# Macroscale Cotton Yield Estimation in Beed District (Kharif-2023) using Multi-Model Ensembles

**Upasana Singh, Bhargav Sonawane, Priyanka Shamraj and Ashutosh Pawar**

*Ashutosh Pawar: Head, GIS and RS, Semantic Technologies and Agritech services Pvt. Ltd., Pune.*
*Upasana Singh: Sr. GIS and RS executive, Semantic Technologies and Agritech services Pvt. Ltd., Pune.*
*Priyanka Shamraj: Agrometeorologist, Semantic Technologies and Agritech services Pvt. Ltd., Pune.*
*and Bhargav Sonawane: Jr. RS executive Semantic Technologies and Agritech services Pvt. Ltd., Pune.*
*Department of GIS and Remote Sensing, Semantic Technologies and Agritech Services, Pvt. Ltd. Pune.*

---------------------------------------------------------------------***---------------------------------------------------------------------

**Abstract**

This study, conducted by the Semantic Technologies and Agritech Services, Pvt. Ltd., Pune, GIS and Remote Sensing Team in Pune during the Kharif-2023 season, focuses on estimating Cotton crop yield in Beed District. Following the methodology outlined in the YESTECH manual under the Pradhan Mantri Fasal Bima Yojana (PMFBY), the research addresses significant weather-induced yield losses in the region. The study targets Revenue Circle (RC) level assessment using a multimodal approach, incorporating various models for precise yield forecasting. The achieved accuracy, measured with Root Mean Square Error (RMSE) below ±30% at the RC level, demonstrates the effectiveness of the ensemble approach. The findings highlight the utility of such models in decision-making for agricultural stakeholders, insurance companies, and government policies, especially in rainfed regions facing cotton productivity challenges under diverse climate change scenarios.

**Keywords:** Remote Sensing, GIS, Net Primary Productivity (NPP), Machine Learning, (DSSAT-4.8), Cotton, Beed, Yield Simulation, Revenue Circle, Cotton Productivity.

## Introduction:

In today's dynamic agricultural landscape, unpredictable weather patterns, such as erratic rainfall, rising temperatures, and extreme events, pose significant threats to crop growth and yield stability. Consequently, farmers increasingly turn to drought-resistant crops and intensive irrigation, exacerbating soil degradation and amplifying economic vulnerability due to unstable production. Average cotton productivity in Maharashtra, is likely falls between 187 kg/ha and 443 kg/ha. The average cotton productivity in Beed district is 7.15 quintals per hectare (715 kg/ha).

Agriculture serves as the cornerstone of global economies, sustaining livelihoods for billions while presenting critical challenges in accurately predicting crop yields. Traditional methods relying on historical data and manual observations often struggle to address the dynamic nature of modern agricultural challenges. However, the integration of advanced technologies such as software applications, remote sensing, GIS, and AI/ML algorithms has revolutionized crop yield estimation, offering unprecedented accuracy and insight.

Accurate crop yield estimation holds immense significance across sectors in the contemporary landscape. Firstly, in the insurance realm, precise estimates facilitate fair risk assessment, enabling insurers to develop tailored products that alleviate financial burdens on farmers during crop failures. Secondly, in economic forecasting, reliable predictions inform commodity markets, trade agreements, and pricing mechanisms, promoting stability and ensuring food security. Thirdly, governments leverage accurate estimates to formulate effective policies, including subsidy allocation, resource distribution, and strategic interventions during adverse conditions or pest outbreaks, fostering sustainable practices and rural development. Additionally, anticipating potential shortfalls supports proactive food distribution, enhancing access, and averting scarcity. Lastly, for farmers, precise estimates enable informed decisions on crop selection, resource allocation, and market participation, enhancing productivity and livelihoods.

The adoption of advanced methods for crop yield estimation signals a transformative step towards building agricultural resilience. By harnessing the synergy between software applications, remote sensing, GIS, and AI/ML technologies, stakeholders empower informed decision-making, paving the way for sustainable agricultural practices and economic prosperity. This report emphasizes the significance of employing advanced methods for estimating crop yield and its implications across diverse domains.

The objectives include estimating the area under major kharif crops in the Beed district, crop classification using remote sensing and GIS techniques, and estimating crop yield through a combination of models including RS and GIS, Artificial Intelligence, Google Earth Engine, ground truth, and DSSAT software.

In conclusion, by recognizing the multifaceted implications of accurate crop yield estimation, societies can collaboratively strengthen global food security, economic stability, and the welfare of farming communities.

## 1. Material and Methods:

**Study area:**

Study was carried out at Semantic Technologies and Agritech Services, Pvt. Ltd., Pune during *kharif* season 2023 for particular assignment. For this study, all revenue circles (RC) in the districts of Beed of Maharashtra state were used as experimental sites. Field level data like ground truth, Crop cutting experiments were carried out.

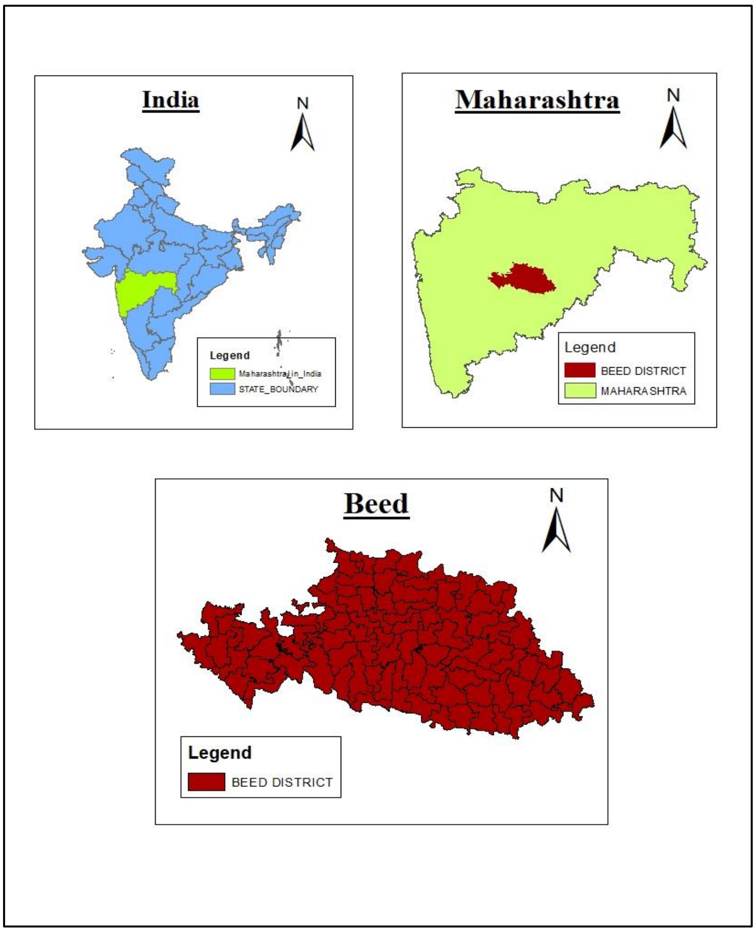


**Figure 1: Study Area**

**Geography and Climate for Beed District:**

Beed district, located in the state of Maharashtra, India, spans an area of approximately 10,693 square kilometres. Its geographical coordinates are approximately 18.9906° N latitude and 75.7531° E longitude, with an average elevation of 540 meters above sea level. The district experiences a semi-arid climate with hot summers and cool winters. The annual rainfall typically ranges from 600 to 800 millimetres, primarily occurring during the monsoon season. Temperatures vary widely throughout the year, with average highs peaking around 40°C during the summer months, while winter temperatures can drop to around 10°C in December and January. Humidity levels tend to be relatively lower during the drier months, especially in winter, with higher humidity levels experienced during the monsoon season. The predominant soil types include black soil, red soil, and alluvial soil, supporting the cultivation of crops such as cotton, sorghum, pulses, and soybeans. Beed is bordered by the districts of Ahmednagar, Osmanabad, Aurangabad, and Jalna. The major rivers flowing through the district include the Godavari and the Sindhphana.

