Final Report
June 2020 – February 2021

Pilot studies on GP level Rice Crop yield estimation using Technology (Kharif 2020) using Sentinel 1&2 Satellite data

Submitted to
Mahalanobis National Crop Forecast Center

ICRISAT INTERNATIONAL CROPS RESEARCH
INSTITUTE FOR THE SEMI-ARID TROPICS

Pilot studies on GP level Rice Crop yield estimation using Technology (Kharif 2020) using Sentinel- 1&2 satellite

Submitted to

Mahalanobis National Crop Forecast Center

Executing Agency	:	International Crops Research Institute for the Semi-Arid Tropics (ICRISAT)
		Patancheru 502 324, Telangana,
		India
		Fax : + (91) 40 30713074/30713075
		Phone : + (91) 40 30713466/30713071
		Email : icrisat@cgiar.org
Partner Institution	:	Mahalanobis National Crop Forecast Centre Near Krishi Vistar Sadan Pusa Campus, New Delhi-110012, India Email ncfc@gov.in
Project PI	:	Dr Murali Krishna Gumma
Project Start	:	June 2020
Project Completion	:	February 2021

Executive Summary

The Government of India plans to optimize Crop Cutting Experiments (CCEs) and Gram Panchayat crop yield estimations using different technologies including satellite derived metrics and crop modelling techniques. The present study for Kharif season (2020) aims to Rice crop yield estimations in 25 districts of six states viz. Andhra Pradesh, Karnataka, Odisha, Tamil Nadu, Telangana and Uttar Pradesh. The study will use comprehensive and existing environmental, weather and management data along with satellite derived crop spatial data. This information will be modelled using statistical optimization techniques and DSSAT crop modelling to assess the yield estimations.

The project will be executed by ICRISAT in partnership with Mahalanobis National Crop Forecasting Center (Ministry of Agriculture, India)

Objectives:

1. Rice Crop extent mapping for the study districts

2. Conduct and assess crops cutting experiments using spatial statistical optimising technique for Rice crop of *kharif* season in the study districts.

3. Crop yield estimation based on DSSAT crop simulations.

Target Areas:

The pilot study allocated following twenty-five districts in six states for gram panchayat level rice crop yield estimation.

S.No	State	District
1	Andhra Pradesh	Krishna
2		West Godavari
3		East Godavari
4	Karnataka	Raichur
5		Mandya
6		Davangere
7		Mysore
8		Shimoga
9		Yadgir
10	Uttar Pradesh	Ballia
11		Faizabad
12		Deoria
13	Tamil Nadu	Thanjavur
14	Telangana	Jayashankar Bhupalpally
15	Odisha	Angul
16		Baragarh
17		Balangir
18		Dhenkanal
19		Kalahandi
20		Kendujhar
21		Koraput

22		Mayurbanj
23		Nuapada
24		Sundargarh
25		Puri

ICRISAT: Project implementation, monitoring, coordination and reporting

Ground Data for Crop Classification

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 9 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 (Murali Krishna Gumma et al., 2022).

Accuracy Assessment for Crop Classification

Accuracy Assessment of Crop Type Maps Accuracy assessment of the maps was performed based on validation data. A ground survey samples were used to assess the accuracy of the classification map, by generating a confusion matrix, wherein columns represent field-plot data points, and rows represent results of classified rice maps in the confusion matrix. The confusion matrix contains corresponding class changes in a multidimensional table. The statistical approach of accuracy assessment shows multivariate statistical analyses such as Kappa 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.

Rice Crop mapping - Methodology

The process began with preparing NDVI maximum images for every 15 days of every month from June to December and stacked together and the crop mask was prepared using sentinel-1 VH-min by giving threshold value of greater than -25 for easy extraction of croplands as well as transplant rice fields in Google Earth Engine (GEE) Platform.

The NDVI images was prepared using normalised difference of Near Infrared (NIR) and Red (R) bands of Sentinel 2.
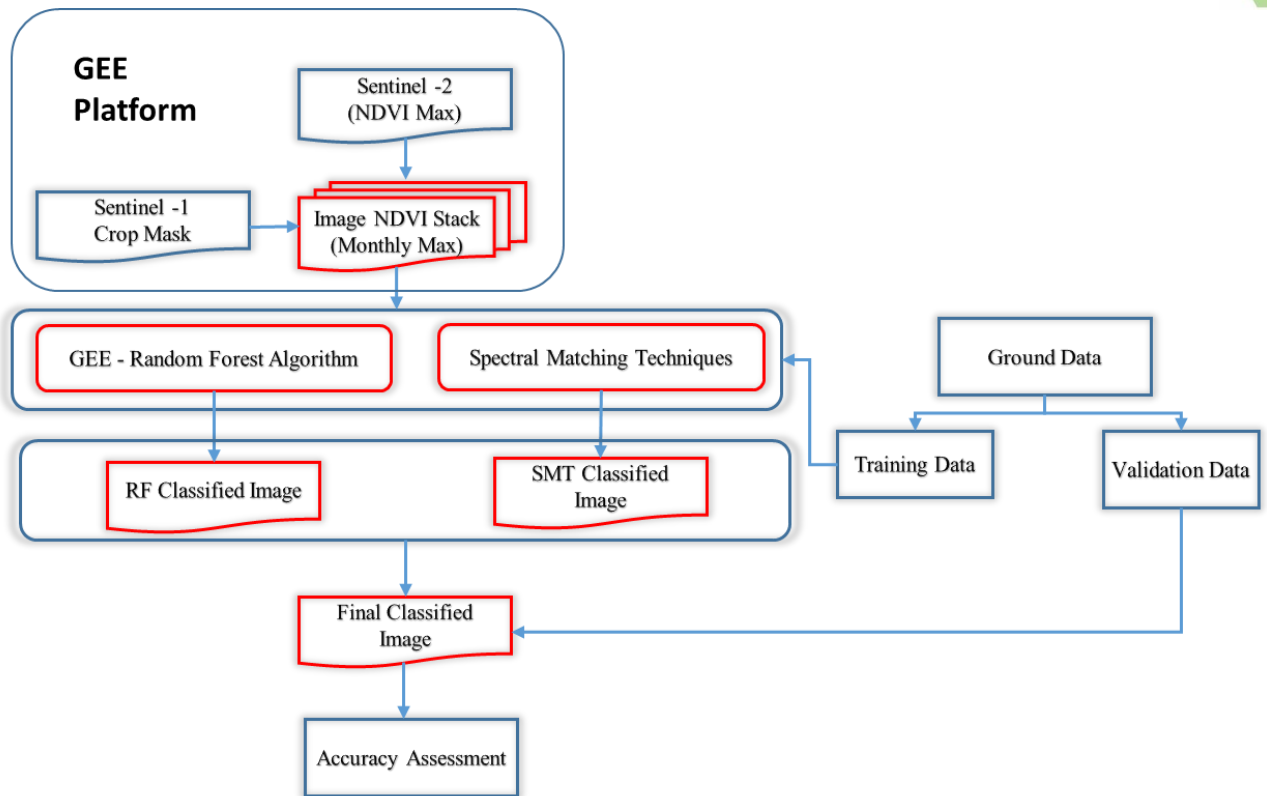
Fig 1: Flow Chart of Methodology of Rice crop mapping

Spectral Matching Techniques:

The stacked image downloaded from GEE consists of every 15 days for entire Kharif season. Unsupervised classification was used to generate initial classes(Murali Krishna Gumma et al., 2020). The unsupervised ISOCLASS cluster algorithm (ISODATA in ERDAS Imagine 2018) run on the stack generated an initial 40 classes, with a maximum of 40 iterations and convergence threshold of 0.99. Though ground survey data was available at the time of image classification, unsupervised classification was used in order to capture the complete effect of all wavelengths over a large area. Use of unsupervised techniques is recommended for large areas that cover a wide and unknown range of vegetation types, and where landscape heterogeneity complicates identification of homogeneous training sites. Identification of training sites is particularly problematic for small, heterogeneous irrigated areas(Murali Krishna Gumma et al., 2022).

Land use/land cover classes were identified based on temporal signatures along with ground survey data. We observed crop growth stages including length of growing periods (LGPs) and cropping pattern from temporal signatures, such as (a) onset of cropping season (e.g., monsoon and winter); (b) duration of cropping season such as monsoon and winter; (c) magnitude of crops during different seasons and years (e.g., water stress and normal years); and (d) end of cropping season.

The process of labelling and class identification was done based on spectral matching techniques (SMTs) (Murali Krishna Gumma, Thenkabail, Deevi, et al., 2018; Murali Krishna Gumma et al., 2016; Murail Krishna Gumma, Uppala, Mohammed, Whitbread, & Mohammed, 2015). Initially, 40 classes from the unsupervised classification were grouped based on spectral similarity or

closeness of class signatures. Each group of classes was matched with ideal spectral signatures and ground survey data, and assigned class names. Classes with similar time series and land cover were merged into a single class, and classes showing significant mixing, e.g., homogeneous irrigated areas and forest, were masked and reclassified using the same ISOCLASS algorithm. This resulted in following classes for each district. We employed a user-intensive method that incorporates both ground survey data and high resolution imagery in order to avoid lumping classes that might be spectrally similar but have distinct land cover(Murali Krishna Gumma, Thenkabail, Teluguntla, & Whitbread, 2018).
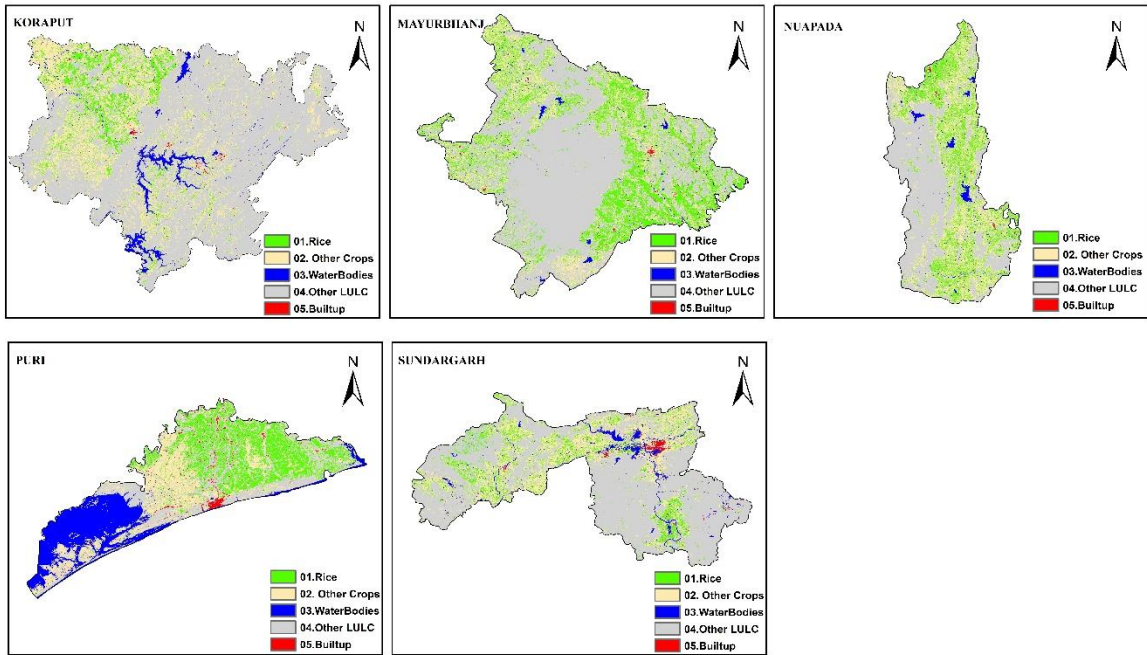
Areas estimated are attached as **Annexure 2**

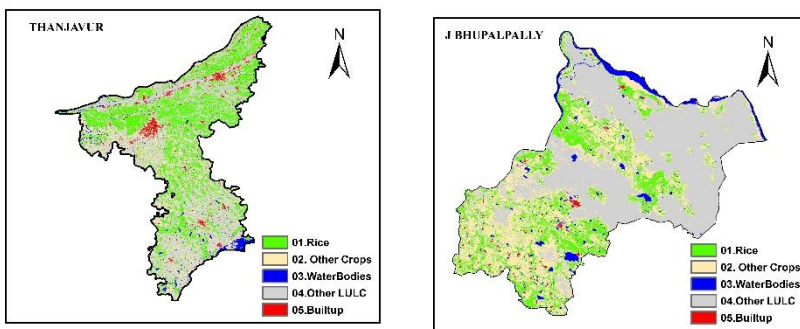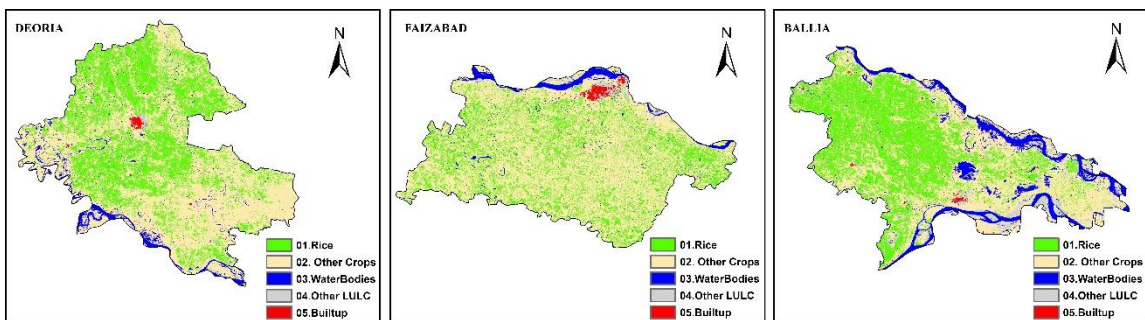Following are the crop type classification images for all study districts (Fig 2):

Andhra Pradesh
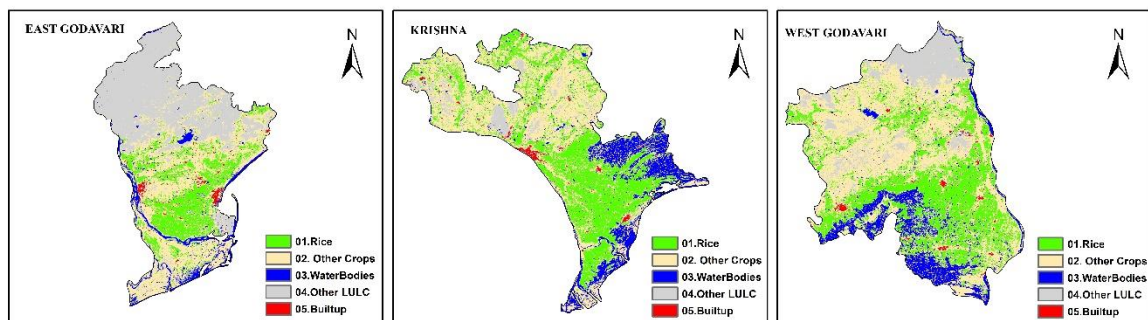
Odisha
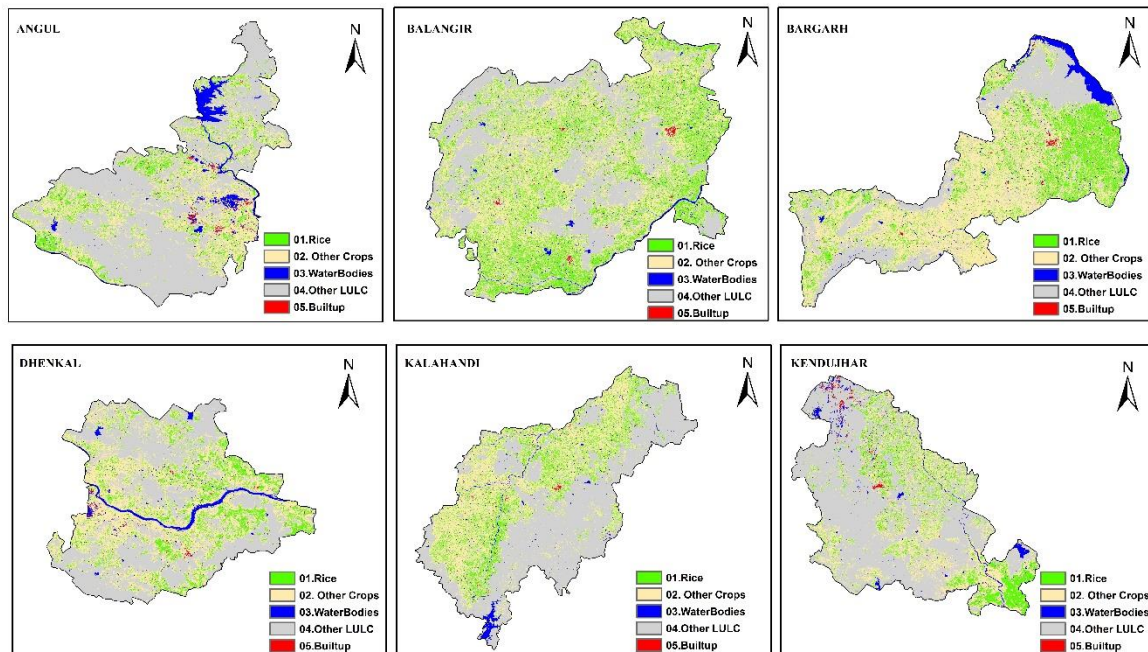
Tamil Nadu & Telangana
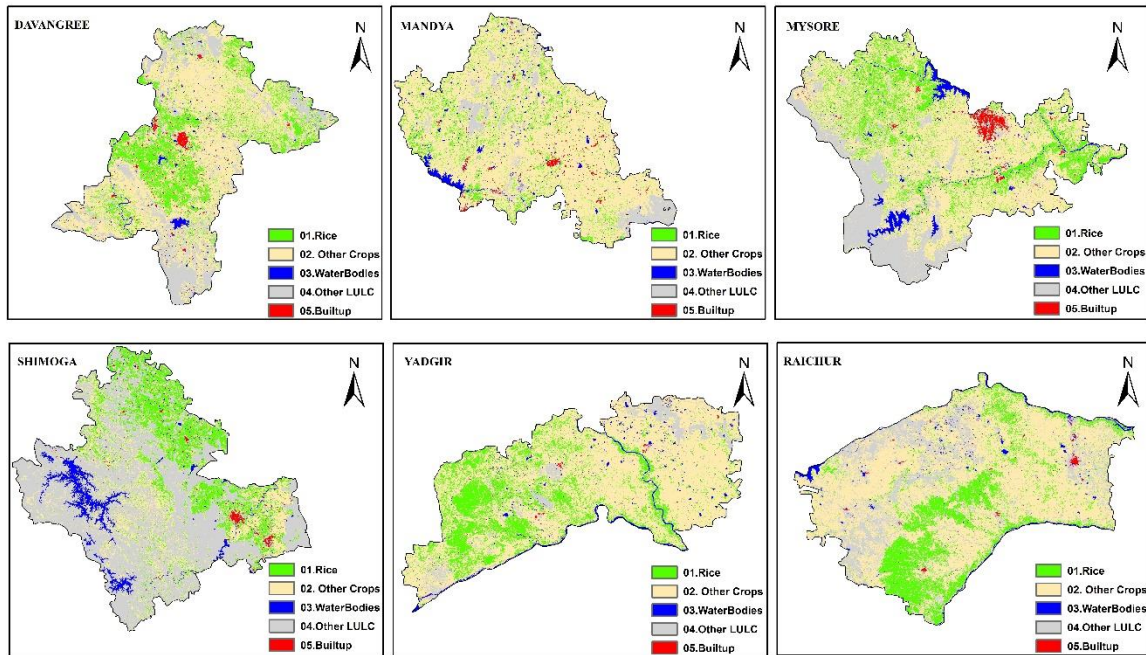

Uttar Pradesh

Karnataka

Fig 2: Crop Classification maps for study districts

CCE's Data Optimisation:

The optimisation of CCE's were carried out using following methodology (Fig.3). The process begin with collection of sentinel 2 NDVI Maximum data (available), climate data and soil map.

