MACHINE LEARNING FOR PREDICTION OF CROP YIELD

A CASE STUDY ON RAINFED AREA RICE YIELD FROM WEST BENGAL

Pabitra Mitra[1], Abhijit Mukherjee[2] and Aditi Chandra[1]

1 - Department of Computer Science and Engineering,
2 - Department of Geology and Geophysics
Indian Institute of Technology, Kharagpur-721302,
West Bengal, India

Contents

CONTEXT

Back to Content

- India is growing fast in population. The demand is high and will increase in coming future.
- Hence, to ensure food security vertical development in agriculture is the need of the hour.
- By area about **75 million ha is rainfed and totally** depends on rains facing the vagaries of monsoon.
- For this a combined structural and methodological approach like variety inception, pesticide & fertilizer management, integrated cropping, rainwater harvesting, efficient irrigation techniques etc. would be required.
- Also, it becomes essential to simulate & predict the crop yield under ambient conditions prior to the implementation stage for effective crop management & desired results, more so in the rainfed area and when India is trending towards precision farming practices
- Since the relations between crop yield and the weather & non-weather factors are non-linear & include decent level challenges, machine learning might prove a successful alternative for yield predictions

Back to Content

Context

Crop yield prediction methods

Statistical Methods

Crop Simulation Methods

Machine Learning

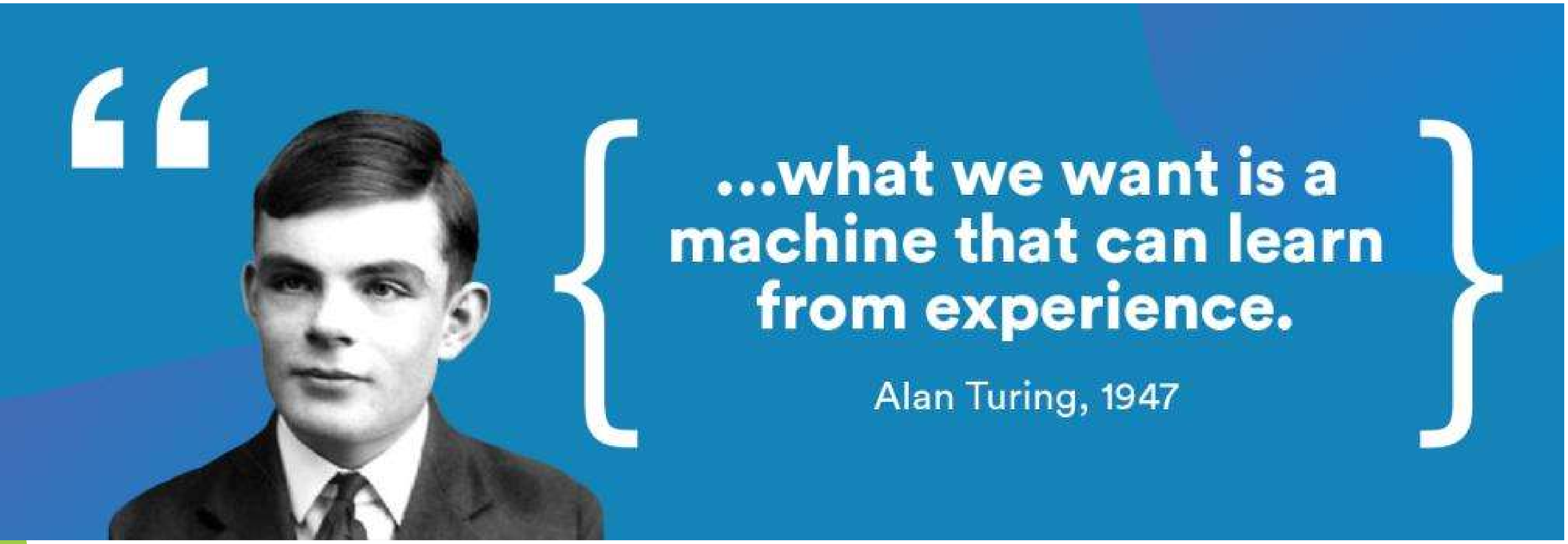
The crop yield prediction methods are – Statistical, Crop Simulation and Machine Learning. Out of these, the Statistical and Crop Simulation models are quite popular.