**Methodology:**

All methodology was followed by the procedure given by yield estimation system based on technology (YES-TECH) under Pradhan Mantri Fasal Bima Yojana (PMFBY).

Methodology used is multimodal approach for estimation of crop yield was given below. RC wise yield in Tonnes/hector of cotton crop during *kharif* season 2023 was estimated by all following methods.

1) Semi Physical NPP- Net Primary Productivity
2) AI and Machine learning
3) Crop simulation model-DSSAT-4.8
4) Ensemble Model

**1) Semi Physical Net Primary Productivity (NPP):**

**Data and materials used:**
The data and materials used in this study are as follows:

**Table 1: Data used for NPP generation in Semi Physical model.**

| Data | Satellite/Ground | Resolution | Source |
|---|---|---|---|
| Daily insolation/PAR | INSAT-3D | 4km resampled to 1km | MOSDAC |
| 10 days composite fAPAR ver. 2 | PROBA V and SPOT-VGT | 1km | Copernicus Land Service |
| 8 days composite surface reflectance | Terra-MODIS | 1km | MODIS Time Series Tool |
| Paddy Mask | Sentinel 1 | 5m | USGS Explorer |
| Temperature | Gridded data from NASA Power website | 1km interpolated | NASA Power |
| Light-use efficiency | | | Literature |
| Harvest Index | Ground | CCE | |

**Fraction of Absorbed PAR (FAPAR):**
The FAPAR data is from Copernicus Land Service, source link is (https: //land. copernicus. eu/global/index. html). the 10 - day composite product with 1 km data is used. The range of FAPAR lies between 0 and 1. The physical values are retrieved from the Digital Number (DN).
**Photosynthetically Absorbed Radiation (PAR):**
PAR is calculated from daily insolation data. The daily insolation data is converted to 8 - day composite (sum) for the whole period. 50% insolation is considered as PAR. This daily insolation data is collected from MOSDAC from INSAT - 3D satellite, source link (www.mosdac. gov. in) for the crop season from 2018 to 2022.

*PAR=* ***8 - day composite * 0.5.***

**Water Stress (Wstress):**
The Wstress is calculated from Land Surface Water Index (LSWI). The MODIS time series tool (MODIStsp) used to download and process the MODIS 8 day composite (MOD09A1) source link is (https: //lpdaac. usgs. gov/products/mod09a1v006), and LSWI is calculated for the entire period with the formula

$$LSWI = (\rho NIR - \rho SWIR) / (\rho NIR + \rho SWIR)$$

LSWI value range from - 1 to 1, and higher positive values indicate the vegetation and soil water stress. Further, the Wstress is calculated from 8 days LSWI output –

$$Wstess = (1 - LSWI) / (1 + LSWImax)$$

The LSWImax value has been taken from the spatial maximum of particular crop mask of the entire district.

**Temperature Stress:**

Temperature Stress (Tstress): The daily average temperature data is downloaded from NASA Power website, source link is (https: //power. larc. nasa. gov/data - access - viewer. html). It is a gridded data with a resolution of 1°0 * 1°0 latitude and longitude.

$$T\ Stress = \frac{(T - Tmin)*(T - Tmax)}{[(T - Tmin)*(T - Tmax) - T - Topt)^2]}$$

Where, Tmin = Minimum temperature required for the photosynthesis (°C).
Tmax =Maximum temperature required for the photosynthesis (°C).
Topt = Optimal temperature required for the photosynthesis (°C);
T = Daily mean temperature (°C).

**Table 2: Data used for cotton crop for Semi-Physical Approach.**

| Sr.No. | Particulars | Values | Source | Sr.No. | Particulars | Values | Source |
|---|---|---|---|---|---|---|---|
| 1 | T maximum | 40°C | (Prasad, et al., 2022) | 4 | LUE | 1.53 | (Prasad, et al., 2022) |
| 2 | T minimum | 10°C | | 5 | Harvest Index | 0.12 | Periodic CCE data. |
| 3 | T optimum | 25°C | | | | | |

On the off chance that air temperature falls beneath Tmin, which is quite a rare chance than Tscalar value will automatically become 0.

**Light Use Efficiency (Ɛ):**

The light use efficiency LUE is used for cotton crop was 1.53 for the study. (Chavan et al., 2018)

**Crop Mask**

The crop mask was derived utilizing Sentinel-1 synthetic aperture radar (SAR) data obtained from the European Space Agency (ESA) Copernicus Hub. Employing the R programming language, we employed the Random Forest algorithm for the generation of the crop mask, implementing hyperparameter tuning techniques and contingency matrix analysis. This methodology was systematically applied across our specified crops within the targeted area of interest.

In terms of accuracy assessment, our results yielded a robust accuracy range of 90% to 95% across all cultivated crops and within various districts. This signifies a high level of precision in delineating and classifying the specified crops within the delineated geographical regions. The meticulous incorporation of Random Forest algorithm, hyperparameter tuning, and contingency matrix analysis has facilitated the generation of a reliable and accurate crop mask, providing valuable insights for agricultural monitoring and management within the designated study area.

**Calculation of NPP and Grain Yield**:

To compute the final Net Primary Productivity NPP and its Grain Yield, the formula and equation is used as follows. The NPP sum has been multiplied with Harvest Index (0.12) to estimate per pixel yield.

***NPP*** *= PAR * FAPAR * Ɛ * Tstress * Wstress* (Logic of Monteith Equation 1972).

Same methodology is followed by Upasana Singh *et.al*. (2023) and also showing same results for all data used to run the model.

**2) Crop simulation model-DSSAT**

Material and method and all file process was carried out by the procedure followed by Hoogenboom, G., et.al (2019) and (2024) Jones, J.W., (2003) and the minimum data requirements for operation, calibration and validation of the Crop models are described below.

Crop simulation model is a mathematical equation or the set of equations, which represents the behaviour of system. We used CROPGRO – for Cotton crop. It is consisting of various subroutines *viz*., Water balance subroutine, Phenology subroutine, Nitrogen subroutine, and Growth and Development subroutine described below.

**Data input to model**

The minimum data requirements for operation, calibration and validation of the Crop models are described below.