The NDVI data with rice crop mask and respective climate and soil data were combined into homogenous stratum and collected random points using stratified sampling. By multiple regression techniques, the number of samples were reduced into half of random samples.
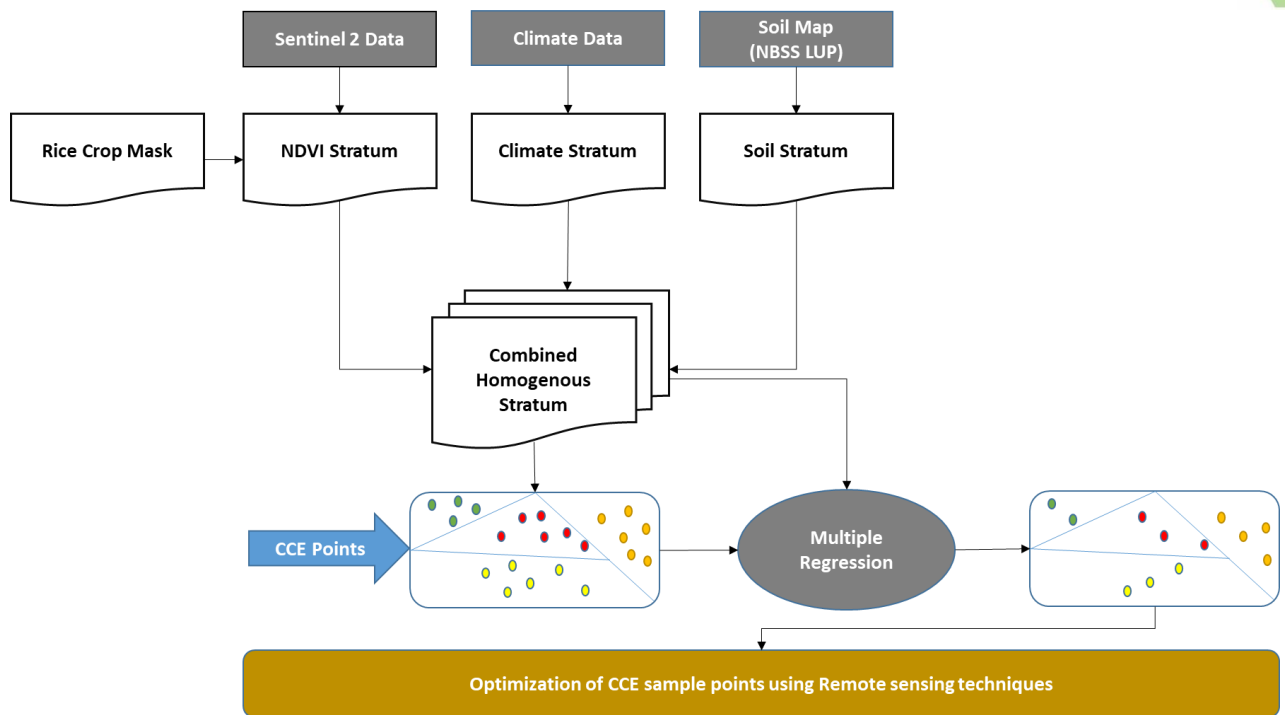
Fig.3 : Optimisation of Crop Cutting Experiments using remote sensing techniques

Using above optimization, we instructed our field staff to collect the possible samples.

In most of the districts, there was same soil type and no significant weather changes available, so, most of the stratification depends upon the NDVI.

Improvements required for stratification:
- High resolution soil data is required for analysis
- There is a need to include agronomical parameters (cultivar types) for smart sampling
- Weather data at possible lower administration level is required;

CCE's Data Collection:

Based on spatial map of Crop extent, optimization and Leaf Area Index (LAI) of rice in their respective areas, the selection of CCE's were shortlisted. LAI indirectly shows the health of the crop, which helps in locating the good crop fields as well as adverse fields for collection of CCE's.

The CCE's was carried out by selecting 5m X 5m plot of field, manually harvested and weighted as shown in following images.
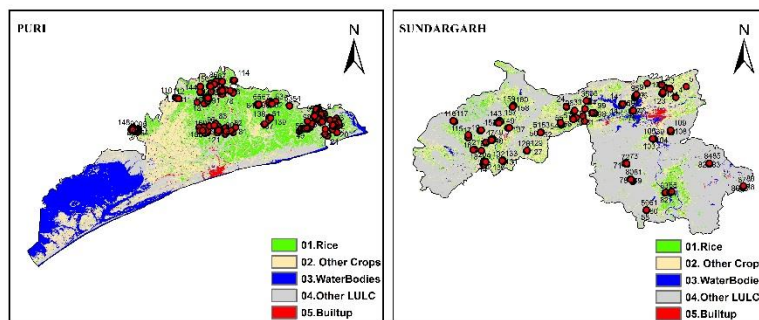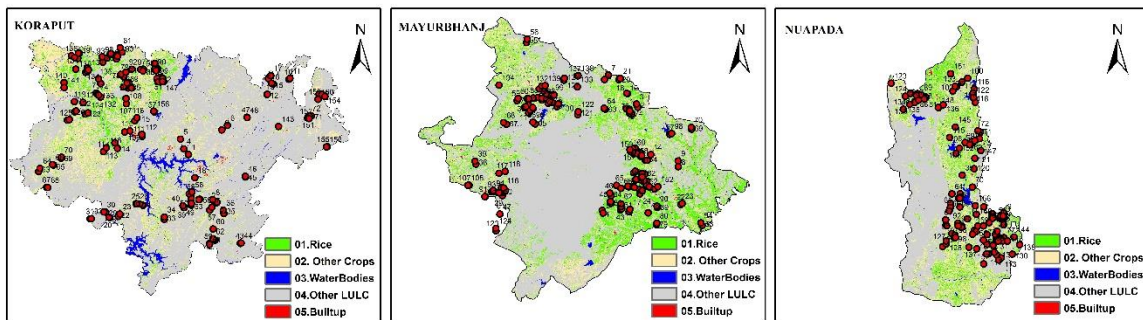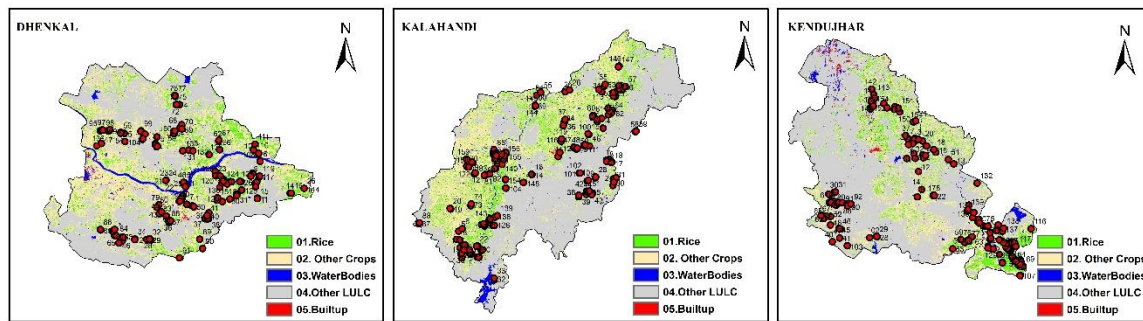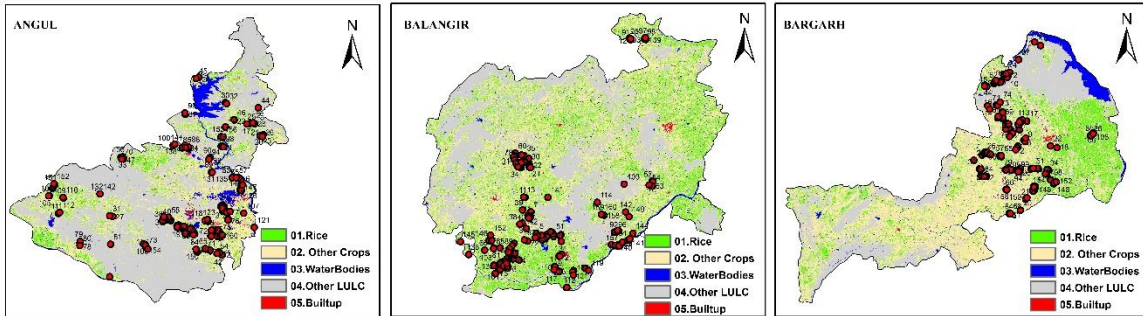
The total number of 160 samples were collected for each district (Annexure 1).

Locations of CCE's collected (Fig 4)

Uttar Pradesh

9
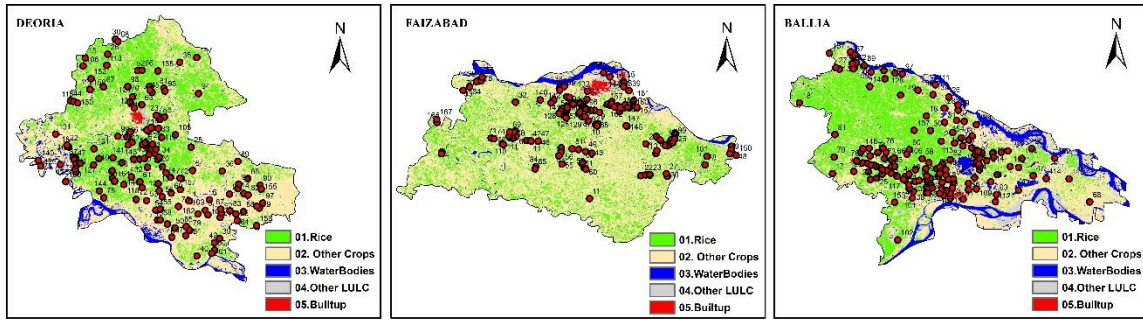
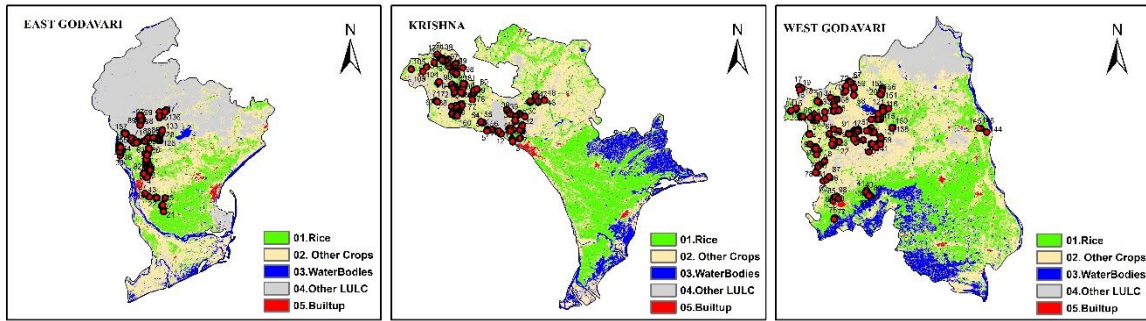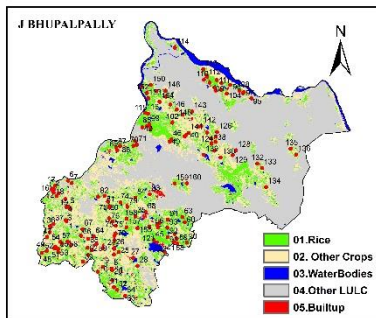**Odisha**

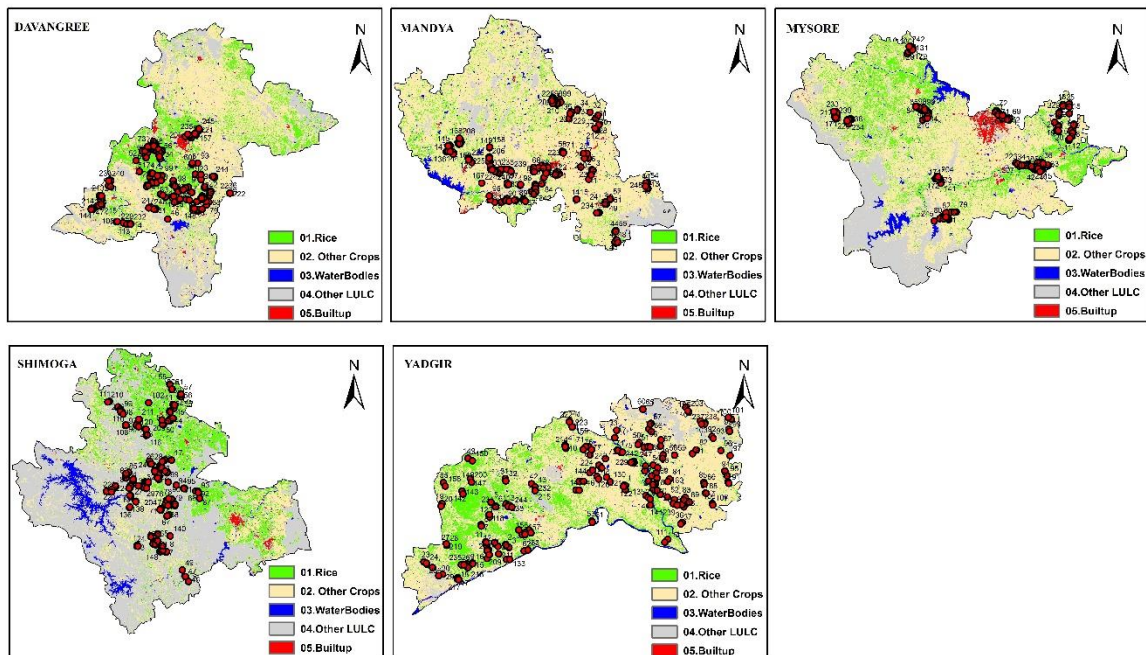**Andhra Pradesh**

Telangana

Karnataka

Fig 4: CCE's distribution across study districts

5. Leaf Area Index

This study used MODIS derived LAI and also sentinel -2 derived LAI index.

- Based on the fact that the spectral response of leaves is unique compared to that of other parts of the plant.
- Vegetation indices – NDVI, EVI, SAVI, etc. – have shown high positive correlation to LAI.
- With a limited field data consisting of LAI values at few locations, regression equations can be arrived at, relating LAI to spectral vegetation indices.
- METRIC (Measuring Evapotranspiration at high Resolution with Internalized Calibration) model has developed a relation between LAI and Sentinel 2-derived Soil Adjusted Vegetation Index (SAVI). According to METRIC model,

$$LAI = \frac{-\ln(\frac{0.69 - SAVI}{0.59})}{0.91}$$

For Landsat-8 images used in this study, SAVI is computed from the formula:

$$SAVI = \frac{(1 + L)\ (B8 - B4)}{L + B8 + B4}$$

Where L is a soil factor, taken to be 0.1, B8 in the spectral reflectance in band 8 (Near Infrared) and B4 is the spectral reflectance in band 4 (Red).

Due to Coarse resolution of MODIS, the study uses LAI derived from Sentinel 2. Compared both values and used the optimised values.

LAI values were extracted for every CCE location and validated against the DSSAT crop model LAI

Integration of remote sensing LAI products with crop simulation models for better crop yield estimation

1. Introduction

Timely and accurate prediction of crop yield is important for agricultural land management and policy making. Several studies have demonstrated the utilization of satellite data in crop yield estimation. However, majority of studies used methods of empirical nature and they work only for specific locations, crops, cultivars and for a particular crop growth stage. Cropping system models and remote sensing tools are two different methodologies often used to answer some of the agronomic questions at various levels such as field and regional scales. Several researchers used these technologies independently however information derived from remote sensing is used to update cropping systems model simulations in recent times as both these technologies are complementary.

Keeping in view the complimentary nature of these technologies several researchers started integration of remote sensing data with crop growth simulation models found to be a promising option for crop growth monitoring and yield estimation. However, each technology has its own advantages limitations. For example use of remote sensing as a temporal crop analysis tool is limited due to availability of cloud free time-series remote sensing data and difficulties in accurate LAI estimation from remotely sensed data.

Similarly cropping systems models are often limited by data availability such as information on cultivar, management, soil, and meteorological inputs for spatial simulations. Uncertainties associated with spatial simulations can be reduced by periodically readjusting the simulation using spatial information from remote sensing images.

Several remote sensing data assimilation methods at various complexity levels were tried mostly either by directly using remote sensing data in the simulation models, updating the state variables or re parameterization of the model using remote sensing data in recent years.

In this study, we used the technique of re- parameterization of crop simulation models based on the several iterations using remote sensing input such as leaf area index(LAI) as it is supposed to be the highest degree of integration. The essence of the data assimilation approach is to improve the initial parameterization of the crop growth model and augment simulation with the use of remotely sensed observations.

2. Methodology

The methodology (**Fig 5**) includes crop model data mainly soil, weather and crop management data and its integration with remote sensing data.

2.1.Data collection

Crop Cutting Experiments (CCE) is an assessment method employed by governments to estimate the crop yield in the region given cultivation cycle. The traditional method of CCE is based on the yield component method where sample locations are selected based on a random sampling of the total area under study. In the current analysis, we identified few mandals in study districts in six states six states viz. Andhra Pradesh, Karnataka, Odisha, Tamil Nadu, Telangana and Uttar Pradesh., to test the methodology. Data assimilation from remote sensing products such as leaf are index (LAI) in to cropping system models to predict crop yield in CCE sites. We have collected GPS location, date of sowing, irrigated vs rainfed and other management details from CCE location if available.

2.2.Soil data

Biophysical crop simulation models normally require profile-wise soil data. For each CCE location, soil inputs to the model were obtained from a set of soil profile data available from ICRISAT data repository and NBSSLUP data bases. We also used certain parameters in soil as free variable. Soil physical and chemical properties such as texture, hydraulic parameters, bulk density, organic matter and available N were extracted for each location based on the available soil profile data. Additional soil parameters such as soil albedo, drainage constant, and runoff curve number were estimated based on soil texture and converted using the generic soil database available in the DSSAT-models.

2.3.Weather data

13

The weather data such as daily maximum temperature, minimum temperature, rainfall and solar radiation data was collected from Automatic Weather Stations (AWS) stations of respective state authorities. If AWS data not available, NASA power data was used for analysis.

2.4. The Cropping System Model

The Cropping System Model (CSM)–Crop Environment Resource Synthesis (CERES)–Rice crop growth model as provided in the Decision Support System for Agro technology Transfer (DSSAT) were used for yield simulations. 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.

