaneswar, Odisha, India

5 International Crops Research Institute for the Semi-Arid
   Tropics (ICRISAT), Dar es Salaam, Tanzania

Introduction

In India, rice, maize and groundnut are the main rainy season crops that contribute significantly to India's economy. Most of this cropped area is rainfed and these areas have low grain yields due to poor distribution of monsoon rainfall in many years. Though rainfed areas have low grain yields, they hold a lot of potential to increase production (Aggarwal et al., 2008; Lobell et al., 2009; Stuart et al., 2016). Timely and accurate estimation of crop yields is essential for effective agricultural land management, policy-making and sustainability of agricultural food production (Masson-Delmotte et al., 2018). The precise and timely monitoring of potential yields is crucial for decision making as it influences markets, export–import decisions and farm income budgeting (Zhao et al., 2020). Yield

✎ Springer

estimation also plays a critical part in the pricing of insurance premiums and serves as a promising financial risk management solution for smallholder farmers (Choudhury & Jones, 2014). In recent years, several methods are used for prediction of crop yields at a range of scales such as empirical equations, remote sensing and simulation models. Although there are studies validating the use of remote sensing data in estimating crop yields, most have used empirical methods for specific locations, crops, cultivars and crop growth stages. Similarly, researchers used crop models to simulate crop growth and yields at a regional scale. Every technology has its advantages and limitations. For example, the use of remote sensing for temporal crop analysis is limited by the availability of cloud-free time series remote sensing data and difficulties in accurately estimating LAI (Batchelor et al., 2002; Quarmby et al., 1993). Similarly, the adequate simulation of cropping systems is often hindered by lack of sufficiently accurate input data such as parameter values for soil, management, cultivar and meteorological inputs. However, combination of these two approaches has the potential to update and, where necessary, rectify cropping system model simulations for more certain outcomes. (Lobell et al., 2015; Pitman, 2000; Potgieter et al., 2014; Schut et al., 2009; Nearing et al., 2012).

Remote sensing data assimilation methods with various levels of complexity have been tried, either by directly using remote sensing satellite data in simulation models (Doraiswamy et al., 2003; Olioso et al., 2005), by updating state variables or by re-parametrization of the model using remote sensing satellite data (Fang et al., 2011; Jin et al., 2017). Biophysical parameters are very important for environmental analysis as leaf area index (LAI) is one of the important parameters that reflects crop growth stages of vegetation and is widely used in quantitative analysis of crop models (Parker, 2020; Yan et al., 2019). LAI shows the amount of leafs area per unit in two dimensions and is a key parameter for biophysical modeling (Asner et al., 2003; Ren et al., 2014; Stark et al., 2012; Yan et al., 2019). Several studies have extracted biophysical parameters from satellite imagery (Hui & Yao, 2018; Yu et al., 2019) and assimilated them in simulation models. Kang and Özdoğan (2019) successfully assimilated Landsat-derived LAI time series into crop model simulations using ensemble Kalman filter for individual fields or pixels. In this study, we used the technique of re-parametrization of crop simulation models through several iterations using remotely sensed LAI estimates; this technique is supposed to best integrate crop growth conditions. The data assimilation approach helps with initializing parameters of the crop growth model and improves simulations with the help of remotely sensed satellite observations.

The Government of India's Pradhan Mantri Fasal Bima Yojana (PMFBY) program aims at supporting sustainable production in the agriculture sector by providing financial support, stabilizing farmers' incomes and encouraging farmers to adopt modern techniques. The scheme provides insurance to indemnify farmers against crop losses. The Mahalanobis National Crop Forecast Centre (MNCFC) in New Delhi, India, estimates the acreage, yield of crop and crop conditions across the country which helps insurance companies set crop premiums. As a part of this initiative, MNCFC has supported pilot studies to assess quickly and accurate yield estimations throughout India using techniques such as remote sensing and crop simulation modeling.

To contribute to this initiative, we adopted the methodology developed by Fang et al. (2011) for crop yield assessment at the village level (Gram Panchayat) using remote sensing, field data and crop simulation models. Our major objectives are (i) to identify suitable locations for CCEs using crop type mapping and to conduct CCE; (ii) to perform crop yield assessment by assimilating the remote sensing data into crop simulation models; and (iii) to validate the simulation results with CCE yield data and statistics from agriculture departments.

Study Area

The selected study areas are located in the states of Andhra Pradesh, Telangana and Odisha in the southern and eastern (Fig. 1). The following districts were chosen for this study: Anantapur, Krishna and Kurnool in Andhra Pradesh; Mahbubnagar in Telangana; and Puri in Odisha. The study areas are located in two distinct agro-climatic zones such as semiarid regions (Anantapur, Kurnool and Mahbubnagar) and humid and dry sub-humid regions (Krishna and Puri). Most areas in the semiarid regions depend on rainfall for agriculture. Mahbubnagar and Anantapur districts receive an average annual precipitation of 650 mm, whereas other districts receive more than 1000 mm.

In the humid districts Krishna and Puri, rice was selected as the crop occupies a major area, whereas in the semiarid regions of Anantapur, Kurnool and Mahbubnagar districts, groundnut and maize were considered for this study. In each district, five *mandals* (small administrative units) were selected for CCEs (Table 1).
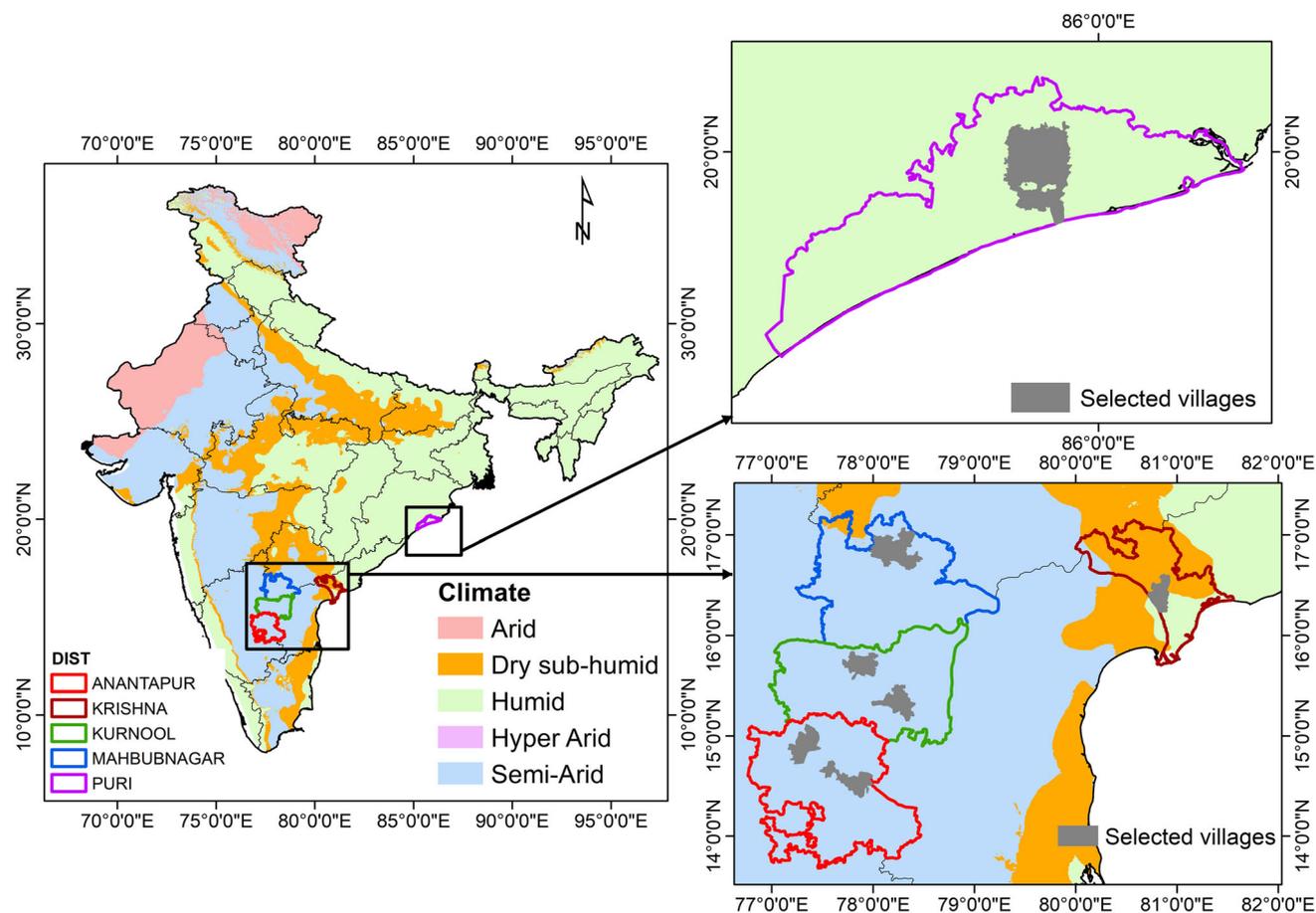
Fig. 1 Location of the pilot study sites and their Climatic zones