**Table 3: Showing List of input required by crop simulation model.**

| Sr.No. | Input variables | Acronym | Source |
|---|---|---|---|
| **1.** | **SITE DATA** | | |
| | Latitude | LAT | NASA power |
| | Longitude | LONG | NASA power |
| | Elevation | ELEV | NASA power |
| **2.** | **DAILY WEATHER DATA** | | |
| | Maximum temperature | TEMPMAX | NASA power |
| | Minimum temperature | TEMPMIN | NASA power |
| | Solar radiation | SOLARAD | NASA power |
| | Rainfall | RAIN | NASA power |
| **3.** | **SOIL CHARACTERISTICS** | | |
| | Soil texture | SLTX | DSSAT website Where Global gridded-soil profile dataset at 10-km resolution was Developed for DSSAT-4.8 Software crop simulation models. |
| | Soil local classification | SLDESC | |
| | Soil depth | SLDP | |
| | Colour, moist | SCOM | |
| | Albedo (fraction) | SALB | |
| | Photosynthesis factor (0 to 1 scale) | SLPE | |
| | pH in buffer determination method | SMPX | |
| | Potassium determination method | SMKE | |
| | **Horizon-wise** | | |
| | Lower limit drained | LL(L) | |
| | Upper limit drained | DUL(L) | |
| | Upper limit drained | SAT(L) | |
| | Saturated hydraulic conductivity | SWCN(L) | |
| | Bulk density moist | BD(L) | |
| | Organic carbon | OC(L) | |
| | Clay (<0.002 mm) ` | CLAY(L) | |
| | Silt (0.05 to 0.002 mm) | SILT(L) | |
| | Coarse fraction (>2 mm) | STONES(L) | |
| | Total nitrogen | TOTN(L) | |
| | pH in buffer | PHKCL(L) | |
| | Cation exchange capacity | CEC(L) | |
| | Root growth factor 0 to 1 | SHF(L) | |

| 4. | MANAGEMENT DATA | | |
|---|---|---|---|
| | Sowing date | YRPLT | Krishi-Dainandini Published by in Vasantrao Naik Marathwada Krishi Vidypeeth , Parbhani, |
| | Plant population at seedling | PLNATS | |
| | Planting method (TP/direct seeded) | PLME | |
| | Row spacing | ROWSPS | |
| | Row direction (degree from north) | AZIR | |
| | Seed rate | SDWTRL | |
| | Sowing depth | SDEPTH | |
| | Irrigation dates | IDLAPL | |
| | Irrigation amount | AMT(J) | |
| | Method of irrigation | IRRCOD | |
| | Fertilizer application dates | FDAY(J) | Krishi-Dainandini Published by in Vasantrao Naik Marathwada Krishi Vidypeeth , Parbhani, |
| | Fertilizer amount N | ANFER | |
| | Fertilizer type | IFTYPE | |
| | Fertilizer application method | FERCOD | |
| | Fertilizer incorporation depth | DFERT | |
| | Tillage date | TDATE | |
| | Tillage implements | TIMPL | |

**Input files**

The files are organized into input, output and experiment performance data file. The experiment performance files are needed only when simulated results are to be compared with data recorded in a particular experiment. In some cases, they could be used as input files to reset some variable during the course of a simulation run. The input files are further divided into those dealing with the experiment, weather and soil and the characteristics of different genotypes. Similarly output files are also further divided into those dealing with the overview, summary, growth, water, carbon and nitrogen balance.

**Soil properties directory file**: The file SOIL.SOL contained the list of different soils with their physical and chemical properties.

**Soil profile initial condition file:** The soil profile initial condition file contained the initial values of soil water, soil reaction and soil nitrogen data pertaining to this situation was entered.

**Irrigation management file:** The Irrigation management file has the provision of date and amount per fixed irrigation (mm) applied depth (cm) of management. Irrigation data pertaining to this situation was entered.

**Fertilizer management file:** The fertilizer management file contained the date, form and amount of nitrogen application. Accordingly, information on fertilizer application was entered in the file.

**Treatment management file:** The treatment management file contained the description of each treatment under separate title and serial numbers. The file also contained dates of planting and emergence, plant population at seeding and at emergence, planting method, planting distribution, row spacing, row direction, planting depth, planting material, transplant age, plants per hill, dates of simulation beginning etc. All needed information was entered for all the treatments.

**Crop cultivars directory file**

For Cotton CRGRO048 contained the list of different cultivars with their genetic coefficients. The modified genetic coefficients viz., CSDVAR, PPSEN, EMG-FLW, FLW-FSD, FSD-PHM, WTPSD, SDPDVR, SDFDUR, PODDUR, THRESH, SDPRO and SDLIP is used. Variety selected was JS-335 which is mostly used in this area.

The genetic coefficients are the most important parameters which represents the genetic characteristics of the cultivar and on which the crop phenology, biomass production partitioning and yield potential of the crop depends. However, the actual performance is controlled by the external factors also.

**Running the crop model:** Once, all the desired files were created carefully the model was run for all the crops cultivars. Each run of model created output files.

**3) Machine learning:**

Methodology and processing of model is described below in details.

**Data Collection and Ground Truthing:**

- Collect remote sensing data (optical and radar imagery) for the study area, covering the growing season of the crops.
- Ground truth data collection using field surveys using CropTech App ( prepared by compony) for accurate calibration and validation.

**Crop Mask Extraction:**

- Pre-process the remote sensing data to correct for atmospheric interference and geometric distortions.
- Apply image enhancement techniques to improve the visual quality of the images.
- Employ supervised or unsupervised classification algorithms to extract crop masks for Cotton fields.

**Generation of Spectral Indices and use of RADAR backscatter:**

- Calculate vegetation indices (e.g., NDVI, NDRE, GNDVI) from the optical remote sensing data to assess crop health and Vigor.
- Utilize backscatter data from radar imagery to analyse surface roughness and other relevant crop information (VV, VH).

**Crop Cutting Experiments:**

- Use of Crop Cutting Experiment (CCE) for Crop with smart sampling methods to efficiently estimate crop parameters for crop.

**Training and Testing Models (Machine Learning):**

- Divide the dataset into training and testing sets, ensuring no overlap between the two.
- Evaluate the model's performance on the testing dataset using evaluation metrics like accuracy, F1-score, and mean squared error (RMSE).

**Model Validation and Final Result:**

- Validate the trained model using independent ground truth data collected during the growing season for Cotton.

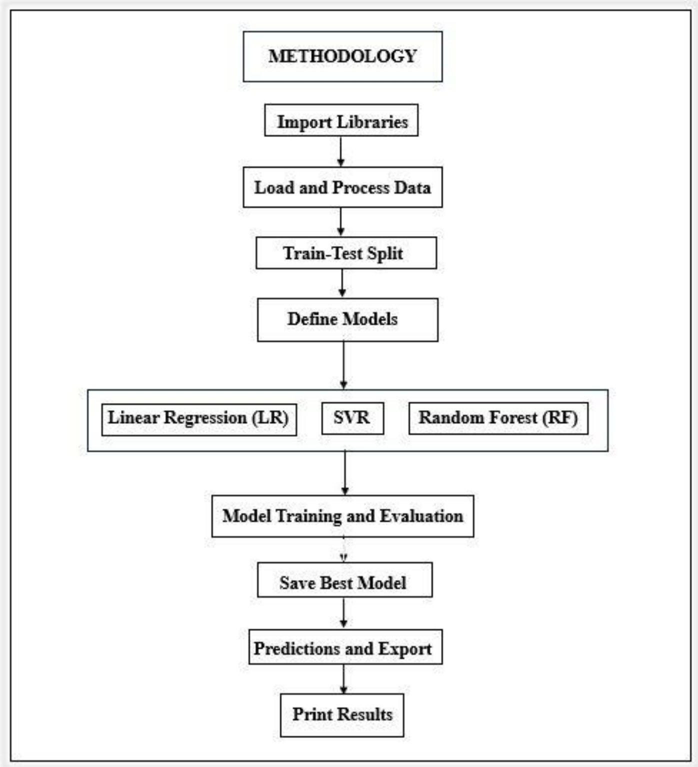


**Figure 2: Methodology used in Machine learning Approach.**

- Assess the model's accuracy and generalization ability to ensure reliable yield estimation.
- Obtain the final crop yield estimation results for Cotton in the study area.