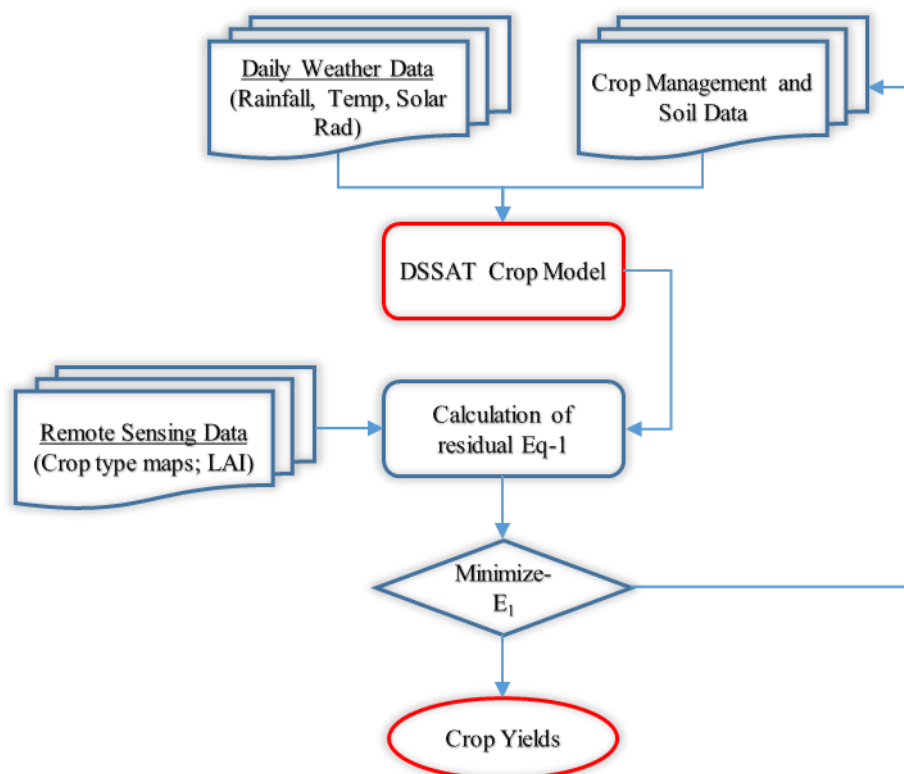**Fig 5.** General methodology of the data assimilation approach integrating remote sensing data with crop growth models for crop yield estimation

2.5. Assimilation of Remote Sensing Data into Crop Growth Model for Yield Estimation

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, Moulin,

14

Cook, & Stern, 2003; Olioso et al., 2005), by updating state variables or by re-parameterization of the model using remote sensing satellite data (Fang, Liang, & Hoogenboom, 2011; Jin et al., 2017)., we used the technique of re-parameterization 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 improve simulations with the help of remotely sensed satellite observations. The optimization process starts with initial model parameterization by adjusting the free parameters so that the model-simulated LAI is in agreement with the Sentinel-2 LAI observations (Eq. 1). 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. The remote sensing LAI data were collected for six times during the crop growth period.

$$= \frac{1}{m} \sum_{i=1}^{m} abs\big[(LAI)_S(t_i) - (LAI)_M(t_i)\big]/(LAI)_M(t_i) \quad \text{--- Equation -1}$$

Where LAIS (ti), LAIM (ti) are the simulated and measured LAI at time t_i, respectively.

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 Sentinel 2 LAI and the crop model-simulated LAI is minimized. 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 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

2.6. Calibration of DSSAT and Validation of yield data at GP level

DSSAT crop models require genetic coefficients, which are cultivar specific for describing processes related to growth and development and grain production. These coefficients allow the model to simulate performance of diverse genotypes under different soil, weather and management conditions. The model was calibrated using field measured values of weather parameters, crop management and soil properties during the cropping season. In our previous studies as a part of Agricultural Model Intercomparison and Improvement Project (AgMIP) phase I & II, we have calibrated CERES-rice model for various cultivars of different duration

As, the model was run at CCE plot level, the observed yield of every CCE collected was validated against the crop model yield generated by re-parameterization of the model free variables using remote sensing LAI data. The rice yields depends mainly on crop management practices followed mainly nitrogen amount and time of application, irrigation application rates,

cultivar duration etc.,   The village mean yield was calculated with collected CCE yield and corresponding simulated yields.

As some times models may underestimate LAI as seen in several published literature and hence we re-parameterized the CERES rice model variables using LAI developed from remote sensing data at regular intervals during crop growth period. However, the accuracy issues for remote sensing LAI may be possible due to due to cloud conditions and varying spectral indices. Further improvements of the Landsat-derived LAI and vegetation index products are necessary, especially during the beginning of the growing season and continued data during the crop growth period. There is also an immediate need to further invest in studying relationship between remote sensing derived LAI product and field LAI observations across locations to understand the accuracy of remote sensing LAI predictions

Future Improvements:
- Improvements in LAI predictions
- Use of remote sensing derived dry matter production and other indices in addition to LAI to re-parameterization of model free variables for improving accuracy of predictions
- Exploring the possibility of establishing a good network of AWS stations for accurate location specific daily weather data for better prediction of crop yields
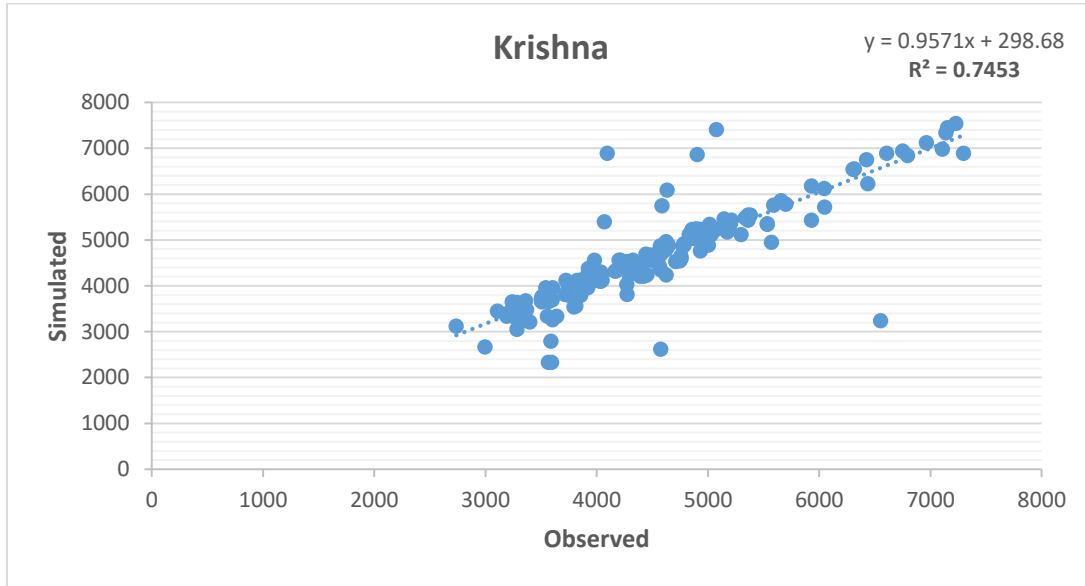
3.0.   Study sites –Results

3.1 East Godavari
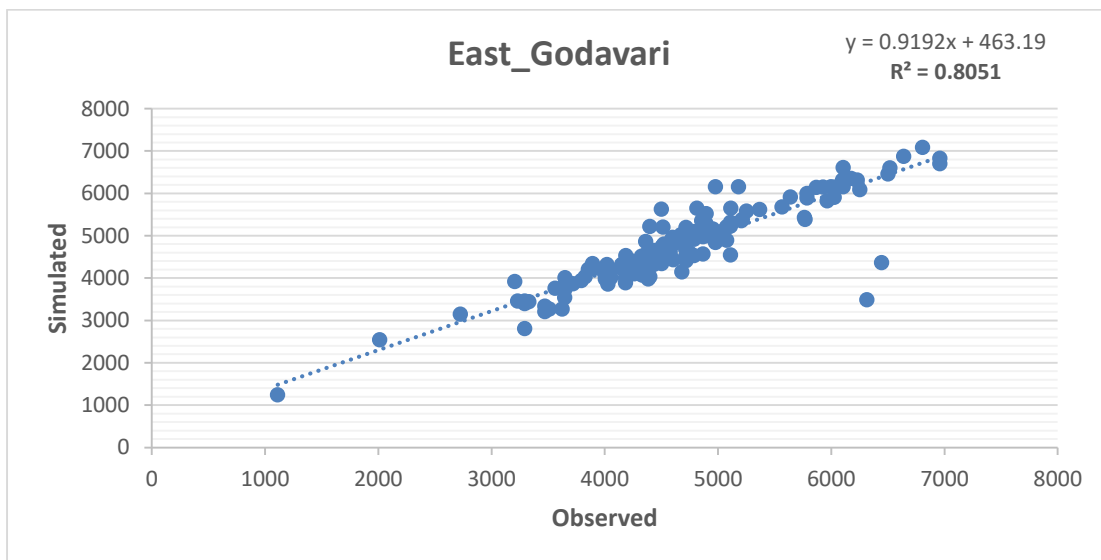
R2	0.805070039	
RMSE	412.4811875	
IoA	0.735168768	
T-test	0.379431032	Non-Significant

3.2 Krishna

Krishna

y = 0.9571x + 298.68
R² = 0.7453

R2	0.745332892	
RMSE	554.2833264	
IoA	0.709312002	
T-test	0.373055539	Non-Significant

3.3 West Godavari

West_Godavari

y = 1.0302x - 143.49
R² = 0.8093
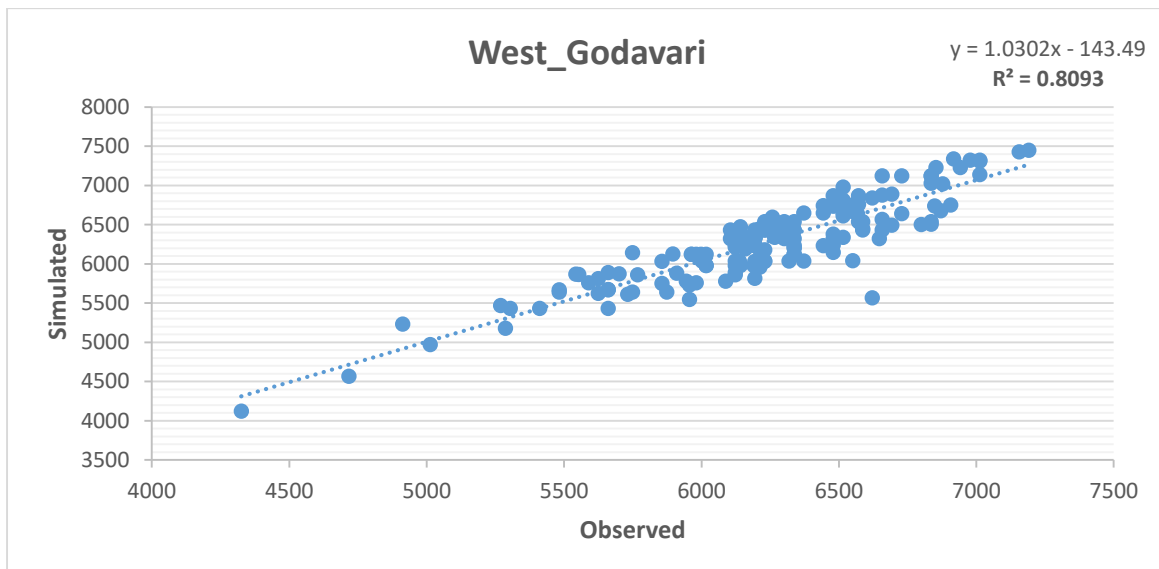
R2	0.809268986	
RMSE	234.8321176	
IoA	0.703002914	
T-test	0.415345658	Non-Significant

3.4 Angul

Angul

y = 1.0468x - 185.08
R² = 0.8096
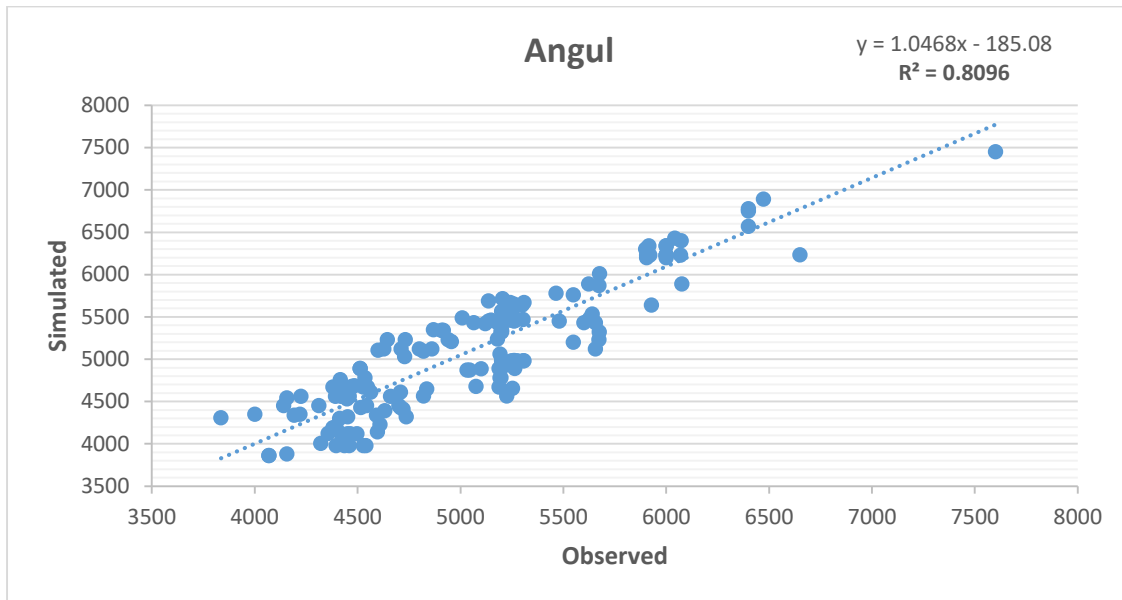
R2	0.809570741	
RMSE	318.5945499	
IoA	0.698830945	
T-test	0.491766652	Non-Significant

3.5 Balangir

Balangir

y = 0.8141x + 405.12
R² = 0.7965
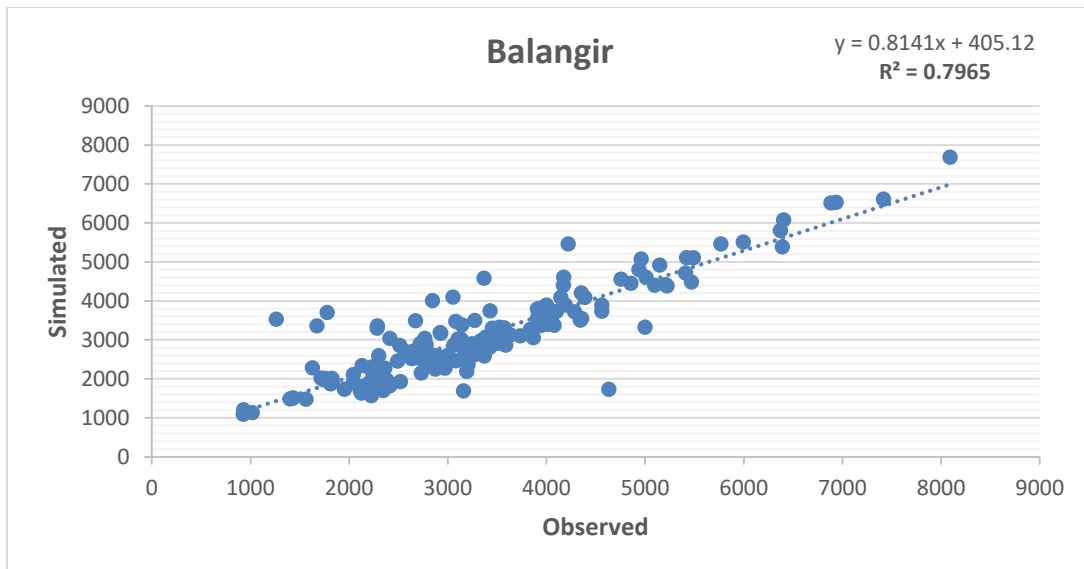
R2	0.796533052	
RMSE	629.9761184	
IoA	0.758597507	
T-test	0.130372327	Non-Significant

3.6 Baragarh

Baragarh

$y = 0.846x + 477.52$
$R^2 = 0.8138$
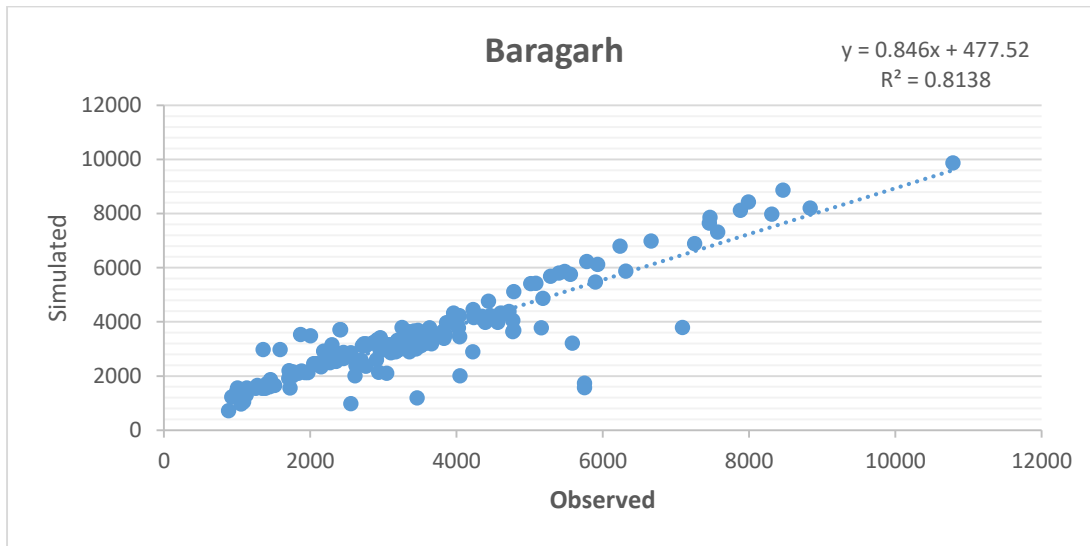
R2	0.813802407	
RMSE	788.8227258	
IoA	0.759050185	
T-test	0.740622404	Non-Significant

3.7 Dhenkanal

Dhenkanal

$y = 1.0508x - 371.12$
$R^2 = 0.83$
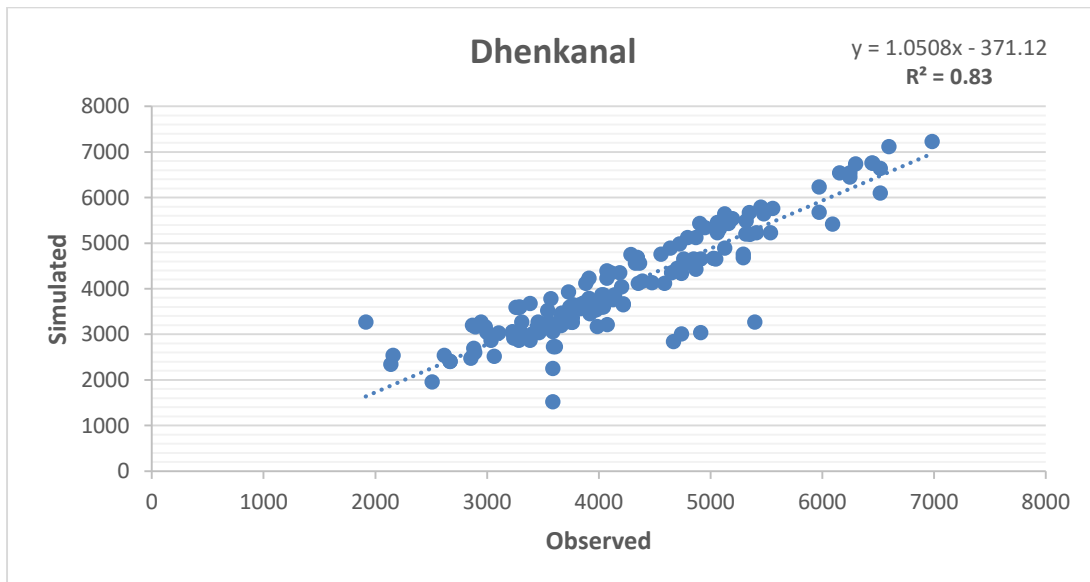
R2	0.829999301	
RMSE	507.5132163	
IoA	0.701486924	
T-test	0.202743099	Non-Significant