Table 1 Crop-wise information for selected study districts

State	District	Type of irrigation	Crop	Crop varieties	Season
Andhra Pradesh	Krishna	Canal irrigation	Rice	Medium to long duration cultivars of 135–150 days	June–November
	Anantapur	Rainfed/groundwater	Groundnut	K-6, JL-24 and Dharani	July–November
Telangana	Mahbubnagar	Rainfed/groundwater	Maize	DHM-117; Dekalb9144	July–October
Odisha	Puri	Canal irrigation	Rice	Medium to long duration cultivars of 135–150 days	July–November

Datasets

Satellite Imagery

We used Sentinel-2 and Landsat 8 data for crop type classification and LAI estimation. Landsat tiles were downloaded from the US Geological Survey (USGS) global land cover facility website (https://earthexplorer.usgs.gov/) for the 2019–20 rainy season (June to October). Sentinel-2 satellite data were downloaded from the European Space Agency (ESA), i.e., Copernicus open access hub (https://sentinel.esa.int/web/sentinel/). Landsat 30 m 16-day cloud-free imagery and Sentinel-2 10 m 12-day

processed satellite images were downloaded from cloud computing Google Earth Engine (GEE) for rainy 2019–20.

Multispectral data from the Sentinel-2 satellite constellation, which is part of the European Copernicus program, were used in this study. It comprises a set of two satellites Sentinel-2A and 2B—that provide high-resolution data, with a high combined revisit time of 5 days at the equator, for monitoring land surface conditions.

Sentinel-2 satellite provided high spatial resolution (optical bands at 10 m) images over a large swath (290 km), high spatial resolution (optical bands at 10 m), and provide data in the optical, near-infrared (NIR) and

short-wave infrared (SWIR) regions of the electromagnetic spectrum.

Ground Data for Crop Classification

Ground data for crop classification were collected during the 2019–20 rainy season (August–October) for 725 sample data points covering about 5000 km of road travel in the selected districts (Fig. 2, Table 2). A stratified systematic sample design was adopted to select the location by considering croplands and road network; data were collected every 3–5 km (Thenkabail et al., 2004, 2005). Ground data samples were acquired within large homogenous areas of a particular land use/land cover (LULC).

Ground data were collected based on preliminary crop classification and near real-time satellite imagery, i.e., Sentinel-2 false color composites with tracking GPS using image processing software. The ground data were collected in a 30 m × 30 m plot and included location, LULC categories, crop type and cropping pattern, methods of irrigation, farmers' interviews (wherever possible), etc. Crop name and location data were collected at each point to validate crop type classification.

Two independent datasets were collected: one for training and another for validation. Table 1 shows sample size of training data for classification and validation data for accuracy assessment in each study sites. Of 725 locations, we used 311 samples for crop classification as training data and 414 samples for validation.

Crop Cutting Experiments (CCEs) for Yield Verification

In the current analysis, we assisted the Government of India in identifying potential locations to conduct CCEs in Krishna, Anantapur and Kurnool districts in Andhra Pradesh, Puri district in Odisha and Mahbubnagar in Telangana to test the methodology. The crop type map overlaid with the NBSSLUP soil map was used to identify the CCE locations. GPS location, date of sowing, irrigated vs rainfed and other crop management details were collected at the locations.

Landsat 8 LAI Data Processing

Leaf area index and soil-adjusted vegetation index (SAVI) were calculated using Eqs. (1) and (2) using Landsat 8 time series surface reflectance data (Schmidt et al., 2013)(Huete, 1988; Qi et al., 1994; Reyes-González et al., 2019; Rondeaux et al., 1996; Thenkabail et al., 2000) downloaded from GEE. In this study, we used the unique spectral response of leaves compared to other parts of the plant. We integrated SAVI and LAI calculated using the following empirical equations.

$$LAI = \frac{-\ln\left(\frac{0.69-SAVI}{0.59}\right)}{0.91} \quad (1)$$

For Landsat images used in this study, SAVI was computed using the formula:

$$SAVI = \frac{(1+L)(B5 - B4)}{L + B5 + B4} \quad (2)$$

Fig. 2 Location of ground data sample points for crop classification and validation
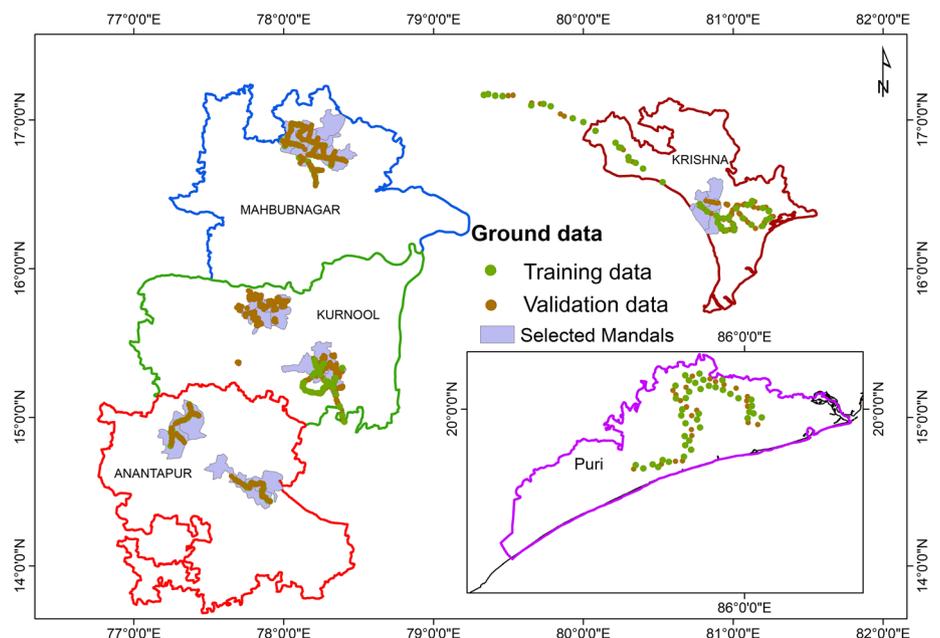
Table 2 Ground data for training and validation of crop classification along with ground data and CCEs collection

District	Training data	Validation data	Major crops sampled	Ground data collection (week)	CCEs (week)
Krishna	43	69	Rice	2nd week of September	2nd week of November
Anantapur	44	91	Groundnut	1st week of September	1st week of November
Mahbubnagar	44	99	Maize	3rd week of September	3rd week of October
Puri	39	23	Rice	2nd week of September	2nd week of November

(where L is a soil factor, taken to be 0.1, B5 is the spectral reflectance in near infrared, and B4 is spectral reflectance in red.)

Soil Data

Soil data for each CEE location used to parameterize the biophysical crop simulation models were obtained from a set of soil profile data available from the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) data repository and the National Bureau of Soil Survey and Land Use Planning (NBSSLUP) databases. Soil properties such as hydraulic conductivity, bulk density and soil texture were extracted from these soil profile data repositories. Additional parameters such as drained upper limit, lower limit and saturation limits were estimated based on soil texture using the generic soil database available in the DSSAT models using pedo-transfer functions. We also used certain parameters such as soil fertility factor (SLPF) in soil as free variables, meaning that they could be modified as required.