The machine learning presents several methods to define rules and patterns in large data sets related to crop yield and has well known predicting capability. In addition it can self improvise the predictive model.

Unfortunately, so far the machine learning methods have not been applied on a large scale in the country, primarily because we are used to the statistical and crop simulation methods

We present here our initial approach through a simple case study on kharif rice yield prediction in rainfed area using one of the machine learning methods - the Artificial Neural Network

Back to Content



Back to Content

OBJECTIVES

- To assess the scope and suitability of machine learning methods for predictions of crop yield in rainfed area (using weather variables),
- To develop initial machine learning models for crop yield predictions in rainfed area
- To analyze the performance of the model developed
- To explore the spatial & temporal granularity, and constraints therein in the database

MACHINE LEARNING METHOD–ARTIFICIAL NEURAL NETWORKS

Artificial neural network (ANN) - one of many computing models of artificial intelligence

ANN can be used for classification and prediction

ANN requires no explicit mathematical equation and assumptions

Advantage of ANN over traditional crop simulation models - most of the intense computations take place during training process.

Once ANN is trained its operation is fast

ANN model constitutes interconnected artificial neurons exchanging actuation signals in the form of activation transition function.

Weights in ANN are generated such that the outputs depend on the actual inputs and the internal state of the network

The multi-layer ANN is often applied in prediction tasks.

Important step in ANN model - selection of network type (its topology), along with the selection of a suitable activation function and a teaching algorithm

For the present study we have constructed a multi-layer neural perceptron (MLP) along with non-linear activation function

For the learning backpropagation algorithm iterative method is implemented.

Backpropagation algorithm attempts to find the minimum of the error function in the weight space using methods based on the gradient descent.

Back to Content

STATUS OF MACHINE LEARNING TECHNIQUES IN CROP YIELD PREDICTIONS

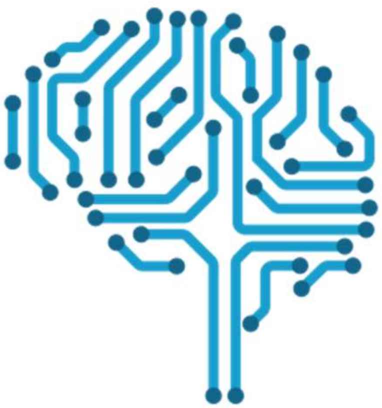
Some of the references related to crop yield where machine learning techniques have been used are

1. McQueen, R.J., et al	Applying machine learning to agricultural data	1995
2.Shearer, S.A., et al	Yield prediction using neural network classifier trained using soil landscape features & soil fertility data	1999
3. Liu, J., et al	A neural network for setting target corn yield	2001
4. Drummond, S. T., Joshi, A.	Predictive ability of neural networks for site-specific yield estimation	2000
5. O'Neal, M. R., et al	Neural network prediction of maize yield using alternative data coding algorithms	2002
6. Drummond, S. T., et al	Statistical and neural methods for site-specific yield prediction.	2003
7. Puteh, S., et al	Backpropagation algorithm for rice yield prediction	2004
8. Kaul,M., et al	Artificial neural network for corn and soybean prediction	2005
9. Irmak, A., et al	Artificial neural network model as a data analysis tool in precision farming	2006
10. Ji, B., et al	Artificial neural networks for rice yield prediction in mountainous regions	2007
11. Li, A., et al	Estimating crop yield from multi-temporal satellite data using multivariate regression & nn techniques	2007
12. Armstrong, L.J ., et al	The application of data mining techniques to characterize agricultural soil profiles	2007
13. Armstrong, L., et al	Data mining can empower growers' crop decision making	2007
14. Diepeveen, D, et al	Identifying key crop performance traits using data mining	2008
15. Vagh, Y. Armstrong, L.	Application of a data mining framework for the identification of agricultural production areas in WA	2010
16. Newlands, N.K, Townley-S. L.	Predicting Energy Crop Yield Using Bayesian Networks	2010
17. Baral, S., et al	Yield Prediction Using Artificial Neural Networks	2011
18. Stastny, J., et al	Agricultural data prediction by means of neural network	2011
19. Gonzalez-Sanchez, A., et al	Predictive ability of machine learning methods for massive crop yield prediction	2014
20. Veenadhari, S., et al	Data mining techniques for predicting crop productivity - A review article	2011
21. Guo, W., Xue, H.	An incorporative statistic and neural approach for crop yield modelling and forecasting	2012
22. Raorane, A., Kulkarni, R.	An effective tool for yield estimation in the agricultural sector	2012
23. Ramesh, D., Vardhan, B.	Data mining techniques and applications to agricultural yield data	2013
24. Ranjeet, T., Armstrong, L.	An artificial neural network for predicting crops yield in Nepal	2014