3.8 Kalahandi

Kalhandi

y = 0.9527x - 83.987
R² = 0.8049
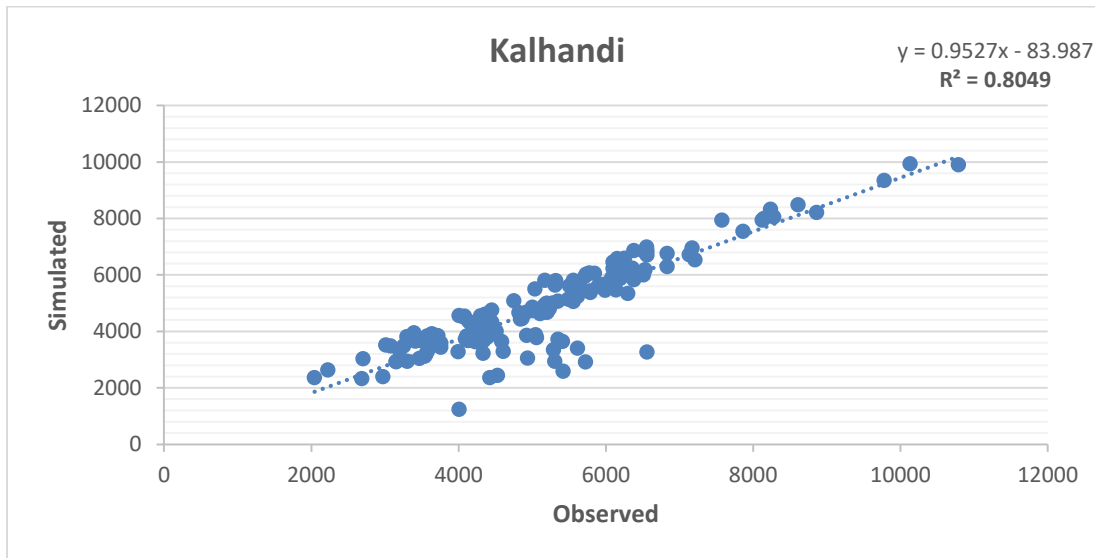
R2	0.807863397	
RMSE	768.1396821	
IoA	0.717659102	
T-test	0.056647698	Non-Significant

3.9 Keonjhar

Keonjhar

y = 0.9072x + 404.42
R² = 0.807
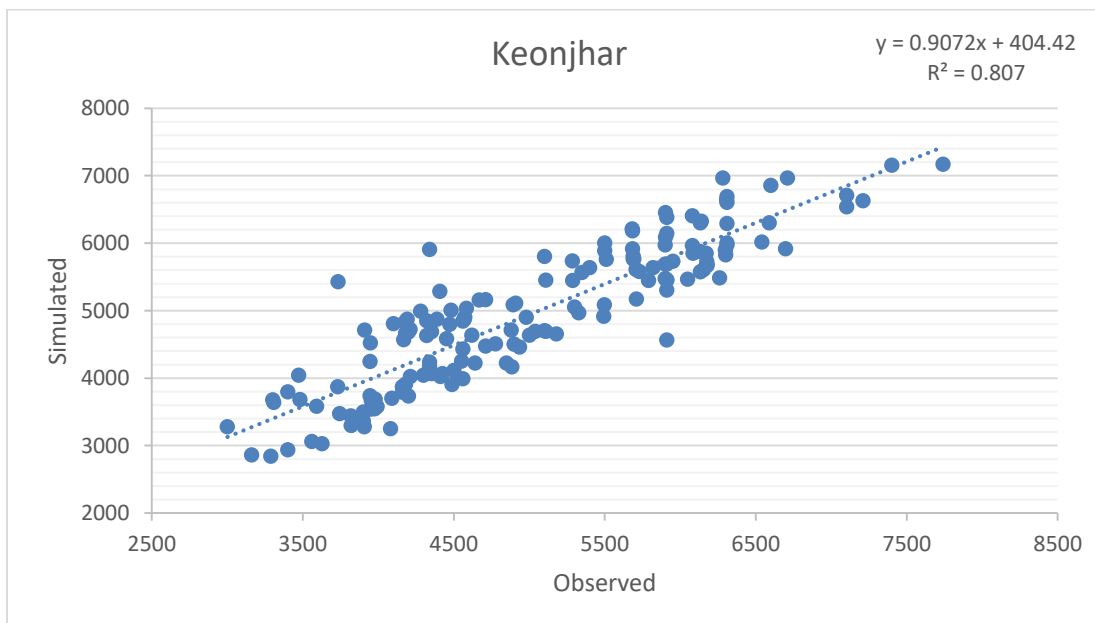
R2	0.807042616	
RMSE	467.5801461	
IoA	0.736499989	
T-test	0.582791212	Non-Significant

3.10 Koraput

Koraput

y = 0.8388x + 699.63
R² = 0.7834
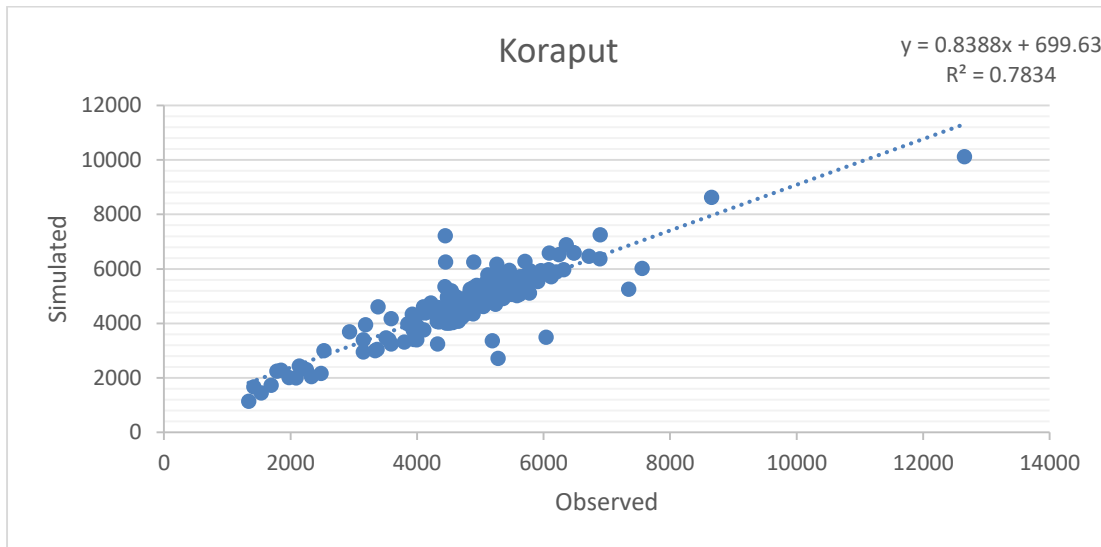
R2	0.783387222	
RMSE	640.3160954	
IoA	0.753353225	
T-test	0.645961675	Non-Significant

3.11 Mayurbhunj

Mayurbunj

y = 0.8875x + 626.26
R² = 0.7585
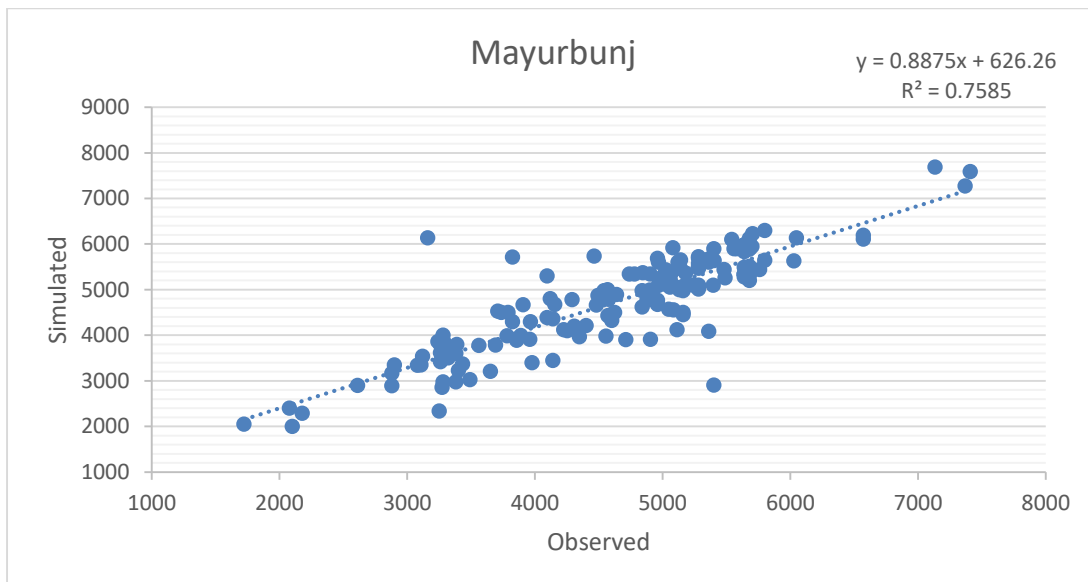
R2	0.758454997	
RMSE	546.6680645	
IoA	0.735784007	
T-test	0.369803373	Non-Significant

3.12 Nuapada
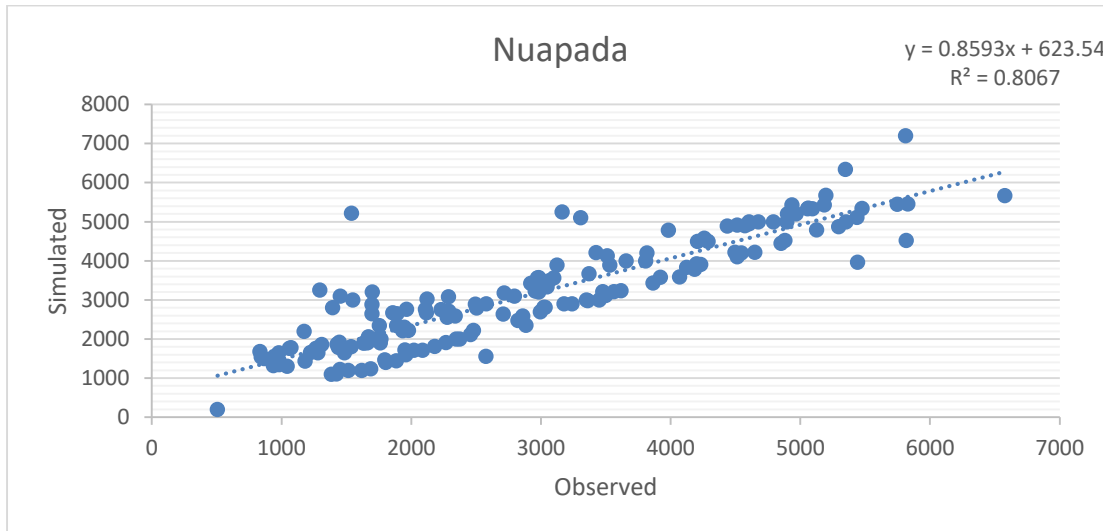
R2	0.806733422	
RMSE	667.3681166	
IoA	0.747602367	
T-test	0.175054281	Non-Significant

3.13 Puri
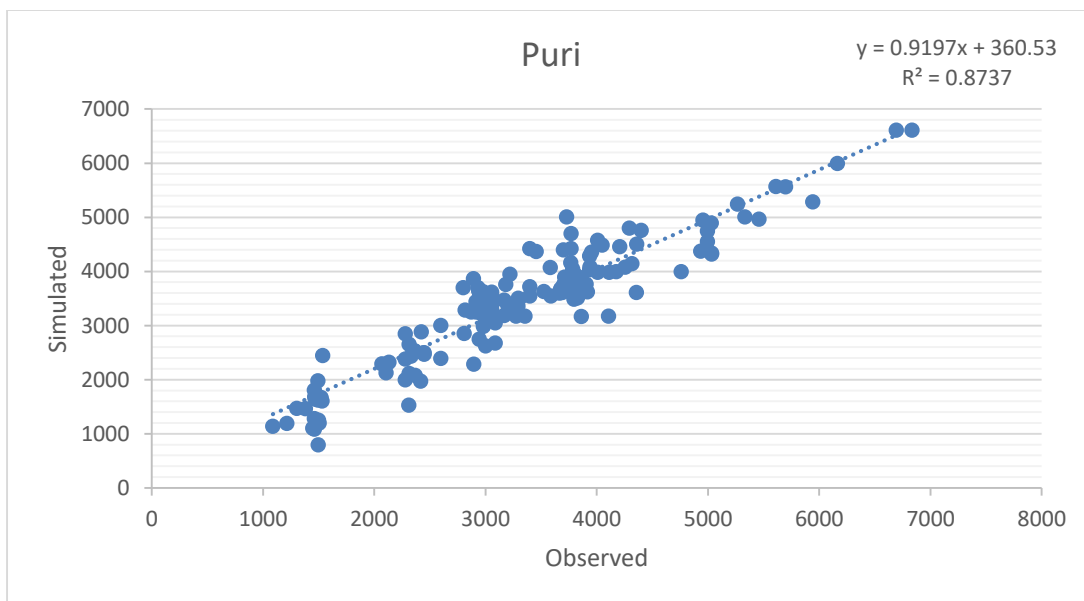
R2	0.873651647	
RMSE	427.6400285	
IoA	0.74687204	
T-test	0.493874411	Non-Significant

3.14 Sundargarh

Sundargarh

y = 0.9606x + 282.43
R² = 0.8
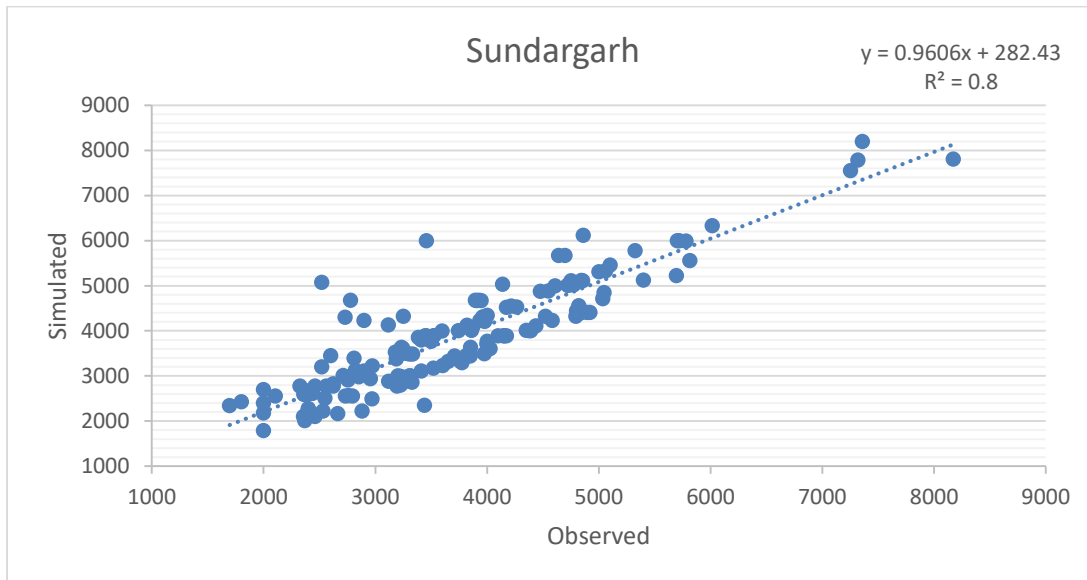
R2	0.799976516	
RMSE	553.7189567	
IoA	0.721454228	
T-test	0.294386185	Non-Significant

3.15 Ballia

Ballia

y = 0.7667x + 784.5
R² = 0.6789
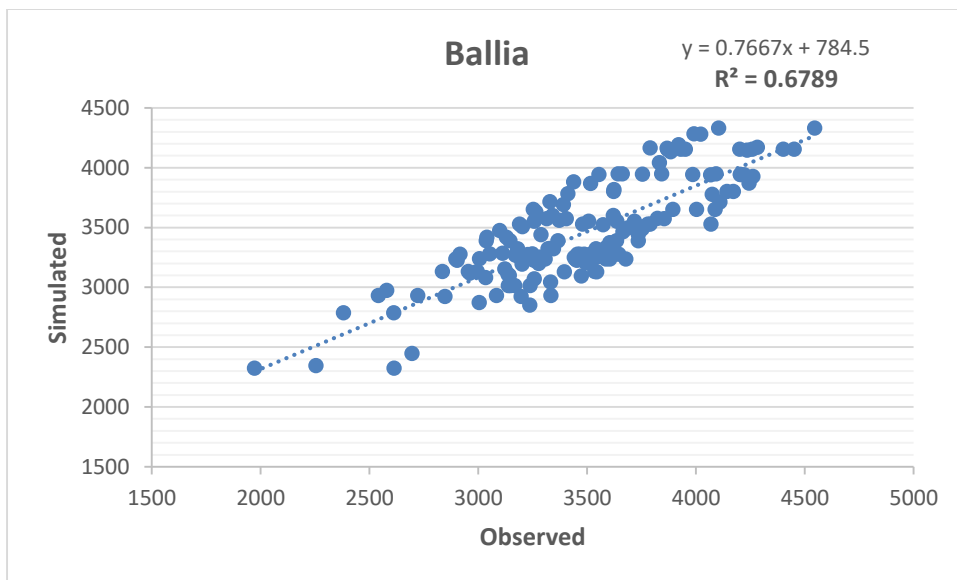
R2	0.678886	
RMSE	258.087674	
IoA	0.75005872	
T-test	0.56214495	Non-significant

3.16 Deoria

Deoria

y = 0.7174x + 780.88
$R^2 = 0.7316$
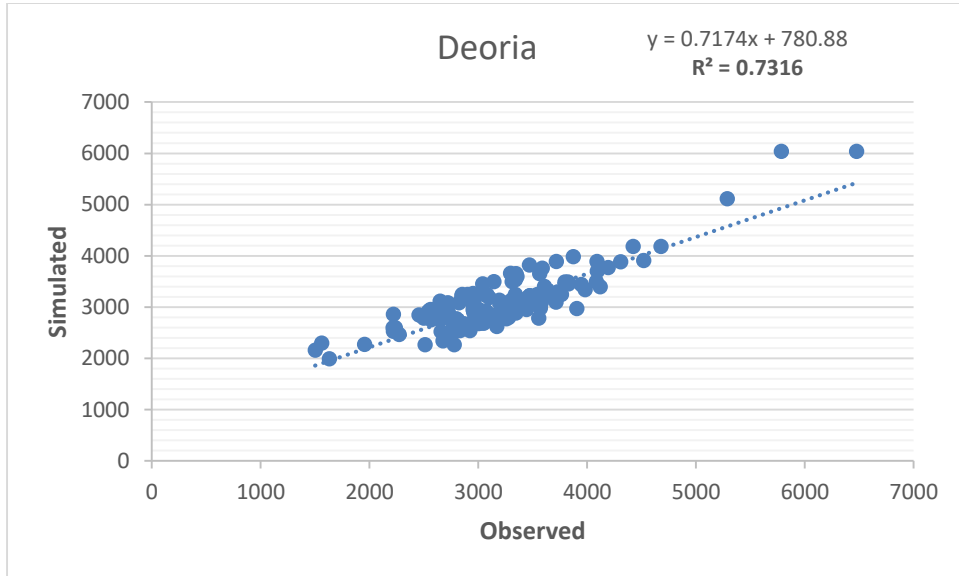
R2	0.73133117	
RMSE	354.233123	
IoA	0.77110869	
T-test	0.05851086	Non-Significant