Weather Data

Weather data such as daily maximum and minimum temperature, rainfall and solar radiation were collected from the Andhra Pradesh Development Planning Society (APSDPS) and Telangana State Development Planning Society (TSDPS) automatic weather station network available near by the testing sites. Weather parameters for Odisha sites were collected from Odisha State Agriculture Department. Weather data were collected for the period June to December 2019.

Methods and Approaches

The overview of methodology starts with three major input parameters (Fig. 3), mainly present daily weather data, crop management data and remote sensing products (crop classification and leaf area index (LAI).

Crop Type Maps

The methodology for mapping the rainy cropping pattern using Sentinel-2 time series and ground data was taken from Gumma et al. (2020b). Crop type classification starts with time series data composition, ideal spectral generation and class identification based on spectral matching techniques (SMTs). Ideal spectra profiles of each crop class were generated using the Sentinel-2 normalized difference vegetation index (NDVI) temporal data. The basis for SMTs is the generation of ideal temporal NDVI (Tucker, 1979) spectra for each LULC class from the ground survey data. The ground data used for ideal spectra generation were collected from homogeneous patches along with in-depth information about the cropping system and irrigation methods. Differences in the phenological growth stages of different plants are reflected in the temporal NDVI profiles, since NDVI can measure growth conditions (greenness of vegetation) (Belgiu & Csillik, 2018; Croitoru et al., 2012).

In the present study, GEE was used for satellite data processing and classification (Gumma et al., 2020a). The Sentinel-2 (S2) collection of GEE is a set of UINT 16-bit images having 13 spectral bands with top-of-atmosphere (ToA) reflectance values scaled by 10,000, as well as three quality assurance (QA) bands. These data are available from 23 June 2015 to the present. Out of the 13 spectral bands, 4—B2, B3, B4 and B8 had a spatial resolution of 10 m; 6—B5, B6, B7, B8a, B11 and B12 had a spatial resolution of 20 m; and 3—B1, B9 and B10 had a spatial resolution of 60 m. GEE was chosen to process and classify the S2 data because the quality of data is assured with cloud masks and user-friendly functions for image processing and classification. Moreover, the main advantage of GEE is its cloud computing capability which reduces the need for the client to have hardware and software capabilities.

Sentinel 2 images for the 2019 rainy cropping season for the months August to December were used for the analysis. The monthly maximum value composites (MVCs) of NDVI were computed and stacked. A time series NDVI MVCs were used to detect phenological stages of vegetation, thus made it, possible to identify the type of vegetation/crop. The stacked NDVI composites were passed

Fig. 3 The data assimilation approach combines remote sensing data with crop growth models to estimate crop yield
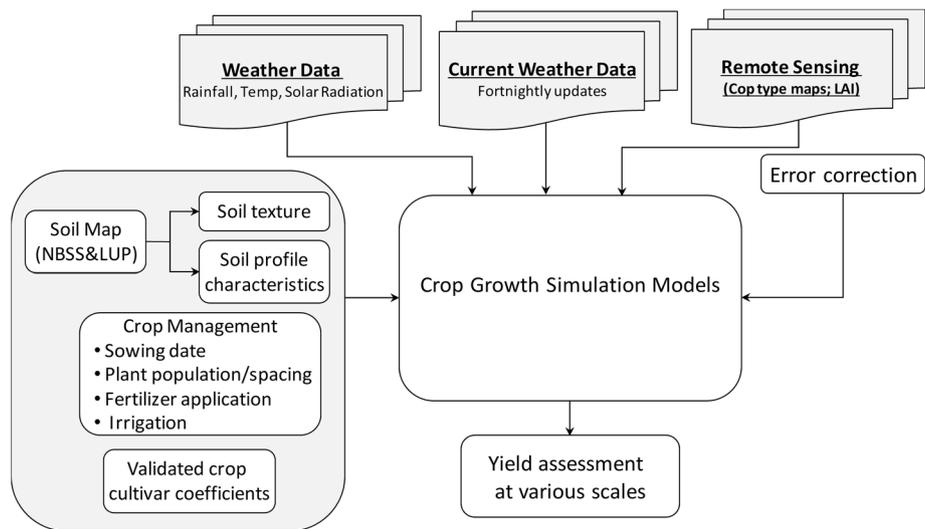
through an unsupervised classifier. The K-means classifier of GEE was used to generate 70 clusters or classes of similar pixels using machine learning algorithm. The NDVI MVC time series spectra of these classes were plotted and compared. Classes with similar spectra were grouped using spectral similarity values (SSVs). Lower SSVs indicate high similarity between two classes. The classes were then identified and labeled using SMT.

Intensive ground truth data were used for class identification. The ideal NDVI MVC spectra for each crop were identified with 90 samples selected from the ground survey data. The sample data helped to effectively identify all the major crop types of the region, and to represent a homogeneous area large enough to avoid confusion while determining the crop type using the time series spectra. The ideal spectra for each class were taken as the average of the spectral curves of all locations of that class. Class spectra were matched with the ideal spectra and labeled with the land use of the matching spectra. Ground data and Google Earth high-resolution images were used to improve class identification.

A large and fragmented landscape dominates almost all parts of India, leading to mixed classes due to unsupervised classification. Such mixed classes were reclassified by masking out and relabeling them using SMT.

Accuracy Assessment of Crop Type Maps

Accuracy assessment of the maps was performed based on validation data. A total of 414 ground survey samples(Table 2) were used to assess the accuracy of the classification map (Jensen, 2004), by generating a confusion matrix, wherein columns represent field-plot data points, and rows represent results of classified rice maps in the confusion matrix (Congalton, 1991). The confusion matrix contains corresponding class changes in a multidimensional table. The statistical approach of accuracy assessment shows multivariate statistical analyses such as Kappa (Cohen, 1960) to relate results from different classifications and regions; it is a degree of agreement between user and reference ground data with a score of homogeneity, or consensus.

Crop Cutting Experiment Procedure

CCEs were carried out based on optimization techniques. The optimization of CCE points was carried out with the help of NDVI, climatic zones, soil data and related classification maps (Fig. 4). Firstly, the strata of NDVI, climatic zones and soil were prepared and combined to identify homogeneous strata. Then collected random sample plots over the homogeneous strata with proportion to sample size PPS (probability proportion to the sample size) such that each stratum contains minimum 30 plots. Based on multiple regression-based model, the numbers of samples were reduced up to significance level. Thus, the optimal number of sample points was selected.

Locations for CCEs were selected using crop type maps, and experiments were conducted in a 5 m × 5 m (as per NSSO Guidelines) homogenous area identified in the selected crop field. The entire harvested crop or subsample weight was recorded. Weight of the biomass and grains by threshing the harvested crop was recorded. Photographs and GPS location of the experiment plot were taken for reference and validation purposes. Crop management data were also collected by interacting with the farmers.