**4) Ensemble Models**

This methodology aims to combine the predictive power of both Machine Learning (ML) models and Crop Simulation Models (CSM) to provide an enhanced and more accurate estimation of crop yields. Here is a structured approach:

1. **Data Collection and Preprocessing**:

- Gather data from both ML, Semi-Physical Approach and CSM approaches as outlined in the above methods.
- Consolidate all input data: weather data, soil properties, crop management practices, spectral indices, RADAR backscatter, and ground truth data.
- Ensure data alignment in terms of temporal and spatial granularity.

2. **Individual Model Generation**:

**a) Machine Learning Approach**:

- Utilize various algorithms like Linear regression, Random Forest, Extra Trees, k-earest neighbours, and neural networks.
- Train these models on the dataset ensuring proper validation and calibration.

**b) Crop Simulation Approach**:

- Use well-calibrated crop simulation models such as DSSAT.
- Simulate the growth and yield of crops using these models based on provided input data.

c) **Semi-physical Models**: A semi-physical model in remote sensing and GIS is a type of model that combines physical principles with remotely sensed data to estimate or predict biophysical parameters, such as crop yield, biomass. These models are often used to monitor and manage natural resources, as well as to assess the impacts of climate change and other environmental stressors.

3. **Ensemble Techniques Application**:

- **Model Averaging**: Calculate the simple mean of predictions from ML, semi- physical model and CSM models.
- **Weighted Averaging**: Assign weights based on individual model performance and calculate the weighted average of predictions.
- **Stacking**: Use a meta-model that takes predictions from individual models as inputs and predicts the final yield.
- **Voting**: Each model votes for a final yield prediction, and the most frequent prediction is considered.

4. **Model Validation**:

- Split the dataset into training, validation, and test sets to avoid overfitting and ensure generalizability.
- Use metrics like Root Mean Squared Error (RMSE), and R-squared (R2) for evaluation.
- Assess performance using the test dataset and ground truth data.

5. **Quality Control**:

- Calculate the normalized RMSE between the observed and ensemble model's estimated yield.
- Ensure RMSE does not exceed acceptable thresholds, refining the model if necessary.

**Validation**:

The accuracy of our model was evaluated based on crop cutting experiment data (CCE data) of PMFBY (Pradhan Mantri Fasal Bima Yojana) for the crop season *kharif*-2023.

**Results and Discussion:**

Following were the results and conclusion for different methods/models used for estimation of yield of soyabean crop in Beed districts of Maharashtra, Revenue-Circle wise.