Back to Content

25. Jabjone, S., Wannasang, S	Decision Support System using Artificial Neural Network to predict rice production in Phimai district, Thailand	2014
26. Khairunniza-Bejo, S., et al	Application of artificial neural network in predicting crop yield: a review	2014
27. Hota, S. K.,	Artificial neural network and efficiency estimation in rice yield	2014
28. Babatunde, O.H., et al	On The Application Of Genetic Probabilistic Neural Networks and Cellular Neural Networks In Precision Agriculture	2014
29. Leng, J., et al	A network that really works- the application of ann to improve yield predictions ..in Western Australia	2014
30. Mobarake, S. A., et al	A Model for the Estimation of Yield and Investigation on Factors Affecting Irrigated Wheat Production …	2014
31. Guo, W.W., Xue, H.	Crop Yield Forecasting Using Artificial Neural Networks: A Comparison between Spatial and Temporal Models	2014
32. Medar, R. , Rajpurohit, V.	A survey on data mining techniques for crop yield prediction	2014
33. Dahikar, S., Rode, S.	Agricultural Crop Yield Prediction Using Artificial Neural Network Approach	2014
34. Mahabadi, N. Y.	Use of intelligent models to predict the rice potential production	2015
35. Dahikar, S.S., et al	An Artificial Neural Network Approach for Agricultural Crop Yield Prediction Based on Various Parameters	2015
36. Kuwata, K.	Estimating crop yields with deep learning and remotely sensed data	2015
37.Ramesh, D. Vardhan, B.	Analysis of crop yield prediction using data mining techniques	2015
38. Ahamed, A., et al.	Applying data mining techniques to predict annual yield of major crops.. in Bangladesh	2015
39. Gandhi, N., et al	Rice crop yield prediction in India using support vector machines	2016
40. Gandhi, N., et al	Predicting Rice crop yield using Bayesian networks	2016
41. Gandhi, N, Armstrong, L.J.	A review of the application of data mining techniques for decision making in agriculture	2016
42 Gandhi, N, Armstrong, L.	Applying data mining techniques to predict yield of rice in humid subtropical climatic zone of India	2016
43. Gandhi, et al	Rice crop yield prediction using artificial neural networks	2016
44. Gandhi, N. Armstrong, L. J.	Rice crop yield forecasting of tropical wet and dry climatic zone of India using data mining techniques	2016
45. Jeong. J.H., et al	Random Forests for global and regional crop yield predictions	2016
46. Gandhi, N., Armstrong, L.	Assessing impact of seasonal rainfall on rice crop yield of Rajasthan, India using Association Rule Mining	2016
47. Gandhi, N., Armstrong, L.	Assessing the Impact of Seasonal Rainfall Patterns on Indian Rice Production using Association Rule Mining	2016
48. You, J., et al	Deep Gaussian process for crop yield prediction based on remote sensing data	2017
49.Patil, D., Shirdhonkar, M.S.	Rice Crop Yield Prediction using Data Mining Techniques: An Overview	2017

Back to Content



Out of the 49 world-wide studies, 17 are related to India, of which 8 have been conducted under the guidance of Prof. Dr. Armstrong

Back to Content

STUDY AREA

West Bengal – Purulia Dist

Study is initiated for a rainfed area kharif rice cropping with no irrigation support – two blocks – Jhalda II & Banduan of Purulia district, West Bengal