Fig. 4 Flow diagram of the optimization of CCEs
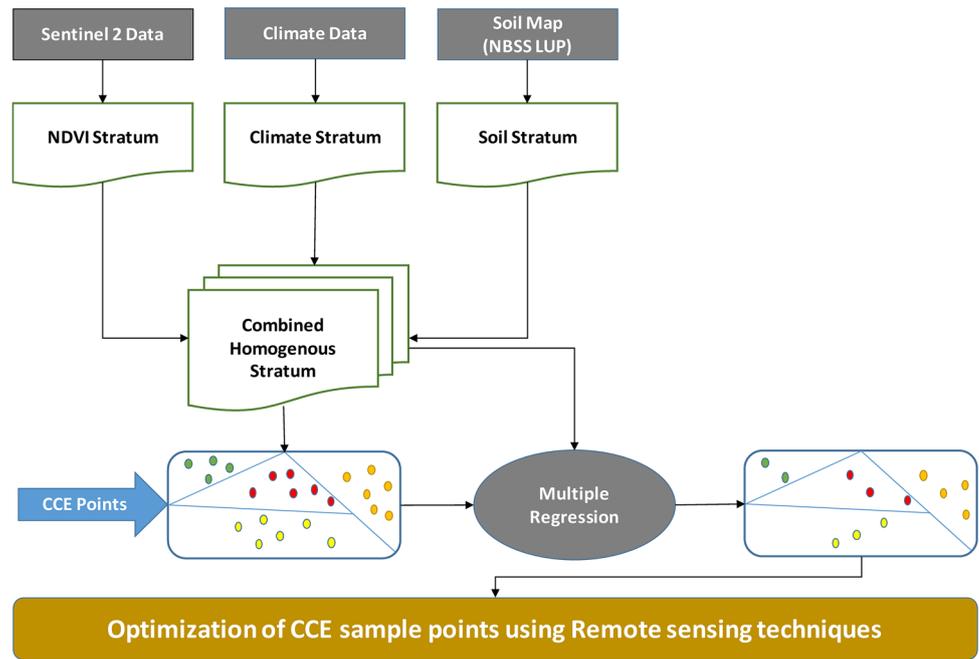
The cropping system model

The cropping system models (CSM) CERES for rice, maize; PEANUTGRO for groundnut, available in the Decision Support System for Agrotechnology Transfer (DSSAT), were used for yield simulations (Jones et al., 2003). Daily biophysical crop information (e.g., LAI, biomass) was generated with the help of the crop growth model. Input data such as soil conditions, weather and management data (planting date, seed rate and fertilizer application rates) were prepared for each CCE location. The most common cultivar types representing a major part in the CCE location were used. Cultivar parameters were sourced from the ICRISAT data bank and from published literature. If the required cultivar parameters were not available, they were matched with cultivar of a similar duration. Nutrient supply, mainly nitrogen, was set as a free variable. The basal application dose was matched with planting date, and the remaining application rates and times were set as free variables.

The optimization process starts with initial model parametrization by adjusting the free parameters so that the model-simulated LAI is in agreement with the Landsat 8 LAI observations (Eq. 3). The simulated LAI values depend on the values of the free variables (e.g., planting date, nitrogen dose, soil profile parameters) that are generated by minimizing the value of the following cost function (Myneni et al., 2002). The remote sensing LAI data were collected for six times during the crop growth period.

$$\sum_{i=1}^{m} \mathrm{abs}[\mathrm{LAI}_S(t_i) - \mathrm{LAI}_M(t_i)]/\mathrm{LAI}_M(t_i) \qquad (3)$$

(where LAI_M and LAI_S are the leaf area index measured and leaf area index simulated at time t_i, respectively.)

By adjusting the free input variables in order to minimize the residuals between Landsat 8 LAI and simulated LAI at each corresponding period (six observations), an optimized set of input parameters were obtained. The free variables were adjusted sequentially based on the cost function (Eq. 3) for each village location. After completing the maximum number of iterations with minimum merit function value, the free parameters were finalized. Finally, with the optimized parameters, the model was executed to update the crop yields. Crop yields were estimated for each CCE location and compared with the observed yields of that particular CCE yield data.

Results

Spatial Distribution of Crops

Figure 5 shows the spatial distribution of the various crops across the selected districts. Groundnut observed to be the major crops in Anantapur district, whereas groundnut occupies nearly 75% of the total crop area. Groundnut was considered for yield assessment in Anantapur district. Sorghum and rice are the major crops in Kurnool district. Maize, millet, pigeonpea and cotton are the four major crops in Mahbubnagar district. Maize and pigeonpea were
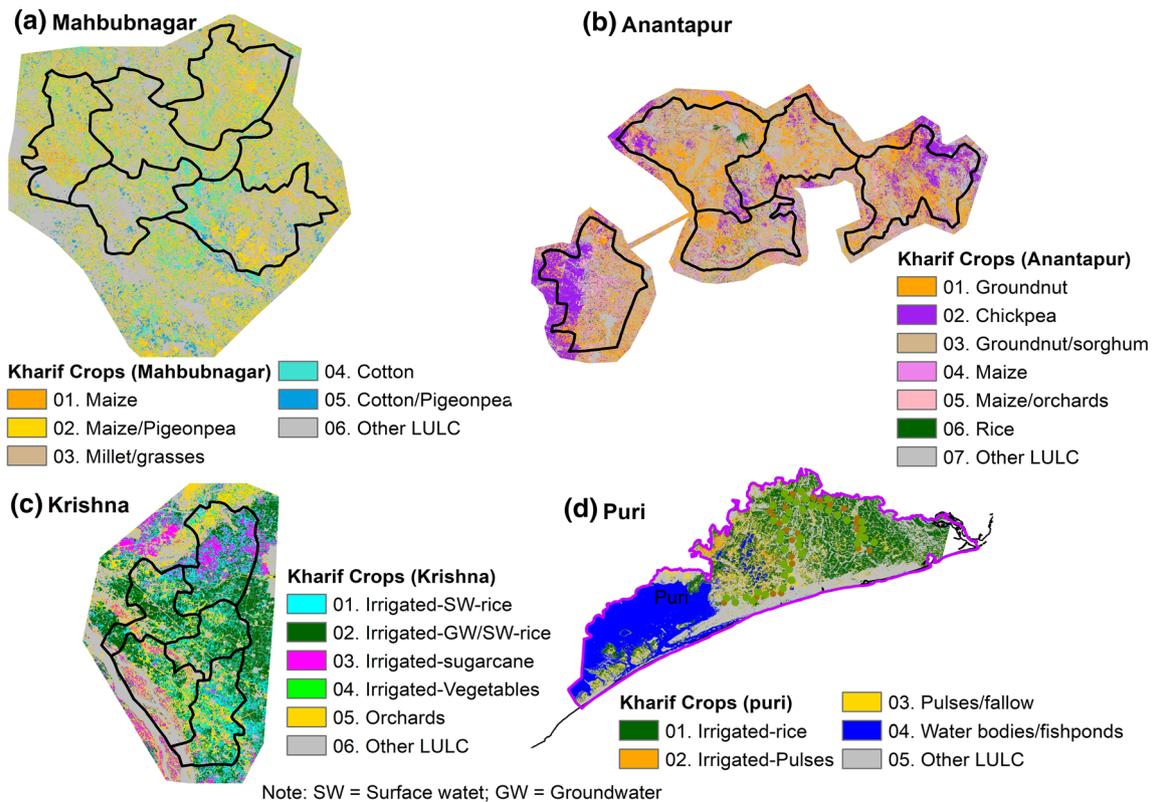
Fig. 5 Crop classification maps of Mahbubnagar, Anantapur, Krishna and Puri districts

widely spread across the chosen mandals, occupying nearly 60% of the total crop area. Maize was considered for yield assessment in Mahbubnagar district. Rice is the major crop across the Krishna and Puri districts occupying nearly 75% and 80% of total crop area, respectively. Most of the selected crops in this study were irrigated crops. Rice was considered for yield assessment in both Krishna and Puri districts.

An assessment of the accuracy of crop type classification showed an overall accuracy of more than 80% in each district (Table 3). In Kurnool, major crops sorghum and rice achieved 86% accuracy, whereas crop type maps achieved more than 90% in both user's and producer's accuracies. In Mahbubnagar, of the major crops cotton and pigeonpea assessed for accuracy maize achieved over 85% in both accuracies. In Puri, rice is the dominant crop and achieved 78% accuracy. In Krishna district, irrigated rice achieved nearly 86% accuracy, whereas in Anantapur, of the two crops chosen to assess accuracy, groundnut achieved more than 88% in both accuracies.