**1) Semi Physical Approach-NPP :**

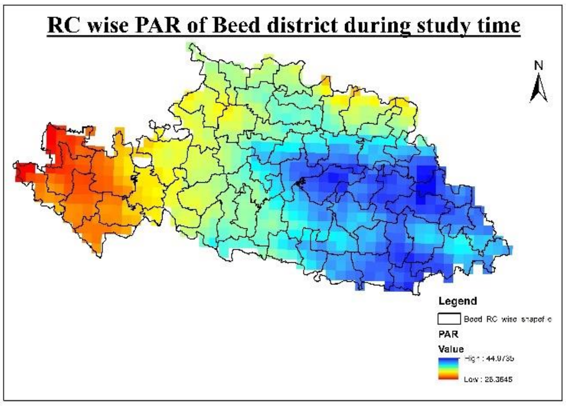


**Fig. 3: PAR for Beed during *kharif* 2023**

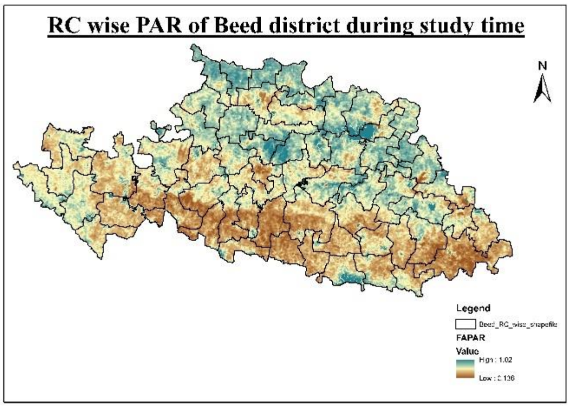


**Fig.4: FAPAR for Beed during *kharif* 2023**

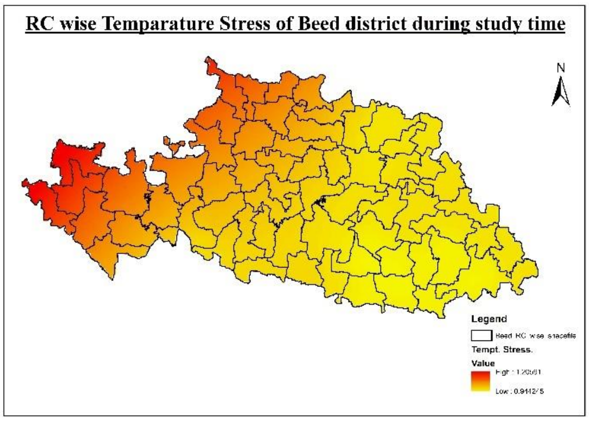


Fig.5: Tstress for Beed during *kharif* 2023

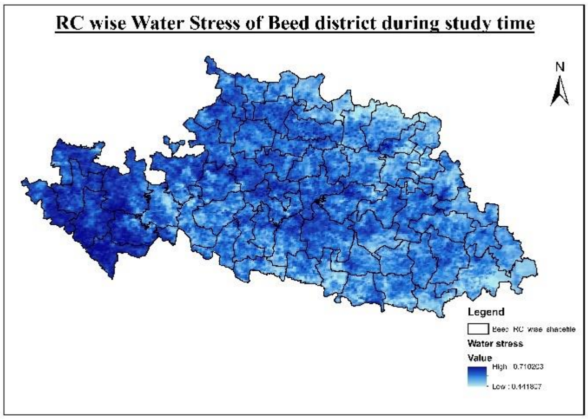


Fig. 6: Waterstress for Beed during *kharif* 2023

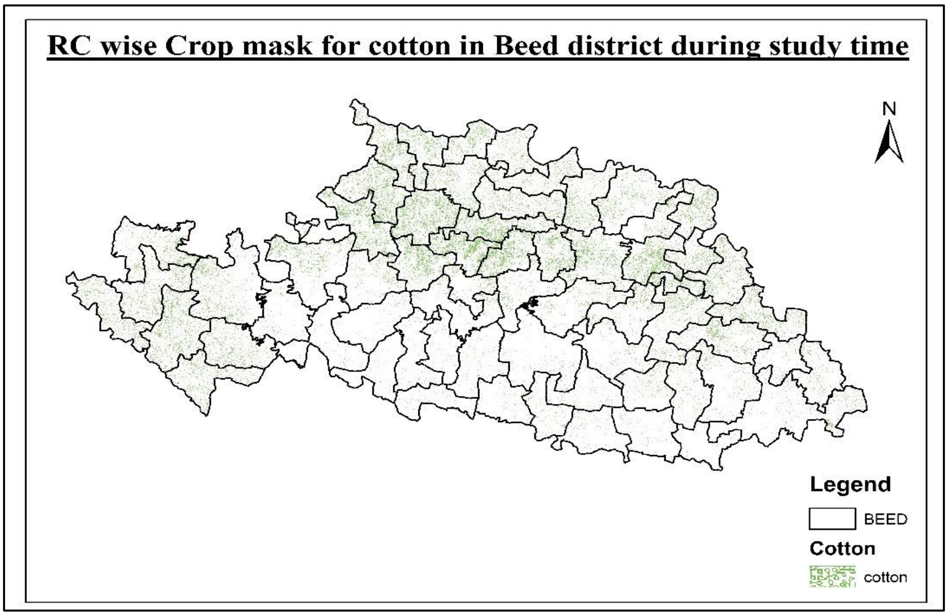


Fig. 7: Cotton Crop Mask of Beed during *kharif* 2023

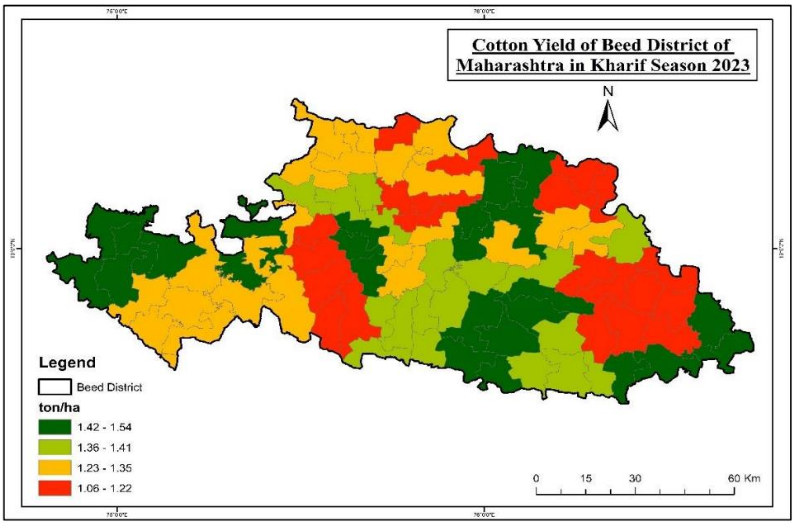


Fig. 8: Cotton yield of Beed during *kharif* 2023 by NPP method

- The average cotton yield across Beed District for the kharif 2023 season was approximately 1.33 tonnes per hectare.
- Yields varied significantly across different regions within Beed District, ranging from 1.06 to 1.54 tonnes per hectare.
- Some of the top yielding areas include Pachegaon, Manjarsumba, and Neknoor, with yields ranging from 1.45 to 1.54 tonnes per hectare.
- Lower yields were observed in areas like Mohkhed, Nagapur, and Pimpalgaon, with yields ranging from 1.06 to 1.09 tonnes per hectare.
- There seems to be variability in yield consistency across the district, with some areas showing stable production while others exhibit fluctuations. Various factors such as soil quality, irrigation practices, and agricultural techniques may contribute to the differences in cotton yields across different regions. Same results were reported by Xiao, X., et.al (2006) and Yao, Y., et.al (2021)

**2) Crop Simulation Model DSSAT-4.8**

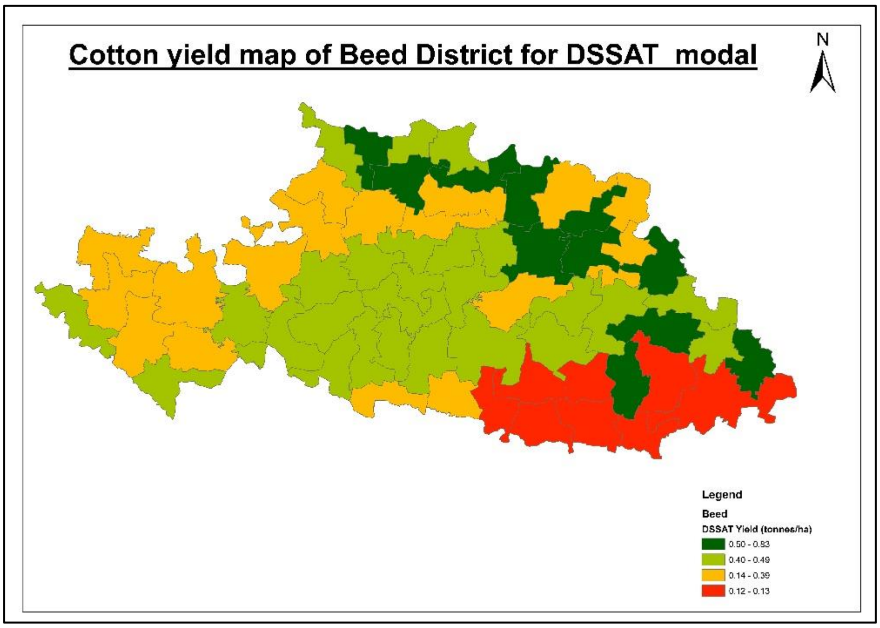


**Fig. 9: Soybean yield in T/ha by DSSAT for Beed during *kharif* 2023**

- The average cotton yield for Beed District in the kharif season of 2023 was 0.437 tonnes per hectare, with yields varying from 0.12 to 0.83 tonnes per hectare across different regions.
- Notable high-yielding areas include Dharmapuri, Majalgaon, and Jategaon, while lower yields were observed in Pimpalgaon, Nagapur, and Yusufwadgao.
- Yield consistency varies across the district, with some areas showing stability in production while others exhibit variability.
- Factors influencing yield disparities encompass soil quality, irrigation practices, fertilization methods, and pest management strategies.
- Opportunities for improvement lie in the adoption of enhanced agricultural practices, infrastructure development, and modern technology to bolster yields and enhance livelihoods for farmers in Beed District. Jadhav, S. D et.al (2018), Bhosale, A. D., et.al (2015) and Deshmukh, S. D., et.al (2013) also elaborated same results for soybean.

**3) Machine learning**

- CCE yield and different indices under study showing accuracy 78 % in Machine learning model. By the method (LR) (logistic regression) accuracy is showing highest value.
- The average cotton yield across Beed District for the kharif 2023 season was approximately 0.731 tonnes per hectare.
- Yields varied significantly across different regions within Beed District, ranging from 0.57 to 0.89 tonnes per hectare.
- Some of the top yielding areas include Nithrud, Dharmapuri, and Majalgaon, with yields ranging from 0.78 to 0.89 tonnes per hectare.
- Lower yields were observed in areas like Madalmohi, Pimpalgaon, and Nagapur, with yields ranging from 0.57 to 0.64 tonnes per hectare.
- There seems to be variability in yield consistency across the district, with some areas showing stable production while others exhibit fluctuations.