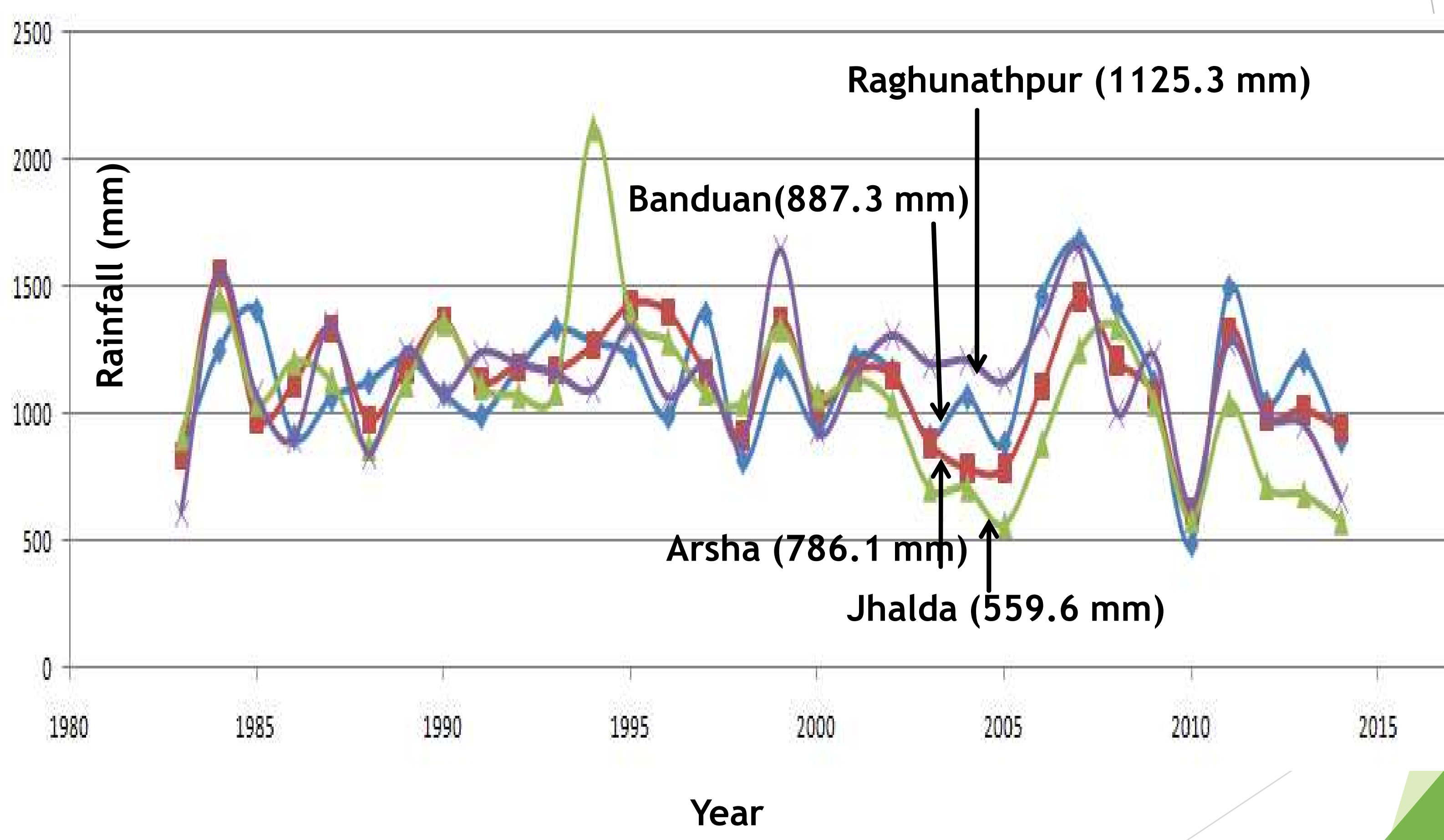
About Purulia District

Back to Content

- Purulia district's total area is 6259 sq km
  - The district is in granite gneiss terrain of Chhotanagpur plateau
  - It has 20 administrative/development blocks
  - Its elevation varies from 312 m to 129 m amsl
  - Its western part is upland and hilly terrain
  - Its eastern and southern parts are lowland
  - The district area is totally rainfed and monocropped
  - Its main crop is kharif rice
  - The district falls under medium to low crop yield category (1500-2000 kg/ha)

- The rainfall data of the district processed at 0.25° X 0.25 °grid representing the blocks indicate a considerable spatial and temporal variations.
- Therefore, the two blocks viz., Jhalda II and Banduan located in the NW upland and SE lowland respectively have been considered for yield prediction through Machine Learning Method.