Crop Cutting Experiments

Crop cutting experiments were conducted based on the finalized crop type map. The location of the CCE sites was determined using crop type map information, rain gauges'

availability, soil types and area statistics of each crops in particular location. Using this information, a representative set of CCE locations was finalized (Figs. 6 and 7). The crop type map contains information about crop type as well as crop growth periods. Based on crop growth periods and the crop, harvesting windows were identified, based on which CCEs were carried out from October to December with the help of local agriculture staff.
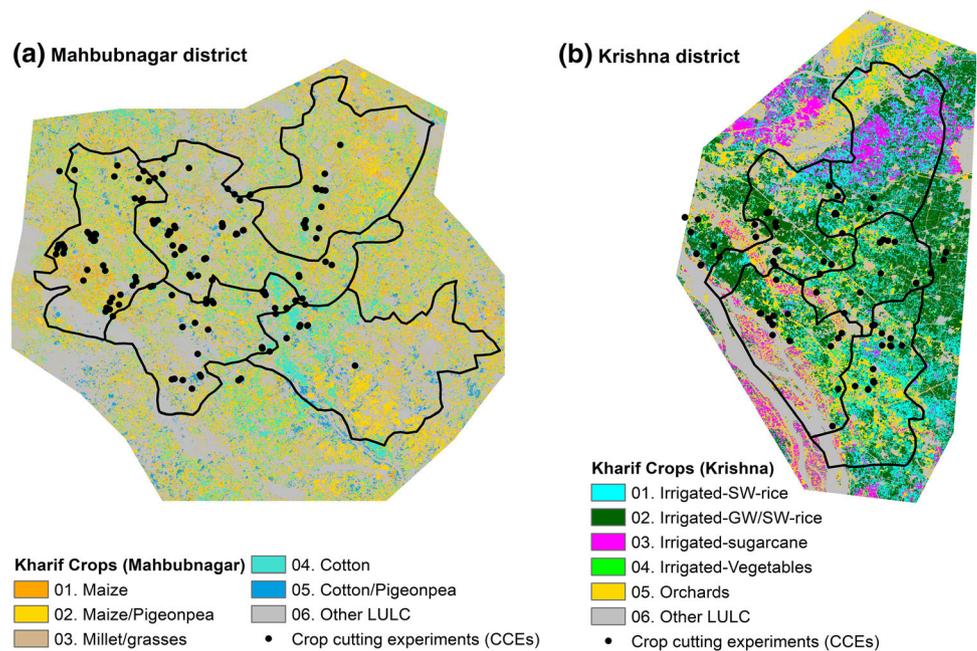
Crop Yields and Limitations

The mean, variance, standard deviation and coefficient of variation were calculated for CCEs collected for crops in their respective districts (Table 4). However, CCE plots were identified using remote sensing techniques along with various parameters such as crop area extent, proximity to roads and accessibility. The CCEs could not be conducted in the proposed locations as farmers were unwilling to harvest the crop in their fields. Neither could we collect information on crop management such as cultivar sown, date of sowing, fertilizer application and dates of irrigation from all the locations.

Table 3 Accuracy assessment of the study sites based on ground reference data

District	Crops	User's accuracy	Producer's accuracy	Overall accuracy
Kurnool	Chickpea	0.93	0.93	0.87
	Sorghum	0.76	0.76	
	Rice	0.84	0.84	
Mahbubnagar	Maize	0.96	0.86	0.85
	Cotton	0.71	0.75	
	Pigeonpea	0.74	0.87	
Puri	Rice	1.00	0.78	0.78
Krishna	Irrigated-SW—rice	0.60	1.00	0.87
	Irrigated-GW—rice	1.00	0.83	
Anantapur	Groundnut	1.00	0.88	0.87
	Millet	0.67	0.80	
	Chickpea	0.33	1.00	

Note: SW = surface water; GW: Groundwater

Fig. 6 Locations of the crop cutting experiments (CCEs) suggested, using remote sensing and other ancillary data

(a) **Mahbubnagar district**

(b) **Krishna district**

Kharif Crops (Mahbubnagar)
- 01. Maize
- 02. Maize/Pigeonpea
- 03. Millet/grasses
- 04. Cotton
- 05. Cotton/Pigeonpea
- 06. Other LULC
- • Crop cutting experiments (CCEs)

Kharif Crops (Krishna)
- 01. Irrigated-SW-rice
- 02. Irrigated-GW/SW-rice
- 03. Irrigated-sugarcane
- 04. Irrigated-Vegetables
- 05. Orchards
- 06. Other LULC
- • Crop cutting experiments (CCEs)

Rice Yields

In Krishna district, CCEs were performed in 94 locations. Observed yields were compared with simulated yields from the crop model with Landsat 8 LAI assimilated data. The observed yields were scattered between 3560 and 7123 kg/ha and the simulated yields clustered between 3697 and 6,213 kg/ha. Observed yields of rice were found to be in proximity to the simulated grain yield over the several CCE locations (Fig. 7). Similarly, observed yields in various CCE locations across Puri district were clustered between 260 and 7684 kg/ha, while simulated yield was scattered between 1154 kg/ha and 6514 kg/ha. It was also observed that yield simulations in rice improved with greater number of LAI observations. Figure 8 compares observed rice yields with yields simulated under different locations with Landsat 8 derived LAI data assimilation methods. The predictions were strong with a R^2 value of 0.76 in Puri district and R^2 value of 0.71 in Krishna district. The root-mean-square error (RMSE) of the yield was 696 kg/ha Puri district and 386 kg/ha, respectively, in Krishna districts (Table 5). The accuracy of simulations was significantly higher in Krishna district than in Puri district mainly due to lack of information on soil and cultivars information for Puri district.
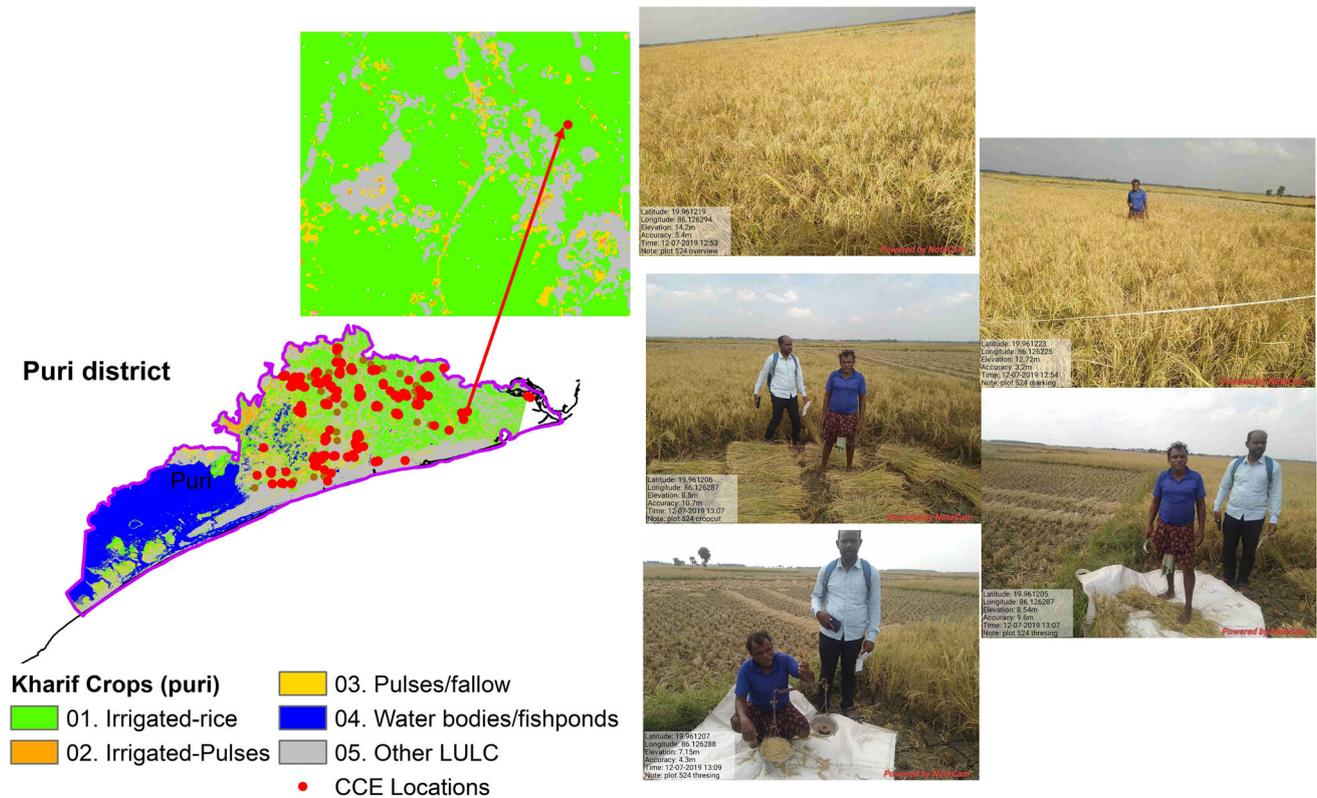
Fig. 7 An illustration of CCEs on the rice crop in Puri district, Odisha