3.17 Faizabad

Faizabad

y = 0.8637x + 453.62
$R^2 = 0.6727$
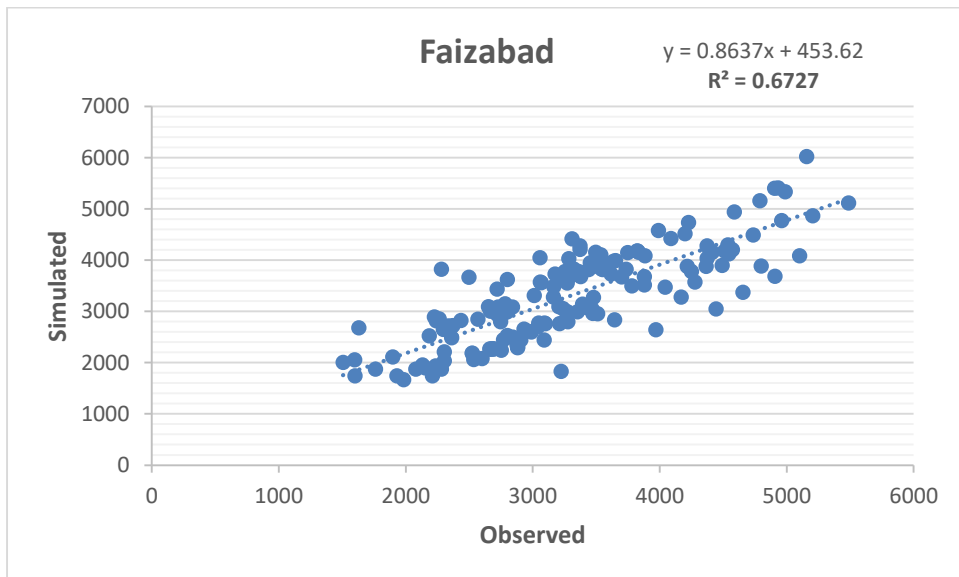
R2	0.67266026	
RMSE	519.864103	
IoA	0.7174809	
T-test	0.95079523	non-Significant

3.18   Davangere
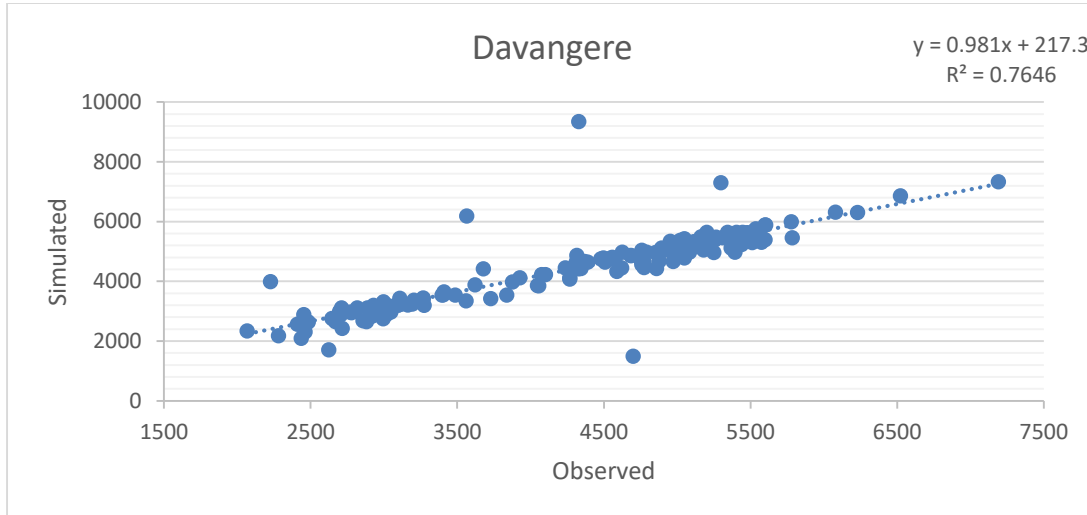
R2	0.764616226	
RMSE	605.4463118	
IoA	0.703079202	
T-test	0.293607104	Non-Significant

3.19 Mandya
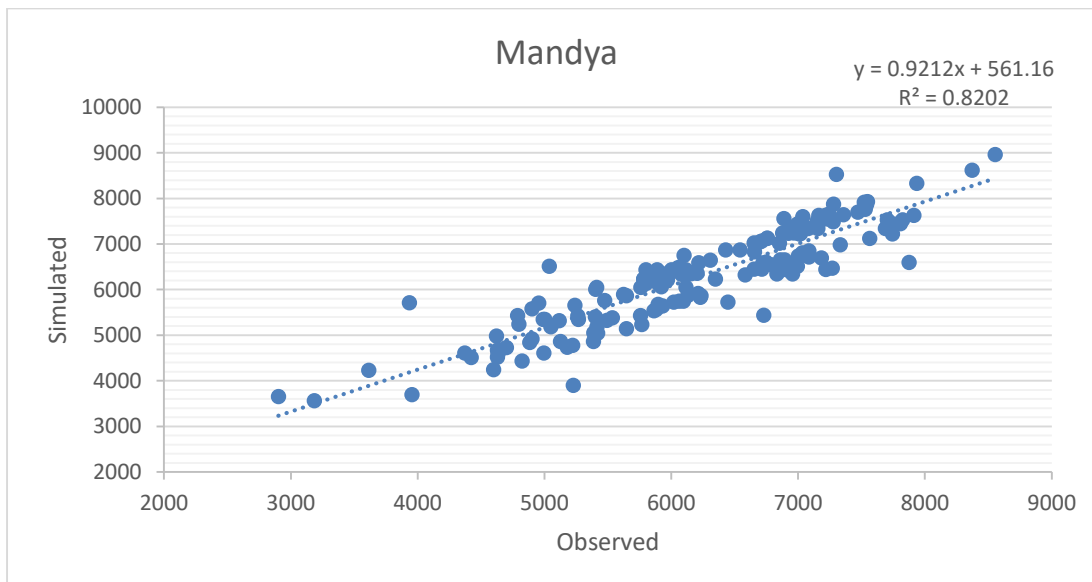
R2	0.820219975	
RMSE	453.4588717	
IoA	0.734492488	

T-test	0.531536949	Non-Significant

3.20 Mysore

Mysore

y = 0.9491x + 425.42
R² = 0.8231
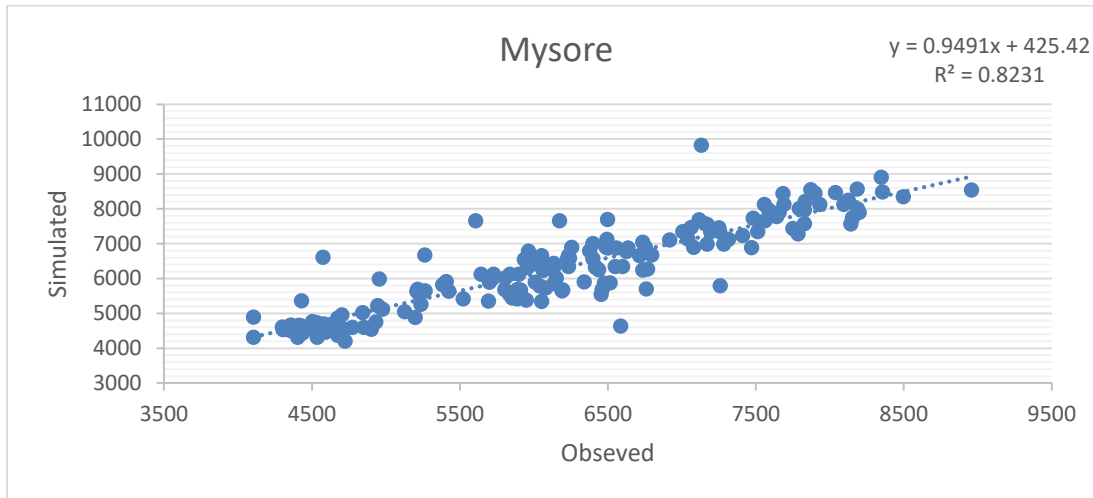
R2	0.823101971	
RMSE	540.5912204	
IoA	0.72907445	
T-test	0.393425463	Non-Significant

3.21 Raichur

Raichur

y = 0.7223x + 1649.7
R² = 0.7244
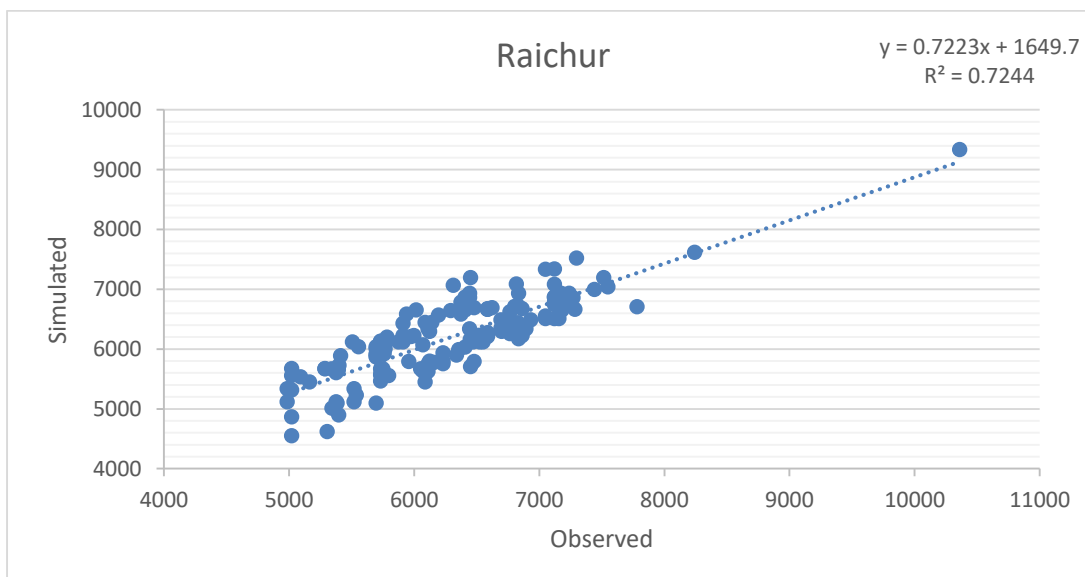
R2	0.724394874	
RMSE	408.170065	
IoA	0.770657505	
T-test	0.166536234	Non-Significant

3.22 Shivmogga
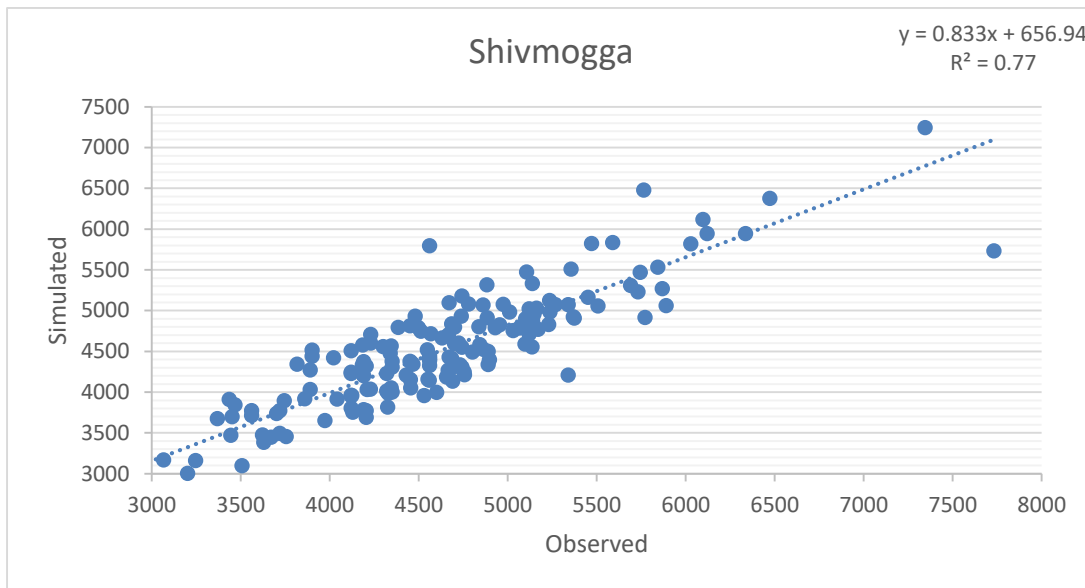
R2	0.767856933	
RMSE	385.6849216	
IoA	0.745696817	
T-test	0.159906211	Non-Significant

3.23 Yadgir
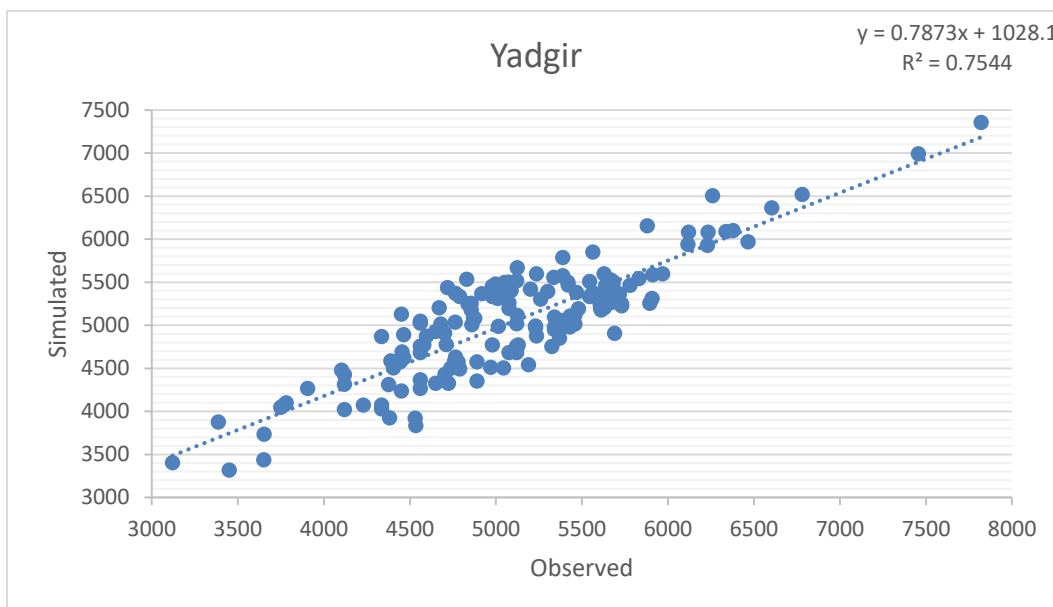
R2	0.762117484	
RMSE	355.0286645	
IoA	0.758430529	

T-test	0.438218204	Non-Significant

3.24 Jayashankar Bhupalpally

Jayashankar Bhupalpally

$y = 0.8831x + 236.52$
$R^2 = 0.8271$
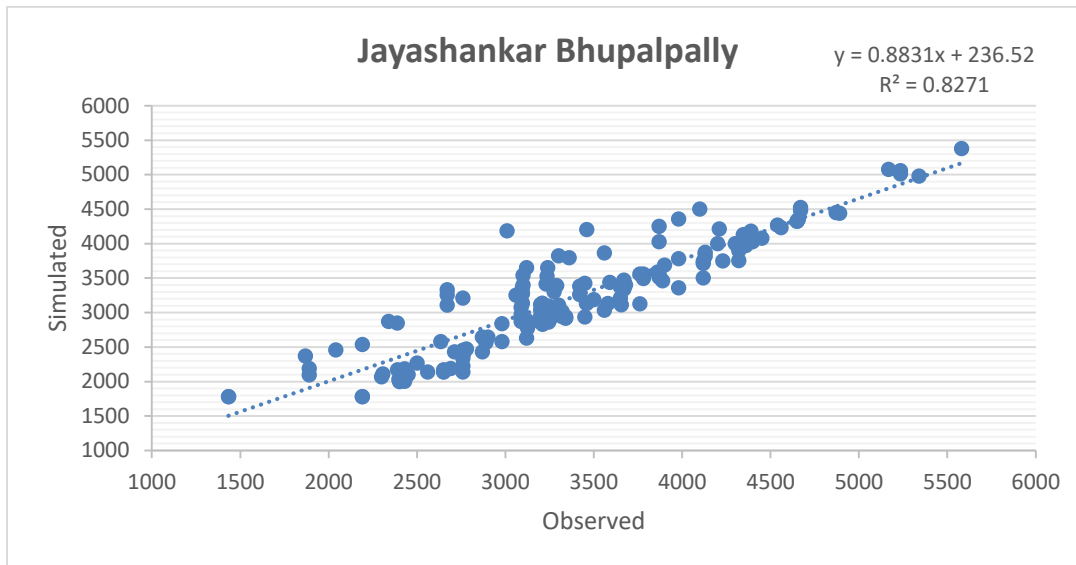
R2	0.827106657	
RMSE	363.6228963	
IoA	0.725159157	
T-test	0.066189445	Non-Significant

4.0. Comparing Sentinel (10m) and Planet scope (3m) Data

This study also compared the classification maps derived from Sentinel (10m) and Planet scope (3m) satellite imagery (Fig 6). This helps in finding the noise in crop classification maps and also helps in attaining good accuracy in classification of other land use land cover.