Variations in Monsoon Total Rainfall in the Year 2005 in 4 blocks of Purulia District, West Bengal



THE VARIABLES & FRAMEWORK

Back to Content

INPUT VARIABLES

OUTPUT VARIABLE

WEATHER

- **Rainfall** (Daily, Weekly, Monthly, Crop Growth Phases)
- **Temperature** (Minimum & Maximum)
- **Solar Radiation**
- **Evapotranspiration**

NON-WEATHER

- **Seed Variety**
- **Soil Type**
- **Fertilizer**
- **Pesticides**
- **Farm Equipment**
- **Labour**

Crop Yield

For the present study we have considered the weather variables only, because it is quite difficult to collect the non-weather data which have inherent irregularity.

However, we aspire to integrate the non-weather variables in future in the model

Back to Content

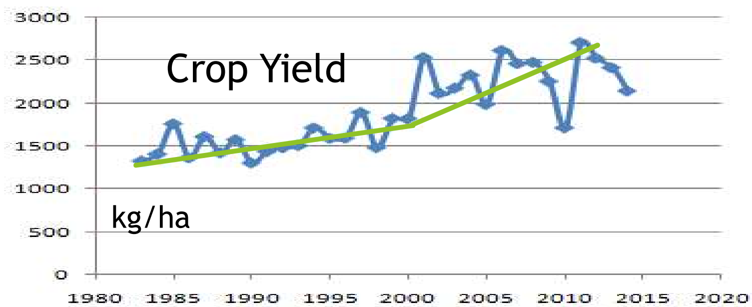
Historical Data Sources

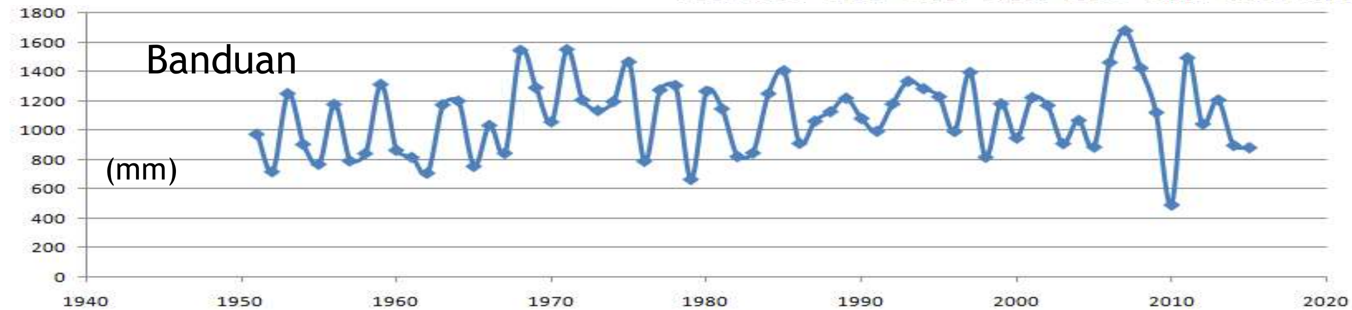
Parameter		Parameter Density	Parameter Frequency	Period	Parameter description	Source
Input variables	Rainfall (mm)	0.25 deg. grid	Daily	1990-2014	Weekly & monthly total, June-Oct	India Met. Dept
	Max.Temp.(°C)	0.50 deg. grid			Monthly average June-Nov	India Met. Dept
	Min.Temp. (°C)					
	Solar radiation (MJ/m^2/day)	1.0 deg. grid			Monthly average June-Nov	NASA Power
Output variable	Kharif rice yield (kg/ha)	District level	Annual			IMD (Agro-Met) &https://data.gov.in/