Table 4 Statistics computed from CCEs in four districts

District	Crop	Mean (kg/ha)	Variance	Standard deviation	Coefficient of variation (%)
Krishna	Rice	5539	360,218.1	600.2	0.11
Puri	Rice	3555	2,559,211	1599.7	0.45
Mahbubnagar	Maize	2189	1,025,800	1012.8	0.46
Anantapur	Groundnut	1187	70,347.36	265.2	0.22

Maize and Groundnut Yields

Observed maize and groundnut yields were compared with simulated yields under different crop management options in each CCE location (Fig. 9). Observed yields in various CCE locations across maize growing locations in Mahbubnagar district ranged between 212 and 5180 kg/ha, and simulated yields between 399 and 4,533 kg/ha. The descriptive statistics are presented in Table 6. It was observed from the yield predictions that assimilation of LAI data from remote sensing products improved yield predictions, especially under rainfed conditions, as the model could simultaneously improve both extreme yields in all the crops studied.

Discussion

This study investigated the potential of combining high-resolution LAI data with crop modeling to assess crop yields at field scale. In the approach used, the initial step involved mapping different crops using high-resolution images with high accuracy. These crop extent maps were very useful in identifying crop types as well as in understanding the crop conditions. We also used crop masking as the base map to identify proposed locations to perform CCEs. The study area mainly comprised smallholder farmers, the sowing dates differed among them. The crop extent map enabled us to identify crop growth stages using NDVI decision tree algorithms and plan the timely
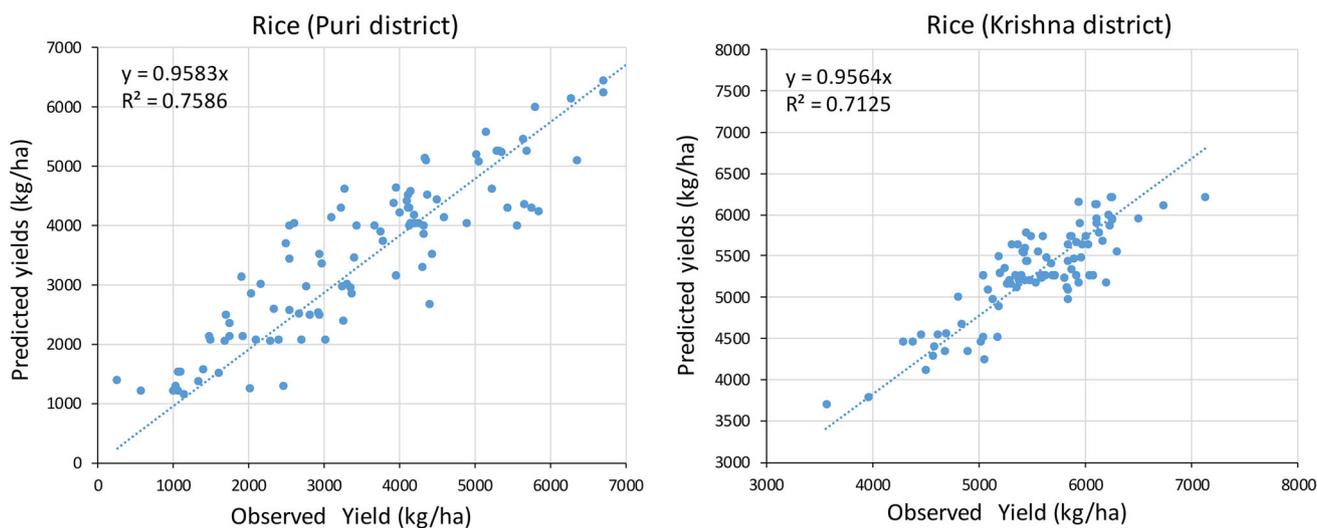
Fig. 8 A comparison of simulated rice yields and observed yields in Puri and Krishna districts

Table 5 Descriptive statistics of the performance of CERES-rice model assimilated with remote sensing LAI data in Puri and Krishna districts

Districts	Variables						
	CCE locations	Observed (kg/ha)	Simulated (kg/ha)	RMSE (kg/ha)	RRMSE	MAE	ME
Puri	101	3555	3537	696	0.19	541	0.76
Krishna	94	5539	5307	386	0.08	307	0.71

RMSE root-mean-square error, *RRMSE* relative root-mean-square error, *MAE* mean absolute error, *ME* modeling efficiency
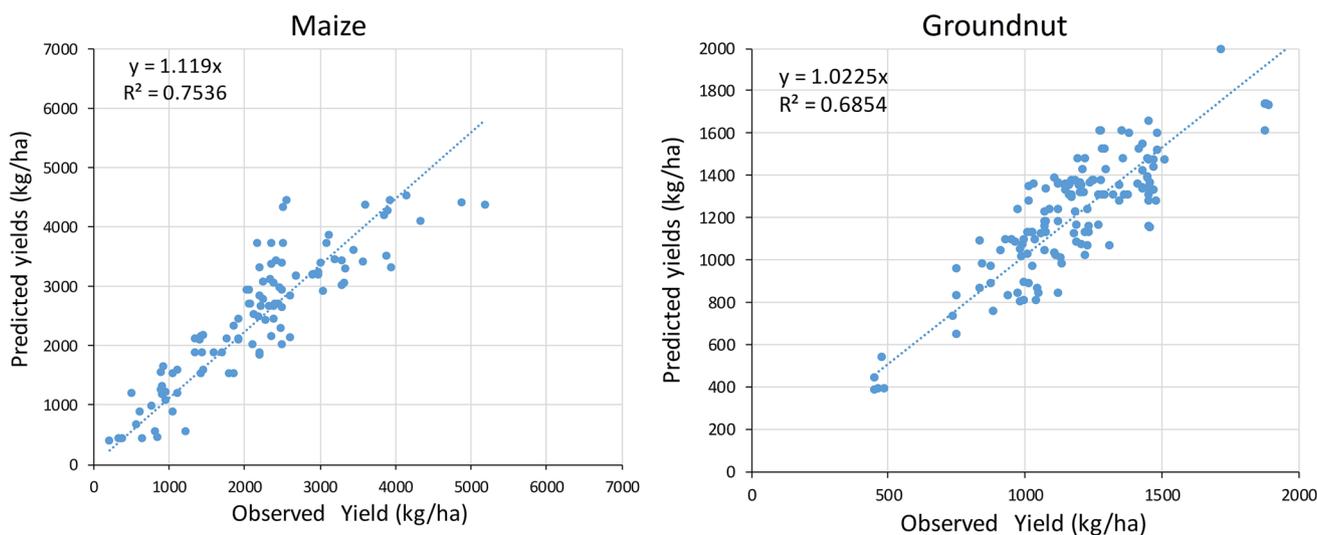

Fig. 9 A comparison of the simulated yields and observed yields of Maize and Groundnut crops

harvesting of crops in different CCEs. Timely harvesting is key for enhanced accuracy in yield assessments.

Various methodologies have been used to estimate LAI in wheat, rapeseed and potato using remote sensing products (Clevers et al., 2017; Li et al., 2017; Wei et al., 2017)

and to estimate LAI and biomass in corn and soybean using remote sensing multispectral data (Kross et al., 2015). This study investigated the potential of assimilating time series Landsat 8-LAI data into the DSSAT crop growth model for better prediction of crop yields at each CCE location.