The comparison was done by picking some area of district and observed the changes in classification.
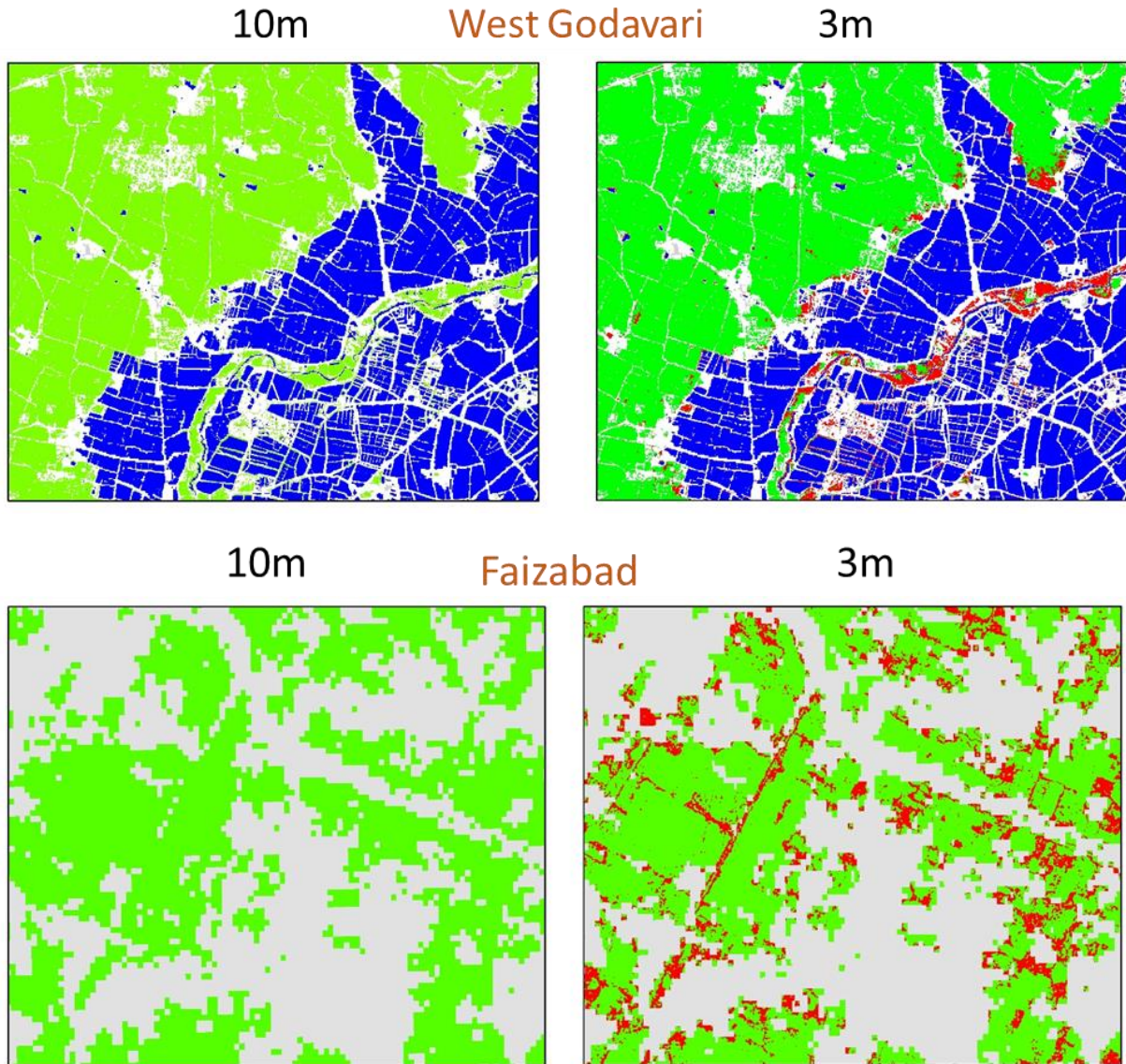
Fig 6: Comparing Sentinel and Planet scope derived classification maps

- For rice dominant areas (West Godavari), 10m has less noise in classification and where trees vegetation is dominant (Faizabad), there is possibility of high noise
- High resolution 3m data is highly useful for identification of rice class, where rice crop is not dominant
- The red colour indicates the other classes which is mixed with rice class in 10m resolution data

5.0.   Spatial Distribution of Yield

Attached as Annexure 1

6.0.   Challenges and improvements

This study indicates the importance of LAI in the data assimilation process and that the incorporation of LAI can improve crop yield prediction. However, the following points need to be considered to further improve the yield prediction.

1. Collection of cloud-free time-series remote sensing data during the cropping season (at least fortnightly if not weekly) for assimilation of data in crop models for improving modeling efficiency.
2. There is a need to study the relationship between remote sensing derived LAI product and final yields of various crops especially in rain-fed regions.
3. Further improvements of the Sentinel 2 -derived LAI and vegetation index products are necessary, especially during the beginning of the growing season and continuous data during the crop growth period.
4. The availability of location-specific weather data is the key for proper simulations with crop simulation models. In some states there exists a good network of AWS stations, however the majority of other locations this is major lacunae.

Annexure 1:

S.NO	STATE	DISTRICT	GT_Training	GT_Validation	GT_Bhuvan	Classification	Classification_Accuracy	CCE	CCE_Bhuvan	LAI	SOIL	WEATHER	YIELD			
													R-Squared	RMSE	Index of Agreement	T-test
1	Andhra Pradesh	Krishna	170	171		✓	85.38	161	159	✓	✓	✓	0.745	554.3	0.709	0.373
2		West Godavari	126	145	456	✓	84.14	161	132	✓	✓	✓	0.809	234.8	0.703	0.415
3		East Godavari	160	138		✓	86.96	163	168	✓	✓	✓	0.8	412.5	0.735	0.379
4	Karnataka	Raichur	47	162		✓	84.32	161	160	✓	✓	✓	0.7243	408.2	0.7706	0.167
5		Mandya	69	171		✓	87.72	171	160	✓	✓	✓	0.8202	453.5	0.7344	0.532
6		Davangere	161	169	59	✓	85.21	161	60	✓	✓	✓	0.7646	605.4	0.703	0.293
7		Mysore	69	161		✓	86.96	174	149	✓	✓	✓	0.823	540.6	0.729	0.393
8		Shimoga	116	146		✓	83.56	166	100	✓	✓	✓	0.767	385.7	0.745	0.16
9		Yadgir	183	175		✓	85.71	161		✓	✓	✓	0.762	355	0.758	0.438
10	Uttar Pradesh	Ballia	209	144	179	✓	88.89	161		✓	✓	✓	0.678	258.1	0.75	0.56
11		Faizabad	203	144	128	✓	86.81	168		✓	✓	✓	0.672	519.9	0.717	0.95
12		Deoria	195	180		✓	86.67	165		✓	✓	✓	0.731	354.2	0.77	0.05
13	Tamil Nadu	Thanjavur				✓					✓	✓				
14	Telangana	Jayashakar Bhupalapally	162	157		✓	85.35	160	80	✓	✓	✓	0.8271	363.6	0.725	0.066
15	Odisha	Anugul	214	141		✓	82.98	161	59	✓	✓	✓	0.809	318.6	0.698	0.491
16		Bargarh	257	170		✓	84.12	161	48	✓	✓	✓	0.813	788.8	0.759	0.74
17		Bolangir	257	168		✓	85.71	161	61	✓	✓	✓	0.796	630	0.758	0.13
18		Dhenkanal	148	173		✓	83.82	161	53	✓	✓	✓	0.8299	507.5	0.701	0.202
19		Kalahandi	177	155		✓	85.16	161	58	✓	✓	✓	0.807	768.1	0.717	0.057
20		Kendhujar	187	174	61	✓	87.93	161	56	✓	✓	✓	0.807	467.6	0.736	0.582
21		Koraput	196	152		✓	84.12	161	72	✓	✓	✓	0.7833	640.3	0.7533	0.646
22		Mayurbanj	256	159		✓	92.45	161	122	✓	✓	✓	0.758	546.7	0.735	0.369
23		Nuapada	223	159		✓	84.28	161	101	✓	✓	✓	0.8067	667.4	0.747	0.17
24		Sundargarh	194	151		✓	86.09	161	53	✓	✓	✓	0.799	553.7	0.721	0.294
25		Puri	44	140		✓	92.14	161	10	✓	✓	✓	0.873	427.6	0.746	0.493
			4023	**3805**	**883**			**3904**	**1861**							

#IFPRI Component – Picture Based Analysis

GP level crop yield estimation by integration of biophysical model with near surface remote sensing derived parameters

1. Introduction

IFPRI's role in the project was to analyze whether the use of smartphone images for targeted crops improves the accuracy of crop model-based predictions of GP-level yields. Past efforts of this approach have focused on studies on wheat in Haryana, where we have used interpolation of images throughout the season—and, as a more scalable approach, applied convolutional neural networks on individual images—to estimate growth stages (Hufkens et al., 2019) and identify visible crop damage in smartphone images of wheat (Ceballos et al., 2020). Since project funding was insufficient to cover image acquisition throughout the season, we focused on images taken at the time of the CCEs, and hence do not analyze results around estimated growth stages (see Afshar *et al.*, 2020 for more information on how visible information on growth stages can improve yield predictions).

Instead, we focus on the use of images to identify visible crop damage and explain gaps between actual and modeled yields. Specifically, we do not estimate yield directly from the smartphone images, but rather assume that whereas potential yields can be predicted largely by biophysical crop models, for which we utilize ICRISAT's crop model outputs. These models will not account for any mechanical damage that may have reduced yields below potential levels. Our objective, then, is to analyze to what extent visible damage identified in the smartphone images can be predicted using machine learning (specifically, using convolutional neural networks or CNNs); and if so, to what extent the gap in predicted or potential versus actual or measured GP-level yields is correlated with the extent to which we observe visible damage in a given GP.

2. Methodology

2.1 Damage Detection

The smartphone images of crops were all captured during the maturity stage, since photos were acquired at the time of the CCEs. These photos were utilized to detect damage and estimate the extent of damage in two ways: by labeling of images based on visual inspections by agronomic experts; and by using the resulting image labels to train convolutional neural networks for classifying images in terms of whether crops were visibly damaged by hazards such as hailstorms, lodging or pests and disease. These types of mechanical damage, often localized, are not captured in crop growth models that predict potential yields but would cause losses relative to potential yields. As final step, we therefore estimated a 'corrected' statistical crop model that accounts for non-simulated mechanical damage visible in the smartphone images of crops, which we would expect to improve accuracy of yield estimation.

Analysis of Images: Image pre-processing, filtering, and labeling

A total of 5,422 images were available. There were multiple types of images. For each of the fields studied, ICRISAT had provided images with a south-west view, a panoramic view, a granular view, and plot marking images. Only panoramic and granular view images are considered for analysis. Images were also rotated, and crops were not always in a near vertical orientation. To ensure that crops were vertically aligned, we rotated images by 90 degrees. To determine whether an image needed to be rotated clockwise or counterclockwise, we developed an algorithm to detect orientation of the images and correct the image orientation. Moreover, all images had a portion of the images rendered unusable because image meta-data were printed over the images. We recommend that applications used when conducting CCEs allow for these image meta-data to be removed from images themselves, and are stored in auxiliary datasets or in the image filenames instead, since the stamps printed over the images reduce value for future image processing.

We pre-processed these images to improve model performance. As a first step, the portion of the image which contained image meta data was cropped away. Given the relatively smaller number of images in the dataset, as a second step, we cropped images so that only crop information was preserved, and extraneous information such as the sky were removed. Additional algorithms were developed to achieve these tasks quickly. Several enhancements were made to the modelling to improve accuracy. First, contrast normalization was done to eliminate the effects of images having been taken at different times of the day. Given that the number of images were small, image augmentation was done on the fly to simulate a larger dataset. These augmentations allowed for mirror images of the input images and small changes in rotation angles, color and contrast characteristics of the images. Finally, we removed blurry and non-significant images.

Out of 5,422 images, we remained with 2,451 images, out of which we selected only images that were in the ripening and maturity stage, since those are the stages during which the CCEs can be conducted and during which principal damage occurs. This leaves us with 2,419 images, for which we assigned labels to the images indicating growth stages, the incidence (and cause of) visible damage, and if damaged, the extent of the damage. In the present analyses, we do not use information on growth stages, because there was little variation in growth stages given that all images were taken at the time of the CCEs; instead, we focus on the incidence, cause and extent of any visible crop damage.

The incidence and causes of visible damage are summarized in Table 1, which provides the number of instances obtained—and associated percentages—for each class of damage. Each image was assigned only a single damage label. Out of 2,419 images, about one fifth was classified not to have any visible damage. More than one third of all images was diagnosed to have visible damage from a disease (or a pest in a minority of cases), and more than one quarter of all images were classified to have damage from weeds. Another major class was formed by images in which crops were visibly damaged by heavy winds or lodging (15 percent). Floods, low plant population, nutritional deficiency and drought were detected in only a small minority of images.

Table 1: Types of damage obtained from image labeling

Damage Type	Nr images	Percentage
Good (no damage)	491	20.30
Disease	877	36.25
Weed	642	26.54
Wind	364	15.05
Pest	21	0.87
Flood	8	0.33
Low plant population	7	0.29
Nutritional deficiency	6	0.25
Drought	3	0.12
Total	**2419**	**100.0**

Agronomists also labeled all 2,419 images in terms of the extent of damage, estimated in percentages (see Table 2). As in Table 1, in about one fifth of all images, we did not find any evidence of visible crop damage (there are 16 images with disease, wind and weed damage whereby experts assessed that the damage was minimal, and as a result, 507 images are classified to have zero percent damage in Table 2, even though in Table 1, there are 491 images labeled to not have any damage). Most images were classified to have some damage, either 10% (more than a quarter of all images), or 20% (more than a fifth of all images). More severe damage was visible in the remaining 30 percent of all images. The image labels presented in Tables 1 and 2 were used to train machine learning models, as discussed in the next section.

Table 2: Extent of damage obtained from image labeling

Extent of Damage	Nr Images	Percentage
0%	507	21.0
10%	648	26.8
20%	521	21.6
30%	302	12.5
40%	136	5.63
50%	148	6.13
60%	42	1.74
70%	30	1.24
80%	52	2.15
90%	33	1.37
Total	**2419**	**100.0**

Damage Category Classification

In training our machine learning models, we combined the categories for drought, flood, low plant population and nutritional deficiency into an 'other' category, which we dropped from the analyses, since these categories included only a small number of observations, and constituted only 1.04% of the total number of images. Alternative model architectures were

also examined, for instance with these other images included, but the final predicted classes including "Good (no damage)", "Disease", "Wind" and "Pest" (with "Other" omitted from the analyses).

Table 3 provides the confusion matrix for the final image classification result, showing that the model assigns the majority of images to the same category as the actual label. For instance, we classify 91.7% of all images without visible damage into the 'good health' category; 95.2% of images with disease damage are accurately classified into the 'diseased' category; 93.8% of images with pest damage are classified into the 'pest' category (all but one of 16 images with visible pest damage); 92.2% of images with weed damage are accurately classified as such; and the model identifies wind damage in 90.1% of all cases where experts indeed identified wind damage. This is encouraging in terms of the scalability of an image-based approach; provided that farmers are able to send in these images, one can use automated image processing techniques to identify the cause of damage.

**Table 3: Confusion Matrix with final damage classes**

Actual/Predicted	Good Health	Diseased	Pest	Wind	Weed	Total
Good Health	354	19	7	5	1	386
Diseased	0	657	5	25	3	690
Pest	0	1	15	0	0	16
Wind	1	16	4	265	8	294
Weed	2	28	2	11	510	553
Not Classified	0	0	0	0	0	0
Total	357	721	33	306	522	1939

Extent of Damage Classification

A second model, using the architecture of the classification model that was developed to predict the incidence and cause of damage, was developed to estimate the extent of damage. The model used root mean square error as the loss function. Threefold cross validation was used again to assess the consistency in the results. Several architectures were considered. The root mean square error obtained in the threefold cross-validation are 6.13%, 7.80% and 7.17%. The correlation coefficients between the actual and predicted and the actual values are 0.801, 0.679, and 0.672. The figure below provides a scatter plot of the observed extent of damage (in percentages, as determined by experts) against the predicted extent of damage by the

convolutional neural network, providing visual evidence of the strong correlation between the extent of damage assessed by experts and the damage predicted by the model.
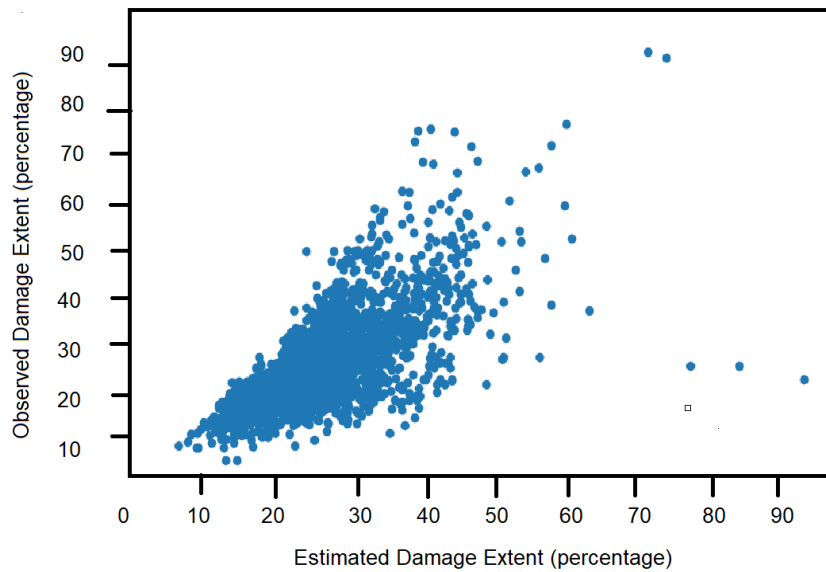
2.2 Building statistical models based on type and extent of damage

Biophysical crop models such as ICRISAT's [DSSAT] are commonly used to predict yield responses to weather, management practices, soil types and cultivar types. However, a challenge to applying this approach for estimating yields in real-world systems is that key input variables (e.g. plot-level management practices) and deviations from potential yields due to mechanical—often localized—damage are not captured by these models, and they are not easily monitored or observable without time consuming and costly field surveys. Crop models predict potential yields and do not estimate mechanical damage, for example damage caused by hailstorms, lodging due to high winds, or pests and disease.

In this study, we address this challenge by combining the predictive capacity of crop models with measurable information derived at low cost from scalable smartphone imagery. The idea is to detect the occurrence of mechanical crop damage on individual plots, so that area-yield predictions based on our statistical model can be adjusted to account for these additional crop losses. To identify whether there is visible mechanical damage, and to assess the extent of the damage, we use smartphone images.

The estimates of crop damage are combined with potential yields predicted by ICRISAT's crop model in the following regression equation:

$$Y_{iv} = \alpha + \widehat{Y_v^m}\beta_1 + \mathbf{D}_{iv}\beta_2 + \varepsilon_{iv},$$

whereby Y_{iv}^m is the yield predicted by the biophysical crop model (estimated by ICRISAT) for site i in village v, $\mathbf{D}_{iv}$ is a vector with (1) the actual image labels (type of damage dummies

and intensity of damage + interactions with type of damage) or (2) predicted labels from the damage classification model, and ε_{iv} is an error term, which we assume is clustered at the village level. We also control for district fixed effects (i.e., we include a separate intercept α for each district).

3. Result and Discussion

3.1 Descriptive analysis of data

Table 4 describes the final image data used in the analyses (for which we have both predicted yields from the ICRISAT model and image data, and excluding the images with damage labels that were classified into the 'other' category), with the percentage of images classified in each of the different health categories at the site and GP level. Data are available for 1,159 GPs, with on average 1.69 images per GP. This is well below the targeted number of 4 CCEs per GP, meaning that the images provide merely a snapshot at the GP level, which will likely have implications for the extent to which GP-level yield measurements correlate with the extent of damage assessed from smartphone images for that GP.

Table 4: Percentage of images classified in health categories and probability of classification in health categories.

Health Category	Labelled	Predicted	Probability of classification
Site Level			
Diseased	35.2	36.8	0.33
Good Health	19.7	18.2	0.24
Pest	0.8	1.7	0.03
Weed	28.2	26.6	0.20
Wind	15.0	15.6	0.20
Total	**1,961**	**1,939**	**1,961**
GP-Level			
Diseased	38.29	39.47	0.34
Good Health	19.38	17.93	0.24
Pest	1.12	1.97	0.03
Weed	27.61	26.08	0.20
Wind	12.51	13.46	0.19
Total	**1,159**	**1,159**	**1,159**

The figure below shows a histogram of CCE and simulated yield values. We find that CCE and simulated yield values are close to one another, indicating that farmers may not have had losses due to mechanical damage at a large scale. This will potentially limit the room for the images

37

to explain additional variation in the CCE yields that is not captured by the simulated yield values estimated using the ICRISAT crop model.