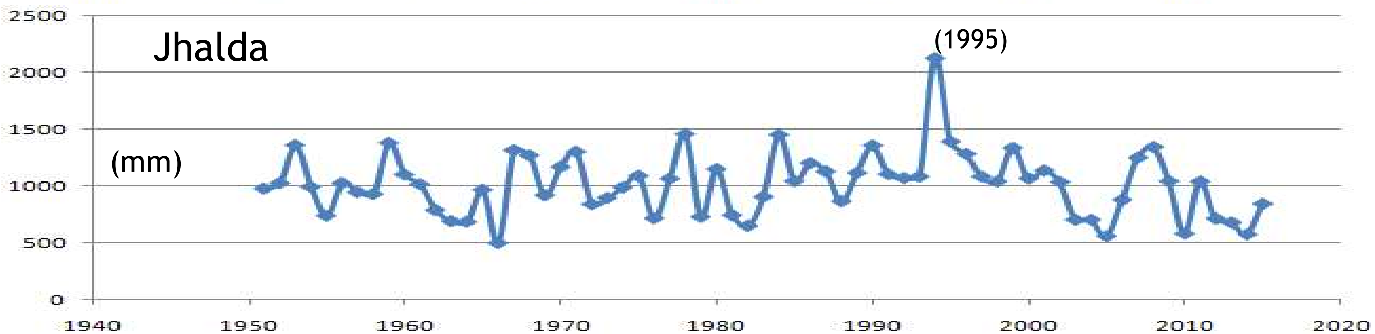
WEATHER VARIABLES & CROP YIELD

Back to Content

Monsoon (June-September) Total Rainfall at Banduan and Jhalda in Purulia District and Kharif Rice Yield of Purulia District