Table 6 Descriptive statistics illustrating the performance of CERES-maize and PEANUTGRO groundnut model assimilated with remote sensing LAI data

Crops	Variables						
	CCE locations	Observed (kg/ha)	Simulated (kg/ha)	RMSE (kg/ha)	RRMSE	MAE	ME
Maize	105	2245	2580	608	0.23	486	0.75
Groundnut	148	1191	1290	157	0.08	134	0.69

RMSE root-mean-square error, *RRMSE* relative root-mean-square error, *MAE* mean absolute error, *ME* modeling efficiency

Using a cost function measuring the distance between the simulated state variables and observed ones, the method employed automatically adjusts the set of model input parameters until the difference between the Landsat 8 LAI and the crop model-simulated LAI is minimized(Fang et al., 2011). Finally, using this optimization algorithm, crop yields were predicted at each CCE location by obtaining a new set of parameters or initial values and allowing a simulation that resembles better observations. The technique we used was a frequently applied re-calibration methodology that enabled us to estimate the yields of rice, maize and groundnut successfully and compare them with observed yields with significant accuracy at each CCE location. The data assimilation approach proved to be reliable and shows great potential in providing yield prediction data at the village level. In this study, since LAI is the only link between the crop growth model and remotely sensed data, the accuracy of the model and final predictions with optimized datasets depends on the quality of remotely sensed LAI data. This study also refined various input parameters, especially nitrogen application rates and timing, by performing various iterations of matching model-simulated LAI with remote sensing outputs as LAI is significantly influenced by nitrogen levels.

With recent advances in remote sensing LAI, this research illustrates the importance of LAI and its use in crop models for yield assessment. The study used 16 days' temporal resolution of LAI which may not adequately characterize crop productivity when the LAI in critical growth stages is hampered by smoothing algorithms. As the study assessed yield in the rainy season, the availability of LAI was affected by climatic conditions. This does not happen in the post-rainy season as remotely sensed LAI very often suffers due to cloud cover. Further LAI alone may not be providing accurate yield assessment as best results have been reported when both LAI and vegetation indices were assimilated in to crop models (Fang et al., 2011). Soil moisture along with remotely sensed LAI assimilated into the DSSAT-CERES model showed promising results for yield estimation in wheat.

Several input parameters were modified one at a time and in combination, to present an accurate description of the environment where the crop has been growing. This provided an opportunity to predict crop yields with reasonable accuracy, particularly rainfed crops such as maize and groundnut. However, the methodology has a few drawbacks as re-calibration and re-setting of the free variable may result in an unreliable parameter setup and it also requires multiple re-runs that need high computational facilities for larger-scale applications. Further, there are potential uncertainties in the remotely sensed products, needing further investigation on improvements in Landsat 8-derived LAI. Depending on Landsat 8 LAI alone may not always be effective, therefore assimilation of multi-objective variables such as evapotranspiration, soil moisture and fraction of absorbed photosynthetically active radiation (FAPAR) is needed to accurately simulate crop yields using multi-objective optimization algorithms (Ma et al., 2013).

Conclusions

With assimilation of remotely sensed satellite data, we could effectively predict crop yields under both irrigated and rainfed conditions at each CCE location. The estimated yields of rice, maize and groundnut in selected districts of Andhra Pradesh, Odisha and Telangana, respectively, during 2019 correlated well with the observed yields obtained from CCE locations The study illustrated the great potential of the data assimilation approach in estimating crop yields. A more rigorous evaluation will be necessary to investigate the model's efficiency for various climatic and atmospheric conditions, and other parameters.

This study demonstrated of the use of crop extent maps in yield assessment by conducting CCEs, mapping croplands during rainy season using NDVI time series and exploiting the SMT approach along with ground survey data. Major crop extents were mapped with higher accuracy during rainy season for the five study sites with ground data. The study showed the potential of high-

resolution temporal images and ground data for mapping cropland at field scale. Mapping crop types is the first step toward characterizing important crop-growing environments in order to understand crop condition and homogeneity, leading to sustainable use of resources and improvement in farming systems. Future studies include yield estimation using satellite data and comparing crop yield simulated from the standalone model.

References

Aggarwal, P. K., Hebbar, K., Venugopalan, M., Rani, S., Bala, A., Biswal, A., Wani, S., (2008). Quantification of yield gaps in rainfed rice, wheat, cotton and mustard in India. Global Theme on Agroecosystems Report no. 43. Patancheru 502 324, Andhra Pradesh, India: International Crops Research Institute for the Semi-Arid Tropics, p. 36

Asner, G. P., Scurlock, J. M. A., & Hicke, J. (2003). Global synthesis of leaf area index observations: Implications for ecological and remote sensing studies. *Global Ecology and Biogeography, 12,* 191–205.

Batchelor, W. D., Basso, B., & Paz, J. O. (2002). Examples of strategies to analyze spatial and temporal yield variability using crop models. *European Journal of Agronomy, 18,* 141–158.

Belgiu, M., & Csillik, O. (2018). Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. *Remote sensing of environment, 204,* 509–523.

Choudhury, A., & Jones, J. (2014). Crop yield prediction using time series models. *Journal of Economics and Economic Education Research, 15,* 53.

Clevers, J. G., Kooistra, L., & van den Brande, M. M. (2017). Using sentinel-2 data for retrieving LAI and leaf and canopy chlorophyll content of a potato crop. *Remote Sensing, 9*(5), 405. https://doi.org/10.3390/rs9050405

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20,* 36–47.

Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment, 37,* 35–46.

Croitoru, A.-E., Holobaca, I.-H., Lazar, C., Moldovan, F., & Imbroane, A. (2012). Air temperature trend and the impact on winter wheat phenology in Romania. *Climatic Change, 111,* 393–410.

Doraiswamy, P. C., Moulin, S., Cook, P. W., & Stern, A. (2003). Crop yield assessment from remote sensing. *Photogrammetric Engineering & Remote Sensing, 69,* 665–674.

Fang, H., Liang, S., & Hoogenboom, G. (2011). Integration of MODIS LAI and vegetation index products with the CSM–CERES–Maize model for corn yield estimation. *International Journal of Remote Sensing, 32,* 1039–1065.

Gumma, M. K., Thenkabail, P. S., Teluguntla, P. G., Oliphant, A., Xiong, J., Giri, C., Pyla, V., Dixit, S., & Whitbread, A. M. (2020a). Agricultural cropland extent and areas of South Asia derived using Landsat satellite 30-m time-series big-data using random forest machine learning algorithms on the Google Earth Engine cloud. *GIScience & Remote Sensing, 57,* 302–322.

Gumma, M. K., Tummala, K., Dixit, S., Collivignarelli, F., Holecz, F., Kolli, N. R., et al. (2020b). Crop type identification and spatial mapping using Sentinel 2 satellite data with focus on field-level information. *Geocarto International.* https://doi.org/10.1080/10106049.2020.1805029.

Huete, A. R. (1988). A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment, 25,* 295–309.

Hui, J., & Yao, L. (2018). A method to upscale the Leaf Area Index (LAI) using GF-1 data with the assistance of MODIS products in the Poyang Lake watershed. *Journal of the Indian Society of Remote Sensing, 46,* 551–560.

Jensen, J. R. (2004). *Introductory digital image processing: a remote sensing perspective* (3rd ed., p. 544p). Prentice Hall.

Jin, X., Li, Z., Yang, G., Yang, H., Feng, H., Xu, X., Wang, J., Li, X., & Luo, J. (2017). Winter wheat yield estimation based on multi-source medium resolution optical and radar imaging data and the AquaCrop model using the particle swarm optimization algorithm. *ISPRS Journal of Photogrammetry and Remote Sensing, 126,* 24–37.

Jones, J. W., Hoogenboom, G., Porter, C. H., Boote, K. J., Batchelor, W. D., Hunt, L., Wilkens, P. W., Singh, U., Gijsman, A. J., & Ritchie, J. T. (2003). The DSSAT cropping system model. *European Journal of Agronomy, 18,* 235–265.