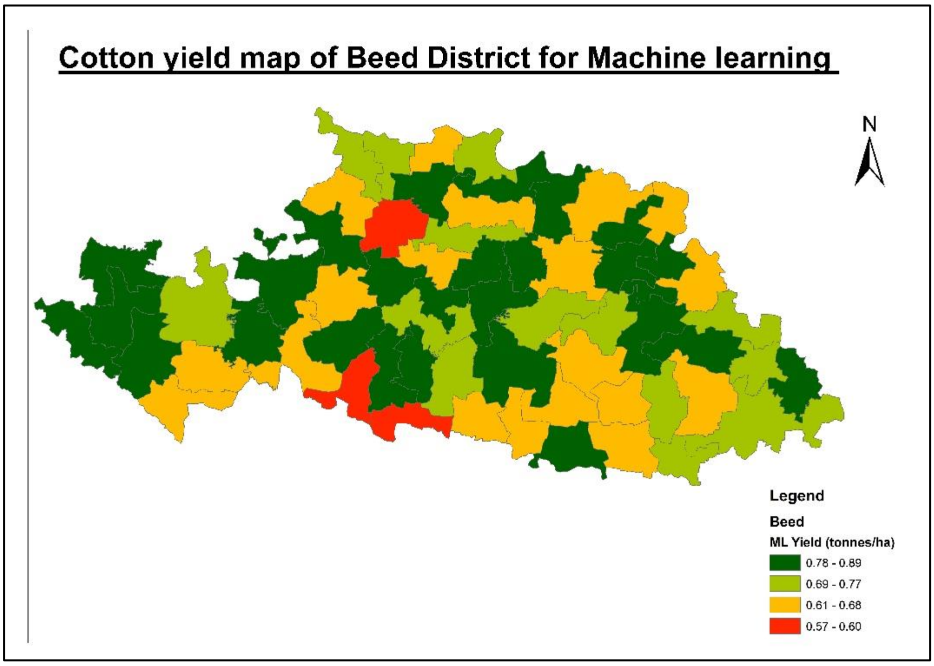


**Fig. 10: Cotton yield in T/ha by ML for Latur during *kharif* 2023.**

- Various factors such as soil quality, irrigation practices, and agricultural techniques may contribute to the differences in cotton yields across different regions.

**4) Ensemble Model:**

- The Ensemble Yield represents a combination of all above three predictive models or methods to estimate soybean crop yield.

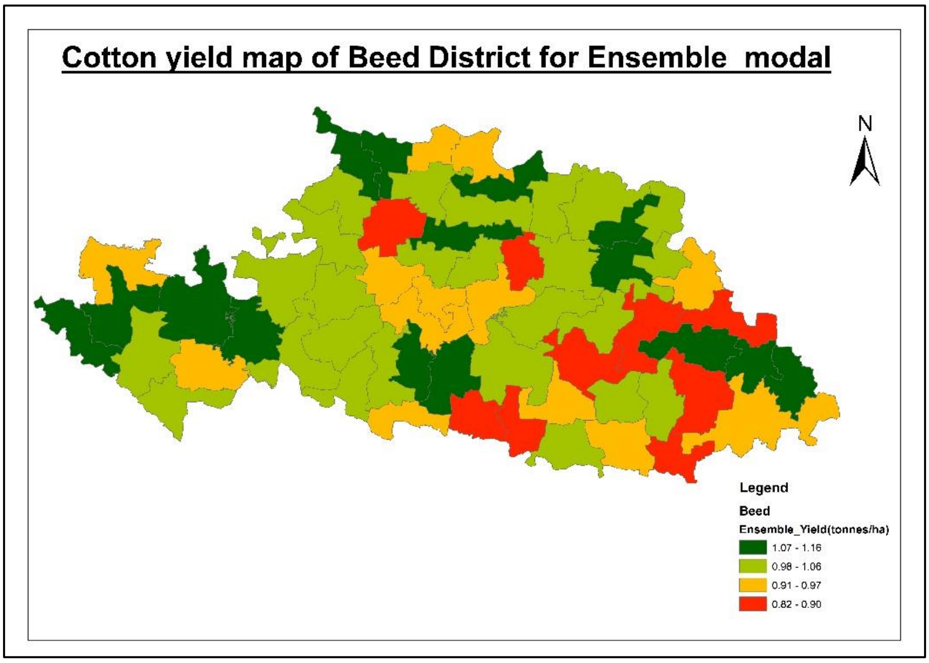


**Fig. 11: Cotton yield in T/ha by Ensemble Model for Beed during *kharif* 2023.**

- Statistical approach give weightage during kharif 2023 as following to different models.

| Model Used | DSSAT Yield | Semi-Physical Yield | Machine Learning Yield |
|---|---|---|---|
| Weightages in % | 16.65 | 53.71 | 29.64 |

- The average cotton yield across Beed District for the kharif 2023 season was approximately 0.992 tonnes per hectare.
- Yields varied significantly across different regions within Beed District, ranging from 0.82 to 1.16 tonnes per hectare.
- Some of the top yielding areas include Dharmapuri, Jategaon, and Majalgaon, with yields ranging from 1.09 to 1.16 tonnes per hectare.
- Lower yields were observed in areas like Nagapur, Pimpalgaon, and Madalmohi, with yields ranging from 0.87 to 1.09 tonnes per hectare.
- There appears to be variability in yield consistency across the district, with some areas showing stable production while others exhibit fluctuations.
- Various factors such as soil quality, irrigation practices, and agricultural techniques may contribute to the differences in cotton yields across different regions. Same results were given by Md Didarul Islam et.al (2023), Liujun Xiao et.al. (2022) and Ayan Das a et.al (2023) in both Machine learning and ensemble approach.

**Table 4: Estimated Yield of Cotton crop in Tones/Hectors with different Models and percent error with ensemble model for year 2023.**