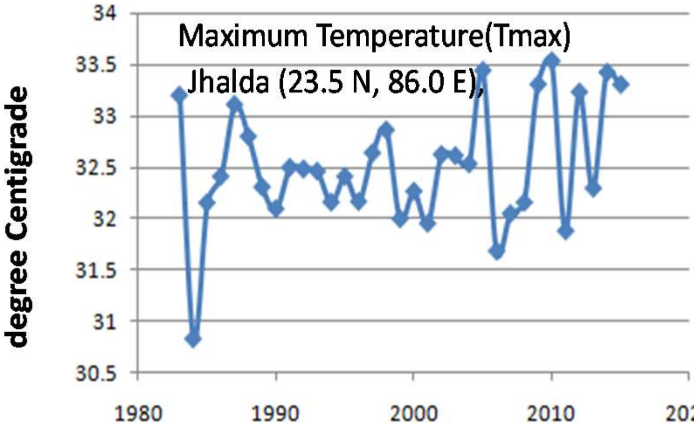


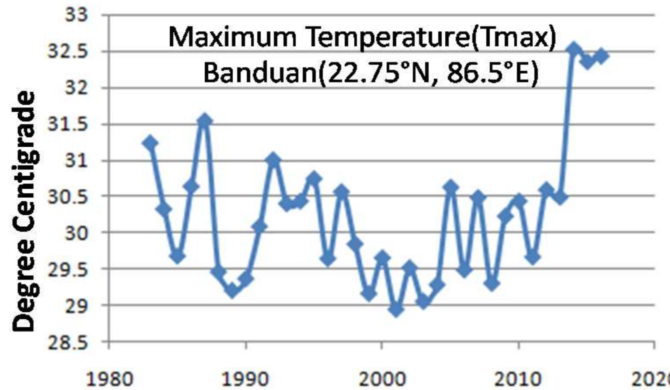


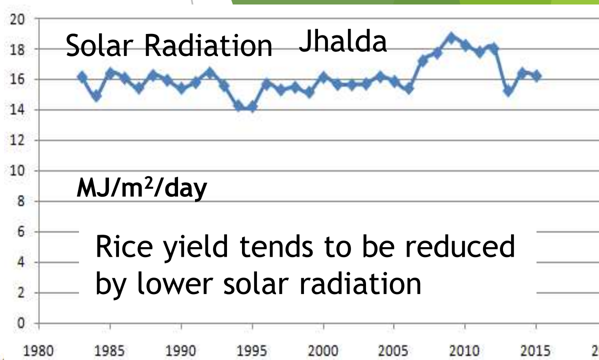


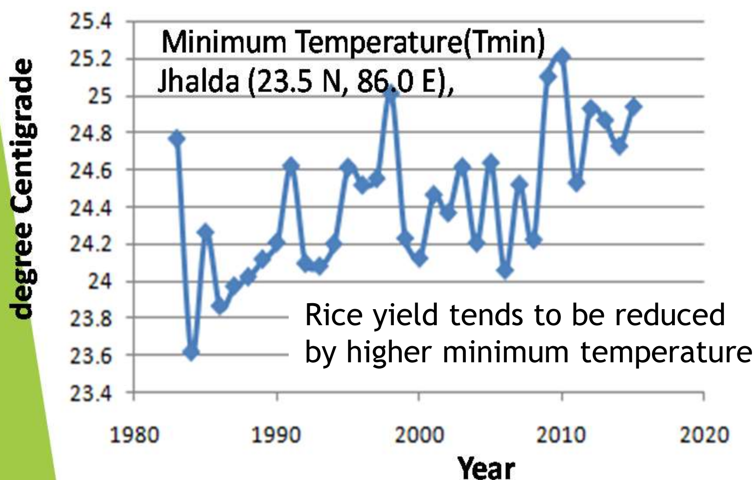
Back to Content

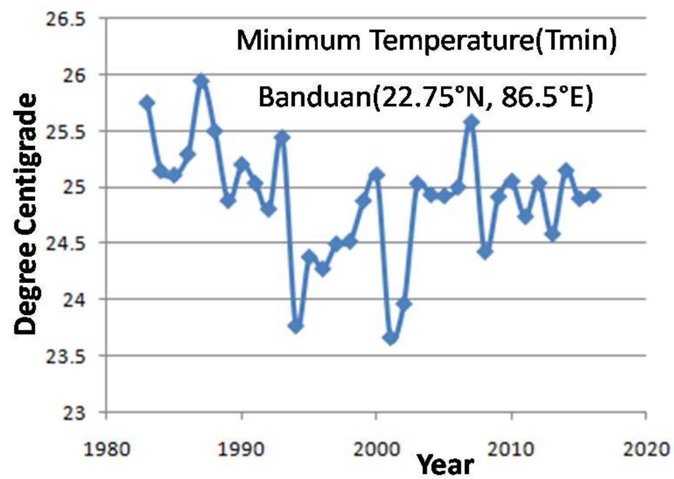
Average maximum (Tmax) & minimum (Tmin) temperatures and solar radiation during monsoon (June-September) at Jhalda & Banduan in Purulia district, West Bengal











METHODOLOGY

Feature Selection

Back to Content

- Data Period - Modern variety rice from 1990s, therefore the period 1990-2014 is considered
- Rainfall data - Availability of daily rainfall data at 0.25°X 0.25° grid interval.
- Spatial Data- Out of 20 blocks 10 blocks are represented by the grids
- Location - Jhalda II, in the Northwest upland, and Banduan in the southern lowland, are selected
- Time granularity- Daily rainfall to monthly rainfall conversion – June to October for 1990 to 2014
- Time Interval - Weekly rainfall computed for 22 weeks (June to October)
- Rainy Day- Total rainy days (2.5 mm or above rainfall)
- Dry Spell - A dry spell -no rain for 4 or more than 4 consecutive days.
- Dry Week- The week with 4 consecutive dry days and less than 17.5 mm rainfall in total as dry week
- Model Input- The months with and without dry spells are defined for the model input by binary numbers
- Temperature- The daily maximum and minimum temperatures are converted to monthly mean
- Solar Radiation- The satellite data based daily solar radiation obtained from the website NASA

Combination of Input Variables

The selection of input variables and their combinations is an important and complex task. The following combinations were initially chosen expecting the combinations will give an insight into variable selections for yield predictions.

COMBINATIONS OF INPUT DATA SETS					
A.	**Weekly rainfall, June-Oct**				
B.	**Weekly rainfall, June-Oct**		**Monthly average Tmax, Tmin June-Nov**		
C.	**Weekly rainfall, June-Oct**		**Monthly average Tmax, Tmin June-Nov**	**Average solar rad. June-Nov**	
D	**Weekly rainfall, June-Oct**	**Monthly average Tmax, Tmin June-Nov**	**Monthly average solar rad. Sept-Nov**	**No. of rainy days June-Sept**	**Dry spells July-Sept**
E	**Weekly rainfall, June-Oct**	**Monthly average Tmax, Tmin June-Nov**	**Monthly average solar rad. Sept-Nov**	**No. of rainy days June-Sept**	
F	**Weekly rainfall, June-Oct**	**Monthly average Tmax, Tmin June-Nov**	**Monthly average solar rad. Sept-Nov**	**Dry spells July-Sept**	
G	**Weekly rainfall, June-Oct**	**Monthly average Tmax, July- Sept**	**Monthly average Tmin, Oct-Nov**	**Monthly average solar rad. June-Nov**	
H	**Weekly rainfall, June-Oct**	**Monthly average Tmax, July- Sept**	**Monthly average Tmin, Oct-Nov**	**Monthly average solar rad. Oct-Nov**	