Kang, Y., & Özdoğan, M. (2019). Field-level crop yield mapping with landsat using a hierarchical data assimilation approach. *Remote Sensing of Environment, 228,* 144–163.

Kross, A., McNairn, H., David, L., Mark, S., & Catherine, C. (2015). Assessment of RapidEye vegetation indices for estimation of leaf area index and biomass in corn and soybean crops. *International Journal of Applied Earth Observation and Geoinformation, 34,* 235–248.

Lobell, D. B., Cassman, K. G., & Field, C. B. (2009). Crop yield gaps: their importance, magnitudes, and causes. *Annual Review of Environment and Resources, 34*(1), 179–204.

Lobell, D. B., Thau, D., Seifert, C., Engle, E., & Little, B. (2015). A scalable satellite-based crop yield mapper. *Remote Sensing of Environment, 164,* 324–333.

Li, H., Chen, Z. X., Jiang, Z. W., Wu, W. B., Ren, J. Q., Liu, B., & Tuya, H. (2017). Comparative analysis of GF-1, HJ-1, and Landsat-8 data for estimating the leaf area index of winter wheat. *Journal of Integrative Agriculture, 16*(2), 266–285.

Ma, G., Huang, J., Wu, W., Fan, J., Zou, J., & Wu, S. (2013). Assimilation of MODIS-LAI into the WOFOST model for forecasting regional winter wheat yield. *Mathematical and Computer Modelling, 58,* 634–643.

Masson-Delmotte, V., Zhai, P., Pörtner, H.-O., Roberts, D., Skea, J., Shukla, P., Pirani, A., Moufouma-Okia, W., Péan, C., Pidcock, R. (2018). Global warming of 1.5 OC: an IPCC special report on the impacts of global warming of 1.5 °C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty. World Meteorological Organization Geneva, Switzerland.

Myneni, R. B., Hoffman, S., Knyazikhin, Y., Privette, J., Glassy, J., Tian, Y., Wang, Y., Song, X., Zhang, Y., & Smith, G. (2002). Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data. *Remote Sensing of Environment, 83,* 214–231.

Nearing, G. S., Crow, W. T., Thorp, K. R., Moran, M. S., Reichle, R. H., & Gupta, H. V. (2012). Assimilating remote sensing observations of leaf area index and soil moisture for wheat yield estimates: An observing system simulation experiment. *Water Resources Research, 48*(5), 1–13.

Olioso, A., Inoue, Y., Ortega-Farias, S., Demarty, J., Wigneron, J.-P., Braud, I., Jacob, F., Lecharpentier, P., Ottle, C., & Calvet, J.-C.

(2005). Future directions for advanced evapotranspiration modeling: Assimilation of remote sensing data into crop simulation models and SVAT models. *Irrigation and Drainage Systems, 19,* 377–412.

Parker, G. G. (2020). Tamm review: Leaf Area Index (LAI) is both a determinant and a consequence of important processes in vegetation canopies. *Forest Ecology and Management, 477,* 118496.

Pitman, J. I. (2000). Absorption of photosynthetically active radiation, radiation use efficiency and spectral reflectance of bracken [*Pteridium aquilinum* (L.) Kuhn] canopies. *Annals of Botany, 85,* 101–111.

Potgieter, A., Power, B., Mclean, J., Davis, P., & Rodriguez, D. (2014). Spatial estimation of wheat yields from Landsat's visible, near infrared and thermal reflectance bands. *International Journal of Remote Sensing Applications, 4,* 134–143.

Qi, J., Chehbouni, A., Huete, A. R., Kerr, Y. H., & Sorooshian, S. (1994). A modified soil adjusted vegetation index. *Remote Sensing of Environment, 48,* 119–126.

Quarmby, N., Milnes, M., Hindle, T., & Silleos, N. (1993). The use of multi-temporal NDVI measurements from AVHRR data for crop yield estimation and prediction. *International Journal of Remote Sensing, 14,* 199–210.

Ren, H., Liu, R., Yan, G., Mu, X., Li, Z.-L., Nerry, F., & Liu, Q. (2014). Angular normalization of land surface temperature and emissivity using multiangular middle and thermal infrared data. *IEEE Transactions on Geoscience and Remote Sensing, 52,* 4913–4931.

Reyes-González, A., Kjaersgaard, J., Trooien, T., Reta-Sánchez, D. G., Sánchez-Duarte, J. I., Preciado-Rangel, P., & Fortis-Hernández, M. (2019). Comparison of Leaf Area Index, surface temperature, and actual evapotranspiration estimated using the METRIC model and in situ measurements. *Sensors, 19,* 1857.

Rondeaux, G., Steven, M., & Baret, F. (1996). Optimization of soil-adjusted vegetation indices. *Remote Sensing of Environment, 55,* 95–107.

Schmidt, G., Jenkerson, C. B., Masek, J., Vermote, E., Gao, F., (2013a). Landsat ecosystem disturbance adaptive processing system (LEDAPS) algorithm description. Open-File Report, Reston, VA, p. 27, doi:https://doi.org/10.3133/ofr20131057

Schut, A., Stephens, D., Stovold, R., Adams, M., & Craig, R. (2009). Improved wheat yield and production forecasting with a moisture stress index, AVHRR and MODIS data. *Crop and Pasture Science, 60,* 60–70.

Stark, S. C., Leitold, V., Wu, J. L., Hunter, M. O., de Castilho, C. V., Costa, F. R., McMahon, S. M., Parker, G. G., Shimabukuro, M. T., & Lefsky, M. A. (2012). Amazon forest carbon dynamics predicted by profiles of canopy leaf area and light environment. *Ecology Letters, 15,* 1406–1414.

Stuart, A. M., Pame, A. R. P., Silva, J. V., Dikitanan, R. C., Rutsaert, P., Malabayabas, A. J. B., Lampayan, R. M., Radanielson, A. M., & Singleton, G. R. (2016). Yield gaps in rice-based farming systems: Insights from local studies and prospects for future analysis. *Field Crops Research, 194,* 43–56.

Thenkabail, P. S., Schull, M., & Turral, H. (2005). Ganges and Indus river basin land use/land cover (LULC) and irrigated area mapping using continuous streams of MODIS data. *Remote Sensing of Environment, 95,* 317–341.

Thenkabail, P. S., Smith, R. B., & De Pauw, E. (2000). Hyperspectral vegetation indices and their relationships with agricultural crop characteristics. *Remote Sensing of Environment, 71,* 158–182.

Thenkabail, P. S., Stucky, N., Griscom, B. W., Ashton, M. S., Diels, J., Van Der Meer, B., & Enclona, E. (2004). Biomass estimations and carbon stock calculations in the oil palm plantations of African derived savannas using IKONOS data. *International Journal of Remote Sensing, 25,* 5447–5472.

Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment, 8,* 127–150.

Wei, C., Huang, J., Mansaray, L. R., Li, Z., Liu, W., & Han, J. (2017). Estimation and mapping of winter oilseed rape LAI from high spatial resolution satellite data based on a hybrid method. *Remote Sensing, 9*(5), 488. https://doi.org/10.3390/rs9050488

Yan, G., Hu, R., Luo, J., Weiss, M., Jiang, H., Mu, X., Xie, D., & Zhang, W. (2019). Review of indirect optical measurements of leaf area index: Recent advances, challenges, and perspectives. *Agricultural and Forest Meteorology, 265,* 390–411.

Yu, Y., Wang, J., Liu, G., & Cheng, F. (2019). Forest Leaf Area Index inversion based on landsat OLI data in the Shangri-La City. *Journal of the Indian Society of Remote Sensing, 47,* 967–976.

Zhao, Y., Potgieter, A. B., Zhang, M., Wu, B., & Hammer, G. L. (2020). Predicting wheat yield at the field scale by combining high-resolution Sentinel-2 satellite imagery and crop modelling. *Remote Sensing, 12,* 1024.