| District | Tehsil | RC | Field CCE | DSSAT Yield | Semi-Physical Yield | ML Yield | Ensemble Yield | RMSE % |
|---|---|---|---|---|---|---|---|---|
| Beed | Ambejogai | Ambajogai | 0.81 | 0.12 | 1.21 | 0.65 | 0.86 | 20.1 |
| Beed | Ambejogai | Ghatnandur | 1.15 | 0.13 | 1.32 | 0.77 | 0.96 | 33.3 |
| Beed | Ambejogai | Lokhandi- Sawargaon | 0.86 | 0.62 | 1.37 | 0.75 | 1.06 | 12.3 |
| Beed | Ambejogai | Patoda M. | 0.58 | 0.13 | 1.21 | 0.77 | 0.90 | -32.6 |
| Beed | Ashti | Ashti | 0.95 | 0.39 | 1.31 | 0.62 | 0.95 | 34.3 |
| Beed | Ashti | Daulawadgaon | 0.53 | 0.35 | 1.21 | 0.81 | 0.95 | -51.9 |
| Beed | Ashti | Dhamngaon | 0.79 | 0.39 | 1.47 | 0.76 | 1.08 | 3.6 |
| Beed | Ashti | Dhanora | 0.72 | 0.39 | 1.47 | 0.79 | 1.09 | -10.2 |
| Beed | Ashti | Kada | 0.70 | 0.39 | 1.34 | 0.79 | 1.02 | -13.8 |
| Beed | Ashti | Pimpla | 0.73 | 0.46 | 1.50 | 0.79 | 1.11 | -7.9 |
| Beed | Ashti | Takalsing | 0.90 | 0.48 | 1.35 | 0.65 | 1.00 | 27.9 |
| Beed | Beed | Beed | 1.01 | 0.43 | 1.14 | 0.80 | 0.92 | 20.6 |
| Beed | Beed | Chousala. | 0.78 | 0.34 | 1.37 | 0.60 | 0.97 | 23.5 |
| Beed | Beed | Limbaganesh. | 0.64 | 0.43 | 1.35 | 0.80 | 1.03 | -23.7 |
| Beed | Beed | Mahlas Jawala. | 0.59 | 0.43 | 1.32 | 0.83 | 1.03 | -41.6 |
| Beed | Beed | Manjarsumba. | 0.92 | 0.43 | 1.50 | 0.79 | 1.11 | 14.0 |
| Beed | Beed | Nalwandi. | 0.82 | 0.44 | 1.14 | 0.80 | 0.93 | 1.6 |
| Beed | Beed | Neknoor. | 0.85 | 0.43 | 1.53 | 0.75 | 1.11 | 11.7 |
| Beed | Beed | Pali. | 0.58 | 0.43 | 1.22 | 0.73 | 0.94 | -26.1 |
| Beed | Beed | Pendgaon. | 0.85 | 0.43 | 1.45 | 0.66 | 1.04 | 22.0 |
| Beed | Beed | Pimpalner | 0.83 | 0.43 | 1.06 | 0.79 | 0.87 | 4.2 |
| Beed | Beed | Rajuri (N) | 0.60 | 0.43 | 1.21 | 0.81 | 0.96 | -34.2 |
| Beed | Dharur | Dharur | 0.77 | 0.43 | 1.14 | 0.65 | 0.88 | 15.5 |
| Beed | Dharur | Mohkhed | 0.80 | 0.42 | 1.06 | 0.80 | 0.88 | -0.2 |
| Beed | Dharur | Telgaon | 0.73 | 0.41 | 1.40 | 0.77 | 1.05 | -5.3 |
| Beed | Georai | Chaklamba | 0.74 | 0.39 | 1.35 | 0.68 | 0.99 | 8.3 |
| Beed | Georai | Dhondrai | 0.59 | 0.65 | 1.44 | 0.72 | 1.09 | -21.1 |
| Beed | Georai | Georai | 0.60 | 0.65 | 1.26 | 0.81 | 1.02 | -34.5 |
| Beed | Georai | Jategaon | 0.49 | 0.78 | 1.44 | 0.78 | 1.13 | -60.6 |
| Beed | Georai | Madalmohi | 1.00 | 0.39 | 1.18 | 0.57 | 0.87 | 42.8 |
| Beed | Georai | Pachegaon | 0.94 | 0.39 | 1.54 | 0.75 | 1.11 | 19.9 |
| Beed | Georai | Revki | 0.53 | 0.45 | 1.31 | 0.63 | 0.97 | -19.4 |
| Beed | Georai | Sirasdevi | 0.92 | 0.39 | 1.43 | 0.64 | 1.02 | 30.7 |
| Beed | Georai | Talwada | 0.67 | 0.45 | 1.22 | 0.73 | 0.94 | -7.7 |
| Beed | Georai | Umapur | 0.76 | 0.45 | 1.45 | 0.77 | 1.08 | -1.4 |
| Beed | Kaij | Bansarola | 0.65 | 0.13 | 1.31 | 0.65 | 0.92 | 0.6 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Beed | Kaij | Hanumant Pimpri | 0.81 | 0.12 | 1.21 | 0.65 | 0.86 | 20.1 |
| Beed | Kaij | Hoal | 0.99 | 0.13 | 1.44 | 0.65 | 0.99 | 34.6 |
| Beed | Kaij | Kaij | 0.96 | 0.13 | 1.31 | 0.67 | 0.92 | 29.9 |
| Beed | Kaij | Nadurghat | 0.86 | 0.34 | 1.06 | 0.65 | 0.82 | 24.2 |
| Beed | Kaij | Wida | 0.97 | 0.43 | 1.34 | 0.82 | 1.03 | 16.0 |
| Beed | Kaij | Yusufwadgao | 0.97 | 0.13 | 1.37 | 0.78 | 0.99 | 20.1 |
| Beed | Manjlegaon | Dindrud | 0.58 | 0.38 | 1.34 | 0.81 | 1.02 | -39.7 |
| Beed | Manjlegaon | Gangamasla | 0.92 | 0.38 | 1.47 | 0.65 | 1.05 | 29.9 |
| Beed | Manjlegaon | Kitti Adgaon | 1.01 | 0.58 | 1.35 | 0.65 | 1.01 | 35.9 |
| Beed | Manjlegaon | Majalgaon | 0.80 | 0.58 | 1.44 | 0.83 | 1.12 | -3.7 |
| Beed | Manjlegaon | Nithrud | 0.52 | 0.58 | 1.37 | 0.89 | 1.09 | -70.8 |
| Beed | Manjlegaon | Talkhed | 0.66 | 0.58 | 1.31 | 0.79 | 1.03 | -21.0 |
| Beed | Parli | Dharmapuri | 0.46 | 0.83 | 1.47 | 0.78 | 1.16 | -70.8 |
| Beed | Parli | Nagapur | 0.35 | 0.59 | 1.41 | 0.81 | 1.09 | -129.2 |
| Beed | Parli | Parli | 0.56 | 0.46 | 1.53 | 0.74 | 1.11 | -31.7 |
| Beed | Parli | Pimpalgaon | 0.34 | 0.43 | 1.09 | 0.77 | 0.89 | -126.7 |
| Beed | Parli | Sirsala | 0.63 | 0.58 | 1.22 | 0.65 | 0.95 | -4.7 |
| Beed | Patoda | Amalner | 0.63 | 0.49 | 1.43 | 0.79 | 1.09 | -24.1 |
| Beed | Patoda | Daskhed | 0.73 | 0.43 | 1.47 | 0.57 | 1.03 | 20.9 |
| Beed | Patoda | Patoda | 0.56 | 0.42 | 1.45 | 0.64 | 1.04 | -14.1 |
| Beed | Patoda | Therla | 0.64 | 0.43 | 1.40 | 0.79 | 1.06 | -24.1 |
| Beed | Shirur (Kasar) | Raimoha | 0.49 | 0.43 | 1.45 | 0.66 | 1.04 | -34.9 |
| Beed | Shirur (Kasar) | Shirur (Kasar) | 0.68 | 0.39 | 1.32 | 0.80 | 1.01 | -16.8 |
| Beed | Shirur (Kasar) | Tintarwani | 0.51 | 0.39 | 1.31 | 0.81 | 1.01 | -56.6 |
| Beed | Wadwani | Kawadgaon Bu. | 0.68 | 0.37 | 1.44 | 0.65 | 1.03 | 4.0 |
| Beed | Wadwani | Wadwani | 0.38 | 0.37 | 1.34 | 0.75 | 1.00 | -101.0 |

In Table 4, the yield estimated by various methods and Actual field CCE yield is presented. The percentage error of yield by the Machine learning model with field CCE, which is provided in the last column. Out of 62 points only 19 points were showing more than ±30% error. Average RMSE was -1. As per mentioned in deliverables in YESTECH manual given by Pradhan Mantri Fasal Bima Yojana, the error (nRMSE) between the observed and modeled yield should not be more than ±30% For district level. Which indicates that the process adopted for RC wise yield estimation for cotton crop is acceptable in Beed district.