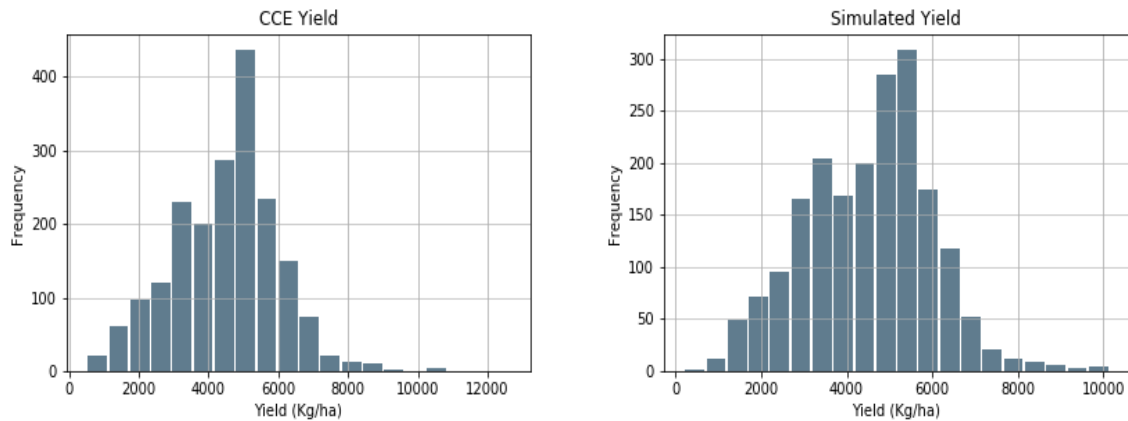**Histogram of Simulated Yield distribution and CCE yield distribution.**

Table 5 provides other data used in the analyses, including simulated yields from the ICRISAT model (a measure of potential yields), actual yields measured through the CCEs, and the percentage of crop damage based on the image labels (labelled crop damage) versus the predictions from the convolutional neural networks (predicted crop damage), along with an estimate of the quantity lost based on the actual versus predicted image labels. Column (1) provides these statistics for all sites, including sites both with and without crop damage visible in the smartphone image. Column (2) includes only sites with damage. Columns (3)-(6) disaggregate this further into sites with visible damage due to disease, pest, weed and wind, respectively.

We do not find that sites with damage have significantly lower yields based on the CCEs compared to sites without damage. In fact, sites in which agronomists identified diseases, yields are above average. Yields are lowest for sites with damage due to weed, which we classify as mismanagement. Simulated yields follow this trend, raising the question whether the modeled yields capture the types of damage that a program such as the PMFBY would want to insure. Overall, there is no major gap between simulated or potential yields and CCE yields, indicating that there is little room for mechanical damage observed from smartphone images to explain a discrepancy between potential and actual yields, with one exception: in the case of pest damage, we find that CCE yields are lower than average, but simulated yields are above average; and the estimated quantity lost is above average for these cases based on both the labelled and predicted estimates of crop damage. Here, images could potentially explain a discrepancy between potential and actual yields, although the number of images with visible pest damage is small, limiting the extent to which we can make inference based on the data.

Table 5: Descriptive statistics

| | | All sites (1) | Sites with damage | | | | |
| | | | Disease | | | | |
			All (2)	(3)	Pest (4)	Weed (5)	Wind (6)
CCE yield	Mean	4503	4497	4708	4540	4286	4383
	Std. dev.	1504	1535	1510	1511	1534	1524
Simulated yield	Mean	4509	4502	4716	4950	4315	4311
	Std. dev.	1480	1509	1504	1491	1472	1511
Labelled crop damage:							
Percentage damage	Mean	21	26	24	29	21	40
	Std. dev.	20	19	19	19	13	24
Quantity lost[*]	Mean	949	1179	1187	1550	904	1621
	Std. dev.	992	977	1083	1168	652	1040
Predicted crop damage:							
Percentage damage	Mean	23	26	26	21	22	38
	Std. dev.	21	22	22	23	17	25
Quantity lost[*]	Mean	1017	1182	1260	1137	937	1526
	Std. dev.	1049	1089	1231	1304	795	1101
Number of observations		**1961**	**1575**	**690**	**16**	**553**	**294**

[*] Calculated as the percentage of damage times the simulated potential yields.

3.2 Performance of statistical yield models for predicting CCE-level yields

As a next step, we analyze to what extent crop damage identified in smartphone images are correlated with deviations between modeled and actual CCE yields at the level of an individual site, before aggregating data at the GP level. To that end, we regress measured yields (from the CCEs) on the yields predicted by the ICRISAT model, and on quantity lost due to crop damage, controlling for district fixed effects (to focus on variation in yields within districts). This is the basic model estimated in Column (1) of Table 6. In Column (2) we also control for the type of damage that is visible in the image of a given CCE site, and in Column (3) we include interaction terms for the type of damage and the percentage of damage that is visible in the crop pictures.

A major challenge in these analyses is that potential yields, estimated with the ICRISAT model, are so closely correlated with measured yields that there is little room for images to explain any further variation. Even when controlling for district fixed effects in Column (1), potential yields explain 79.6 percent of the variation within districts. A model that indeed performs that well in predicting actual yields would reduce the need for any additional technology, such as

smartphone images to estimate localized mechanical damage, but with an R-squared of 0.796 after controlling for variation between districts, we believe that the model used to estimate potential yields is likely overfitted, and we do not have access to the data on whether a sites has been included as training or validation data point, which could help rule this out.

Table 6: Regressions for CCE-level yield predictions

	Dependent variable: Yield		
	(1)	(2)	(3)
Modeled yields	0.892***	0.894***	0.894***
	(0.0222)	(0.0228)	(0.0229)
Quantity lost	0.000167	0.000111	0.00309
	(0.000179)	(0.000174)	(0.00198)
Health = 1, Diseased		4.992	17.9
		(31.64)	(32.59)
X % damage			-0.00308
			(0.00198)
Health = 2, Pest		-371.3***	-257.1**
		(111.9)	(93.82)
X % damage			-0.00373
			(0.0024)
Health = 3, Weed		-39.4	-47.87
		(36.26)	(43.95)
X % damage			-0.00283
			(0.00194)
Health = 4, Wind		49.84	43.92
		(39.6)	(40.98)
X % damage			-0.00292
			(0.00208)
Constant	466.1***	469.2***	463.0***
	(96.82)	(104.1)	(105.2)
Number of Districts	19	19	19
Mean Yield - CCE	4532	4532	4532
Observations	1,939	1939	1,939
R-squared	0.796	0.798	0.798

Note: Robust standard errors in parentheses clustered by GP. All estimates are controlling for district fixed effects, and the R-squared is the proportion of variation explained within districts.

*** p<0.01, ** p<0.05, * p<0.1

3.3 Performance of statistical yield models for predicting GP-level yields

Table 6 estimates the regression equation specified above with all variables aggregated at the GP instead of CCE level. That is, we regress GP-level yields measured through CCEs on the potential yields in the GP predicted by the ICRISAT model ('Modeled yields'), and quantity

lost due to crop damage for a given GP, as estimated from the smartphone images available for that GP. Column (2) adds the proportion of sites for which images have been classified to have different types of damage, and in Column (3), we add interaction terms for this variable and the average quantity lost as inferred from the smartphone images available for that GP (both variables are demeaned so that the coefficients for the main effects can be interpreted as the marginal effect at the average quantity lost and the average proportion of sites with damage from different types of causes). We again control for district effects to focus on explaining within-district variation.

In the first column, we find a close correspondence between actual yields measured from CCEs and modeled yields: an increase of 100kg in modeled yields is associated with an increase of 93.6 kg in measured yields. The average quantity lost within that GP as estimated from the smartphone images is not significantly correlated with yields measured at the GP level. Losing an estimated 100kg is associated with a negligible increase in average yields of 0.02kg. These findings are robust to controlling for the proportion of images labeled with different types of damage within a GP in Columns (2) and (3). As in Table 6, the only type of damage that is associated with a reduction in measured yields is pest damage; so, pest damage predicts lower yields not only at the CCE level, but also at the GP level. In Column (3), we find increased average yields in GPs in which there has been moderate wind damage, but not in GPs in which there has been more severe damage and a higher quantity was lost.

Table 7: Regressions for GP-level yield predictions

	Dependent variable: Average yield in GP		
	(1)	(2)	(3)
Modelled yields	0.936***	0.938***	0.939***
	(0.0262)	(0.0259)	(0.0278)
Average quantity lost	0.000197	0.000136	0.0000497
	(0.00026)	(0.000263)	(0.000421)
Proportion of sites with damage from Disease		28.48	11.05
		(35.59)	(36.14)
X Average quantity lost			74.44
			(78.86)
Proportion of sites with damage from Pest		-407.6***	-371.2***
		(135.4)	(50.12)
X Average quantity lost			-23.45
			(228.4)
Proportion of sites with damage from Weed		6.422	9.757
		(43.65)	(49.62)
X Average quantity lost			4.671
			(61.84)
Proportion of sites with damage from Wind		59.23	158.8**
		(43.91)	(68.94)
X Average quantity lost			-131.7
			(123.9)
Constant	257.7**	240.3*	239.6*

	(119.7)	(115.3)	(123.7)
Number of Districts	19	19	19
Mean Yield - CCE	4532	4532	4532
Observations	1,159	1,159	1,159
R-squared	0.81	0.812	0.812

Note: Robust standard errors in parentheses clustered by GP. All estimates are controlling for district fixed effects, and the R-squared is the proportion of variation explained within districts.

*** p<0.01, ** p<0.05, * p<0.1

Conclusions and recommendations

Summarizing, we find that convolutional neural networks can accurately identify from smartphone images taken at harvest time (i) whether a crop is damaged, (ii) the cause of the damage, including whether damage is due to natural hazards such as a disease or wind damage, versus management practices resulting in excess weeds; and (iii) the extent of the damage. However, controlling for yields predicted through the ICRISAT crop models, we do not find a strong association between these indicators and either CCE-level or GP-level yields. At least in this case, the type of damage and extent of damage are not major determinants of GP-level yields, and we would not recommend including these images for GP-level yield assessment. Nonetheless, the strong performance of the machine learning models to process smartphone images indicates that this method could still be useful in implementing other components of the PMFBY coverage, for instance localized damage, verifying mid-season adversities or post-harvest losses.

There are a few challenges that we encountered in these analyses, which provide further recommendations for how to use this technology. First, we needed to drop a significant number of images because image meta-data was covering the portion of the image in which crops were visible. We recommend that applications used when conducting CCEs allow for these image meta-data to be removed from images themselves, and are stored in auxiliary datasets or in the image filenames instead, since the stamps printed over the images reduce value for future image processing.

Second, we could only use images from harvest time (taken at the time of the CCEs) in our analyses, since there were no images available from other periods during the season. Mechanical damage can have different impacts on crop yields depending on the growth stage during which it occurs, which would require more continuous monitoring. In absence of such data, we cannot answer the question if information on growth stages would have improved both damage predictions and whether growth stages could have contributed to refined crop models, which predict crop yields based on the interaction between weather and growth stages, rather than relying primarily on weather-based data and assuming the timing of growth stages based on observed sowing dates.

Since the ICRISAT model already explains so much of the variation in yields, the images taken at the time of the CCEs do not add much value in estimating yields. And for localized damage, monitoring mid-season adversities, or post-harvest losses, one could imagine a system whereby
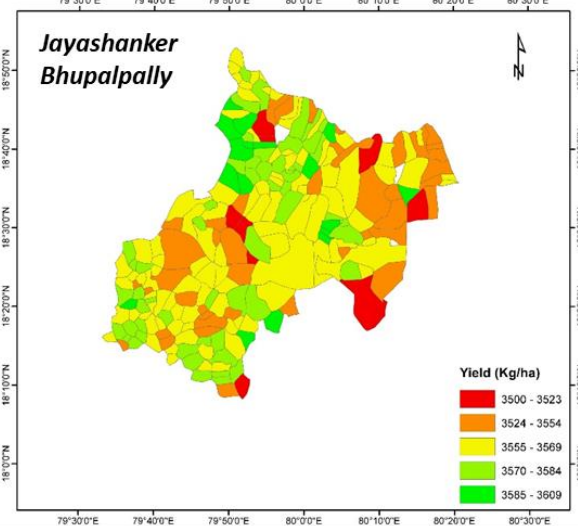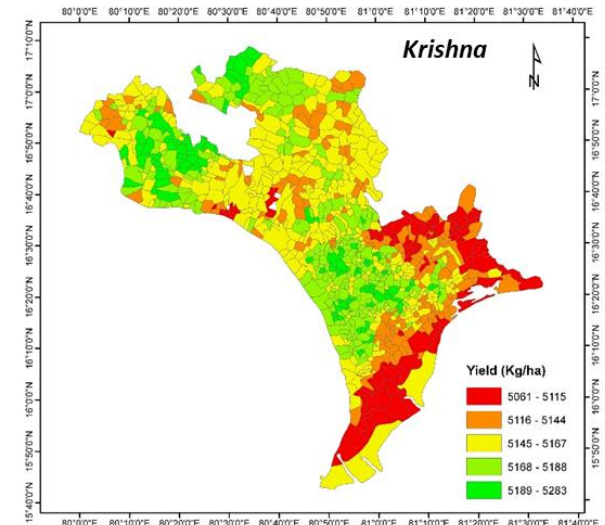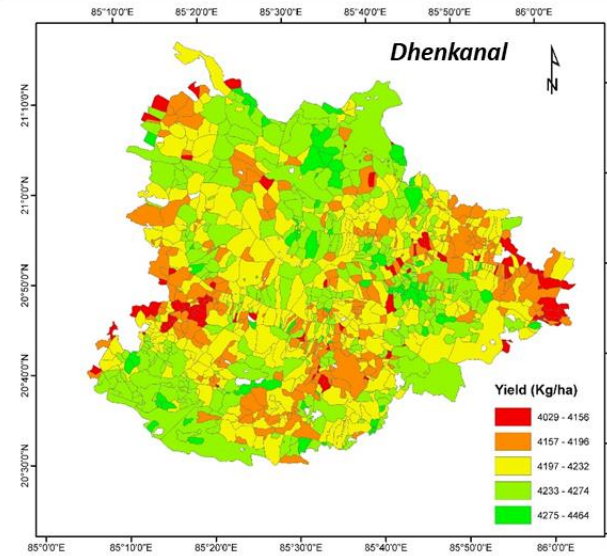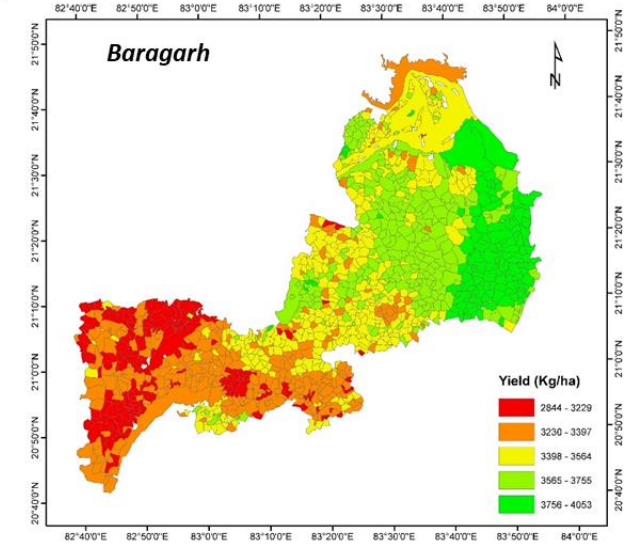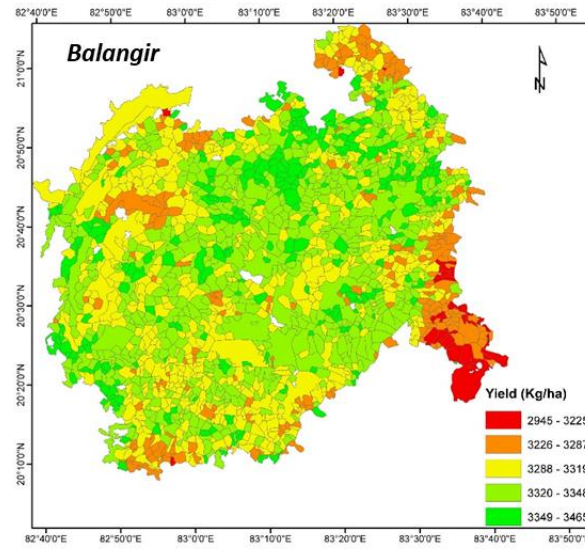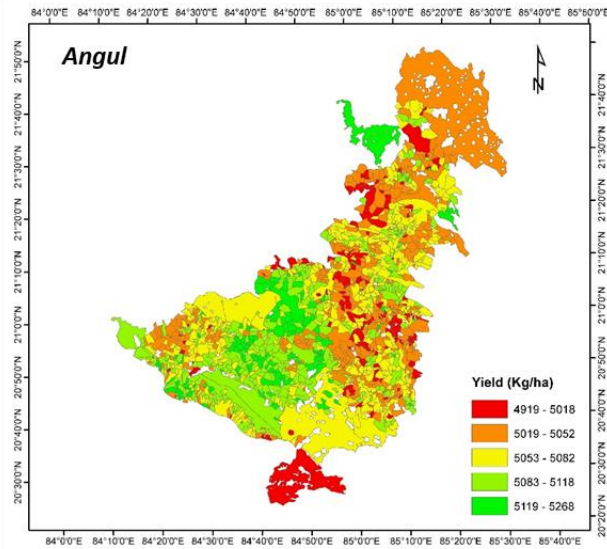
images are taken at the time when the damage occurs; provided that farmers or surveyors can send in these images, our convolutional neural networks are showing that images can be processed accurately and at scale.
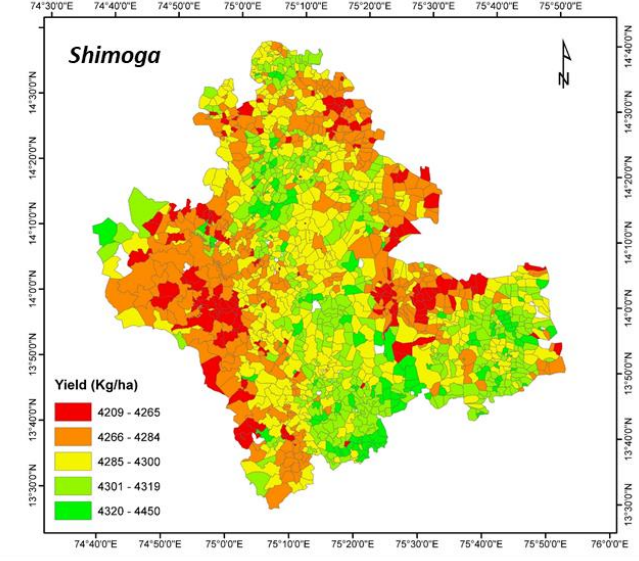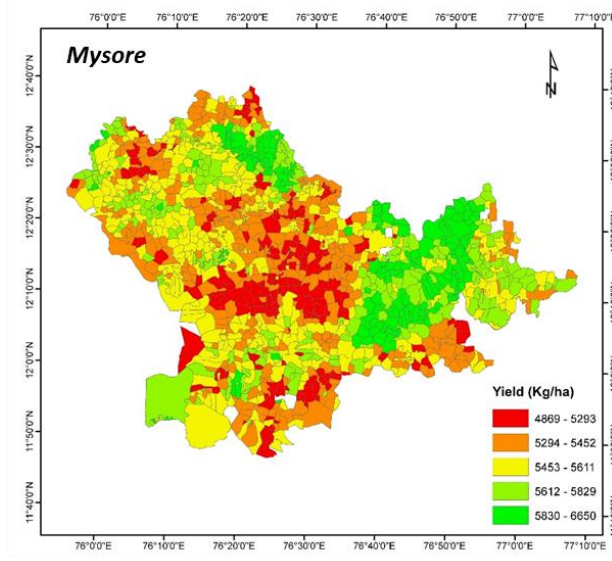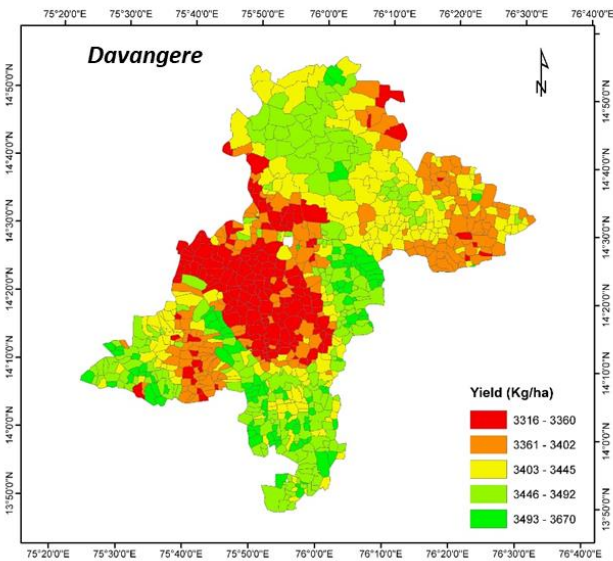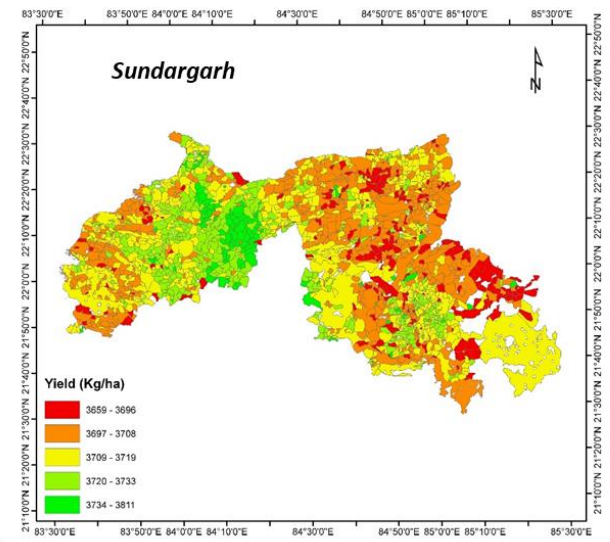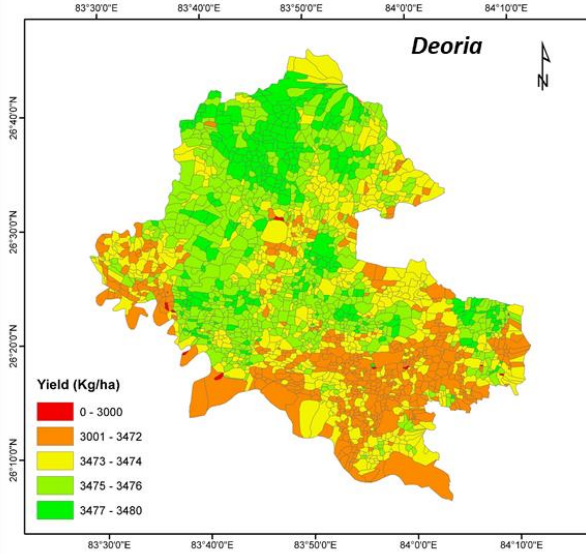
Annexure 1: Spatial Yield Distribution
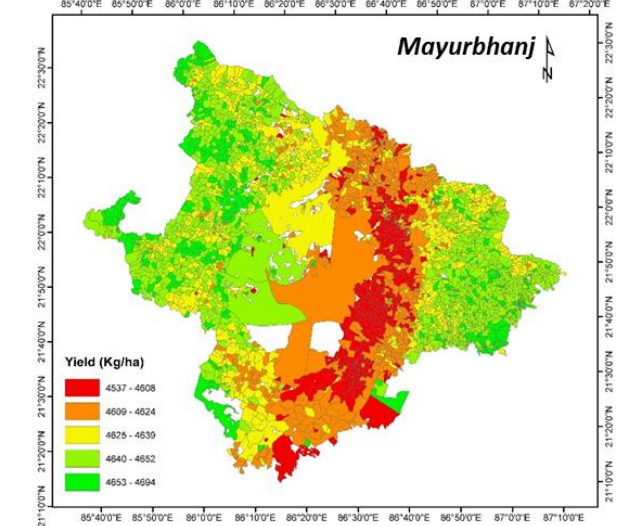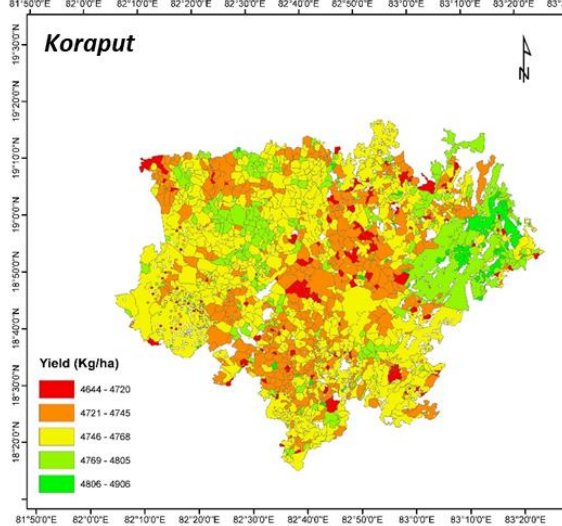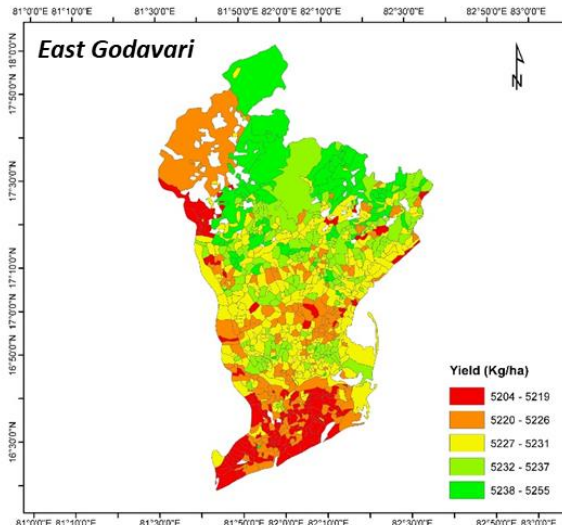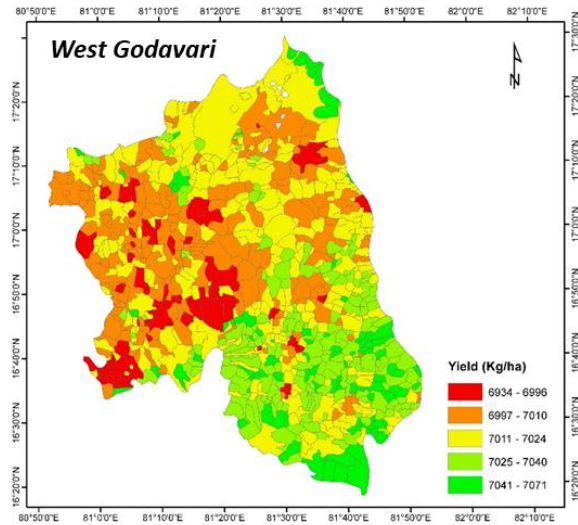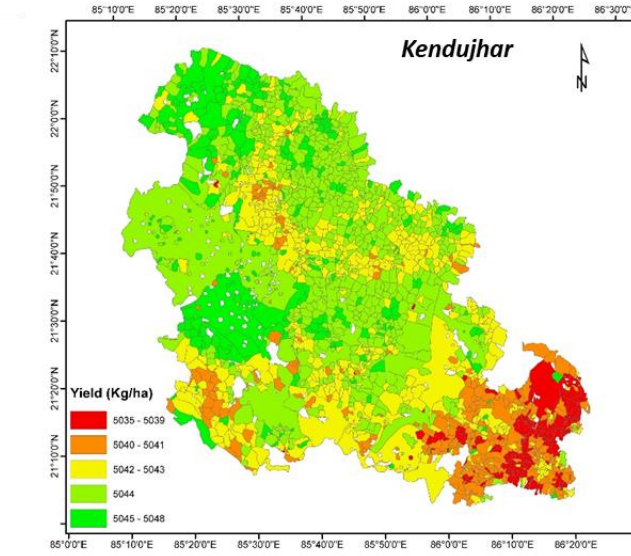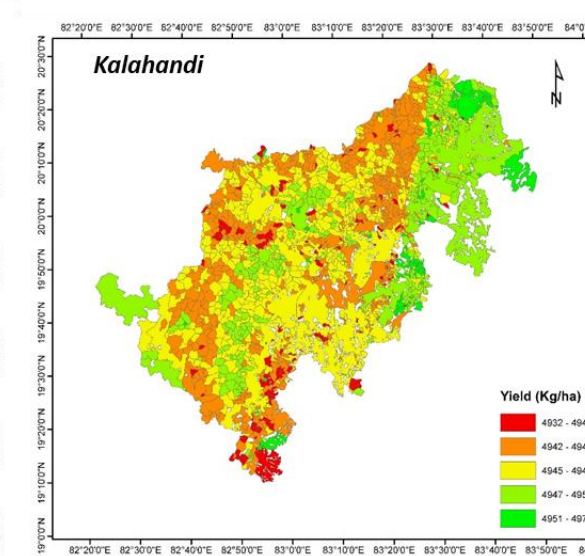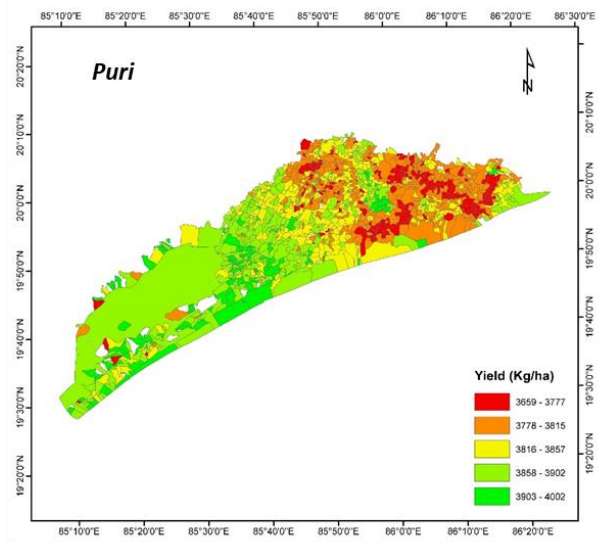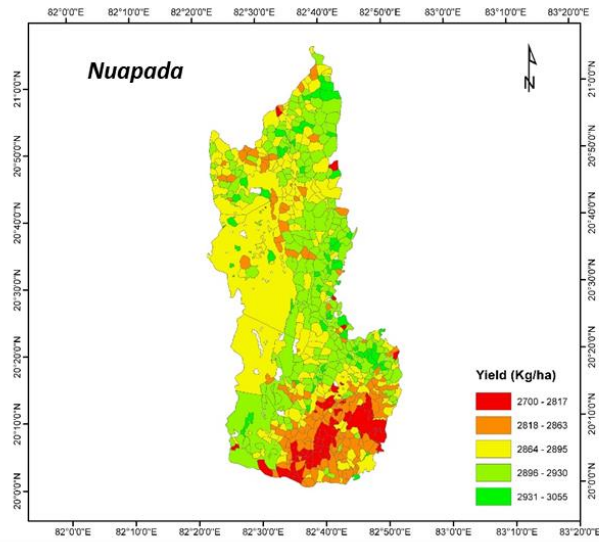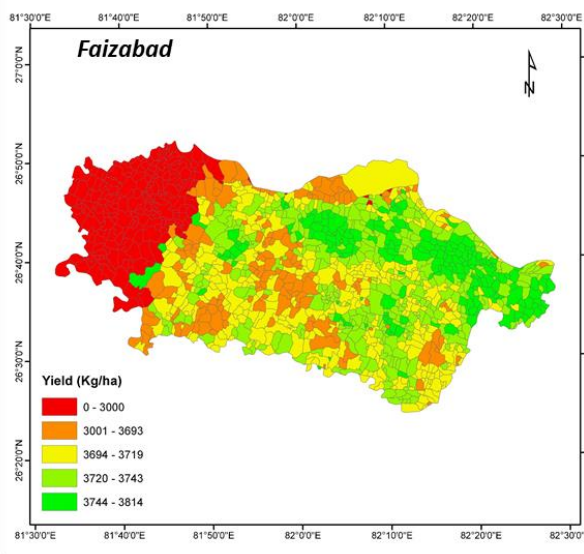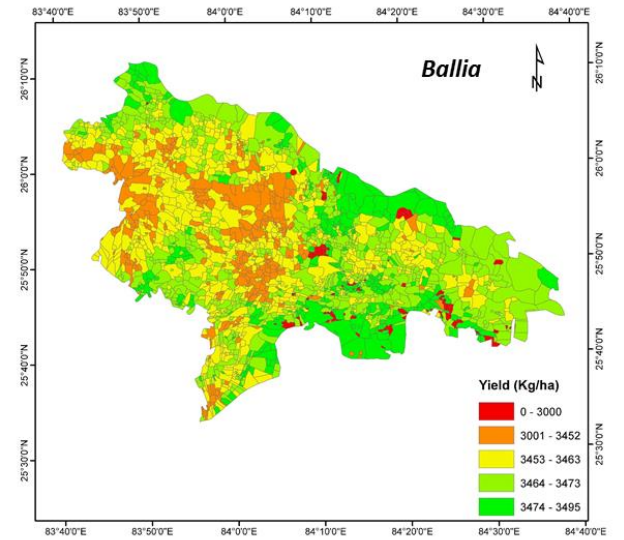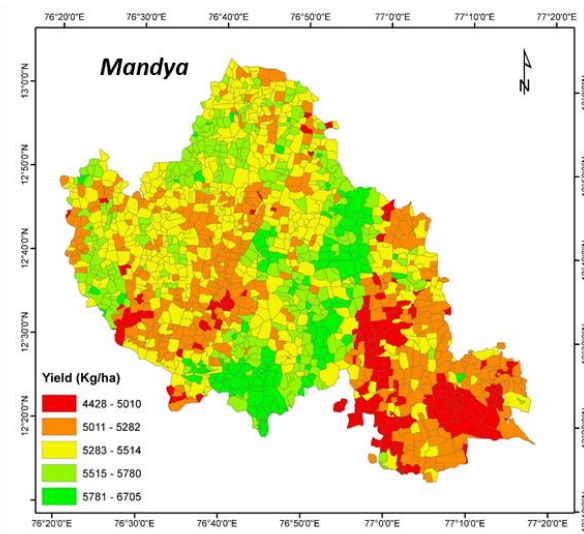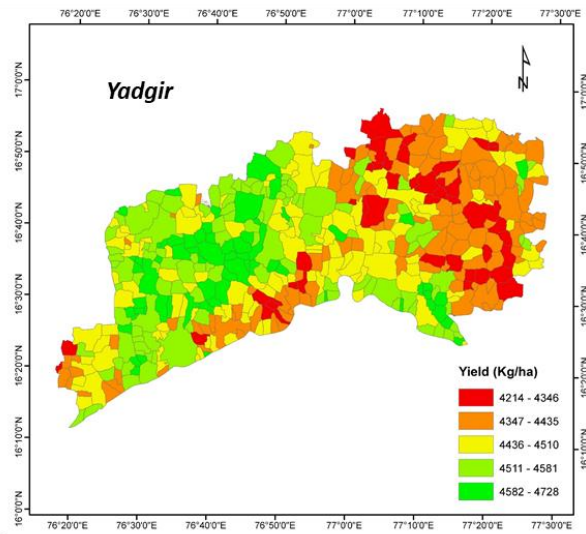
Annexure 2: Paddy Area Estimates

S.no	District Name	Paddy_ Area (Ha)
1	Angul	58060
2	Balangir	145486
3	Ballia	112682
4	Baragarh	108845
5	Davanagere	114717
6	Deoria	89056
7	Dhenkanal	54816
8	East Godavari	183475
9	Faizabad	56780
10	Jayashanker Bhupalpally	47869
11	Kalahandi	99076
12	Kendhujhar	95510
13	Koraput	62773
14	Krishna	280576
15	Mandhya	51784
16	Mayurbhanj	216132
17	Mysore	116300
18	Nuapada	76005
19	Puri	81334
20	Raichur	148407
21	Shimoga	145266
22	Sundargarh	90952
23	Thanjavur	114639
24	West Godavari	213586
25	Yadgir	116336

References

Doraiswamy, P. C., Moulin, S., Cook, P. W., & Stern, A. (2003). Crop yield assessment from remote sensing. *Photogrammetric engineering & remote sensing, 69*(6), 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*(4), 1039-1065.

Gumma, M. K., Thenkabail, P. S., Deevi, K. C., Mohammed, I. A., Teluguntla, P., Oliphant, A., . . . Whitbread, A. M. (2018). Mapping cropland fallow areas in myanmar to scale up sustainable intensification of pulse crops in the farming system. *GIScience & Remote Sensing, 55*(6), 926-949. doi:10.1080/15481603.2018.1482855

Gumma, M. K., Thenkabail, P. S., Teluguntla, P., Rao, M. N., Mohammed, I. A., & Whitbread, A. M. (2016). Mapping rice-fallow cropland areas for short-season grain legumes intensification in South Asia using MODIS 250 m time-series data. *International Journal of Digital Earth, 9*(10), 981-1003.

Gumma, M. K., Thenkabail, P. S., Teluguntla, P., & Whitbread, A. M. (2018). Monitoring of spatiotemporal dynamics of rabi rice fallows in south Asia using remote sensing. *Geospatial Technologies in Land Resources Mapping, Monitoring and Management*, 425-449.

Gumma, M. K., Thenkabail, P. S., Teluguntla, P. G., Oliphant, A., Xiong, J., Giri, C., . . . Whitbread, A. M. (2020). 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*(3), 302-322.

Gumma, M. K., Tummala, K., Dixit, S., Collivignarelli, F., Holecz, F., Kolli, R. N., & Whitbread, A. M. (2022). Crop type identification and spatial mapping using Sentinel-2 satellite data with focus on field-level information. *Geocarto International, 37*(7), 1833-1849.

Gumma, M. K., Uppala, D., Mohammed, I. A., Whitbread, A. M., & Mohammed, I. R. (2015). Mapping Direct Seeded Rice in Raichur District of Karnataka, India. *Photogrammetric Engineering & Remote Sensing, 81*(11), 873-880.

Jin, X., Li, Z., Yang, G., Yang, H., Feng, H., Xu, 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.

Olioso, A., Inoue, Y., Ortega-Farias, S., Demarty, J., Wigneron, J.-P., Braud, I., . . . 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*(3-4), 377-412.