**Conclusion:** This study investigated the effectiveness of various models for predicting cotton crop yields in Beed, Maharashtra, for the kharif season of 2023. The research compared the performance of three models: the Potential Production (NPP) model, the Decision Support System for Agrotechnology Transfer (DSSAT) model, and a Machine Learning model.

The evaluation revealed distinct strengths and limitations in each individual model. While each captured specific aspects of crop growth dynamics, the Machine Learning model demonstrated superior adaptability and predictive accuracy.

To overcome the limitations of individual models and enhance prediction reliability, the study explored an ensemble approach. This approach combined the strengths of all three models, creating a holistic framework that leverages their individual capabilities.

The ensemble model yielded promising results, demonstrating a close alignment with field data. This highlights the potential of such ensemble models to significantly improve the accuracy of crop yield predictions. By minimizing uncertainties associated with individual models, the combined approach provides a more reliable foundation for informed decision-making in the agricultural sector.

In conclusion, this study presents a compelling case for the integration of NPP, DSSAT, and Machine Learning models into an ensemble framework for crop yield prediction. This approach offers a promising avenue for advancing prediction methodologies and ultimately empowers farmers and policymakers with valuable insights to support sustainable agricultural practices in Maharashtra. The findings serve as a foundation for further research and refinement, aiming to continuously improve the accuracy and actionable nature of these predictions.

## References:

Ayan Das, Mukesh Kumar, Amit Kushwaha, Rucha Dave, Kailash Kamaji Dakhore, Karshan Chaudhari, & Bimal Kumar Bhattacharya. (2023). Machine learning model ensemble for predicting sugarcane yield through synergy of optical and SAR remote sensing. Remote Sensing Applications: Society and Environment, Volume 30, April 2023, 100962. https://doi.org/10.1016/j.rsase.2023.100962

Bhosale, A. D., Waskar, D. P., & Shinde, P. B. (2015). Performance of DSSAT model for simulating soybean yield under rainfed condition in Vertisols of central Maharashtra. International Journal of Agriculture, Environment and Biotechnology, 8(3), 604-610.

Deshmukh, S. D., Waskar, D. P., & Shinde, P. B. (2013). Application of DSSAT model for soybean yield prediction in Vertisols of western Maharashtra. International Journal of Current Microbiology and Applied Sciences, 2(8), 555-562.

Hoogenboom, G., C.H. Porter, K.J. Boote, V. Shelia, P.W. Wilkens, U. Singh, J.W. White, S. Asseng, J.I. Lizaso, L.P. Moreno, W. Pavan, R. Ogoshi, L.A. Hunt, G.Y. Tsuji, and J.W. Jones. (2019). The DSSAT crop modeling ecosystem. In: p.173-216 [K.J. Boote, editor] Advances in Crop Modeling for a Sustainable Agriculture. Burleigh Dodds Science Publishing, Cambridge, United Kingdom. https://dx.doi.org/10.19103/AS.2019.0061.10

Hoogenboom, G., C.H. Porter, V. Shelia, K.J. Boote, U. Singh, W. Pavan, F.A.A. Oliveira, L.P. Moreno-Cadena, T.B. Ferreira, J.W. White, J.I. Lizaso, D.N.L. Pequeno, B.A. Kimball, P.D. Alderman, K.R. Thorp, S.V. Cuadra, M.S. Vianna, F.J. Villalobos, W.D. Batchelor, S. Asseng, M.R. Jones, A. Hopf, H.B. Dias, L.A. Hunt, and J.W. Jones. (2024). Decision Support System for Agrotechnology Transfer (DSSAT) Version 4.8.2 (www.DSSAT.net). DSSAT Foundation, Gainesville, Florida, USA.

Jadhav, S. D., Waskar, D. P., & Shinde, P. B. (2018). Evaluation of DSSAT model for soybean yield prediction under different sowing dates and irrigation levels in Vidarbha region of Maharashtra. Journal of AgriSearch, 6(4), 37-42.

Jones, J.W., G. Hoogenboom, C.H. Porter, K.J. Boote, W.D. Batchelor, L.A. Hunt, P.W. Wilkens, U. Singh, A.J. Gijsman, and J.T. Ritchie. (2003). The DSSAT cropping system model. European Journal of Agronomy, 18, 235-265. https://doi.org/10.1016/S1161-0301(02)00107-7

Liujun Xiao, Guocheng Wang, Hangxin Zhou, Xiao Jin, & Zhongkui Luo. (2022). Coupling agricultural system models with machine learning to facilitate regional predictions of management practices and crop production. Environ. Res. Lett., 17(2022), 114027. https://doi.org/10.1088/1748-9326/ac9c71

Md Didarul Islam, Liping Di, Faisal Mueen Qamer, Sravan Shrestha, Liying Guo, Li Lin, Timothy J. Mayer, & Aparna R. Phalke. (2023). Rapid Rice Yield Estimation Using Integrated Remote Sensing and Meteorological Data and Machine Learning. Remote Sensing, 15(9), 2374. Remote Sensing | Free Full-Text | Rapid Rice Yield Estimation Using Integrated Remote Sensing and Meteorological Data and Machine Learning (mdpi.com)

Monteith, J. L. (1982). Solar radiation and productivity in tropical ecosystems. Journal of Applied Ecology, 19(3), 657-666.

Upasana Singh, Gargi Gaydhane and Ashutosh Pawar (2023) A Semi - Physical Approach using Remote Sensing based Net Primary Productivity (NPP), Spatial, Spectral & Temporal Paddy Yield Model Development for the State of Assam. International Journal of Science and Research (IJSR). 12 (8),1175-1785.

VNMKV Diary 2023 Krishi Dainandini 2023 published by Vasantrao Naik Marathwada Krishi Vidyapeeth, Parbhani during 2023.

Xiao, X., Boles, S., Frolking, S., Li, C., Babu, J. Y., Salas, W., & Moore III, B. (2006). Mapping paddy rice agriculture in South and Southeast Asia using multi-temporal MODIS images. Remote Sensing of Environment, 100(1), 95-113.

Yao, Y., Li, Z., Tian, F., & Tao, F. (2021). Remote Sensing-Based Estimation of Maize Yield Using a Semi-Physical Approach: A Case Study in the North China Plain. Frontiers in Plant Science, 12, 662.

Prasad, N.R., Patel N. R., Abhishek Danodia, Manjunath K. R. (2022) Comparative performance of semi-empirical based remote sensing and crop simulation model for cotton yield prediction Modeling Earth Systems and Environment 8(6) (10.1007/s40808-021-01180-x)