Total rainfall June-Oct	Dry spells in the 3 phases of growth	Heavy rainfall in the 3 phases of growth	Monsoon Season average Tmax	Monsoon Season average Tmin	Monsoon season average solar radiation

Back to Content

Methodology

Artificial Neural Network

- Twenty-five years' data considered
- Eighty percent for training and 20% for testing
- R language(open source) used
- R neuralnet package for model generation
- A three layer model generated
- Seven hidden nodes
- Correlation and mean square error calculated
- Validation done using 20% dataset
- Models generated separately for the 2 locations

Back to Content

RESULTS & MODEL PERFORMANCES

Combinations of Input data sets						Location	Correlation (max 1.0)	MSE
A.	Weekly rainfall, June-Oct					Jhalda	0.74	0.20
						Banduan	0.83	0.12
B.	Wkly rainfall, June-Oct	Monthly average Tmax, Tmin June-Nov				Jhalda	0.81	0.14
						Banduan	0.85	0.54
C.	Wkly rainfall, June-Oct	Monthly average Tmax, Tmin June-Nov		s		Jhalda	0.79	0.06
						Banduan	0.85	0.22
D	Wkly rainfall, June-Oct	Monthly average Tmax, Tmin June-Nov	Monthly average solar rad. Sept-Nov	No. of rainy days June-Sept	Dry spells July-Sept	Jhalda	0.82	0,06
						Banduan	0.91	0.23
E	Wkly rainfall, June-Oct	Monthly average Tmax, Tmin June-Nov	Monthly average solar rad. Sept-Nov	No. of rainy days June-Sept		Jhalda	0.78	0.21
						Banduan	0.83	0.27
F	Wkly rainfall, June-Oct	Monthly average Tmax, Tmin June-Nov	Monthly average solar rad. Sept-Nov	Dry spells July-Sept		Jhalda	0.81	0.04
						Banduan	0.89	0.55
G	Wkly rainfall, June-Oct	Monthly average Tmax, July- Sept	Monthly average Tmin, Oct-Nov	Monthly average solar rad. June-Nov		Jhalda	0.79	0.15
						Banduan	0.89	0.49
H	Wkly rainfall, June-Oct	Monthly average Tmax, July- Sept	Monthly average Tmin, Oct-Nov	Monthly average solar rad. Oct-Nov		Jhalda	0.73	0.10
						Banduan	0.87	0.27

Performance Analysis

Five-fold Cross-Validation of the Neural Network Model E

Model E subgroup	Training Data	Test Data	Location	Corre-lation	MSE
1	1990-2009 (initial model)	2010-2014	Jhalda	0.74	0.05
			Banduan	0.82	0.22
2	20-year data other than test data	1990-1994	Jhalda	0.89	0.25
			Banduan	0.85	0.19
3	20-year data other than test data	1995-1999	Jhalda	0.30	0.20
			Banduan	0.67	0.08
4	20-year data other than test data	2000-2004	Jhalda	0.60	0.12
			Banduan	0.90	0.01
5	20-year data other than test data	2005-2009	Jhalda	0.93	0.09
			Banduan	0.64	0.03

Back to Content

Comparison of Yield Prediction

Existing Statistical & Crop Simulation Combined Models vs ANN Models

Under FASAL Scheme of Govt. of India, predicted the kharif rice yield for the year 2015 for Purulia district at F2 stage (pre-flowering) through statistical and crop simulation model is 2468 kg/ha with R^2 0.85 (Singh,2015)*

Through our ANN models predicted kharif crop yield for Banduan for the year 2015 is

2390 kg/ha – Model D
2302 kg/ha – Model E
2270 kg/ha – Model F

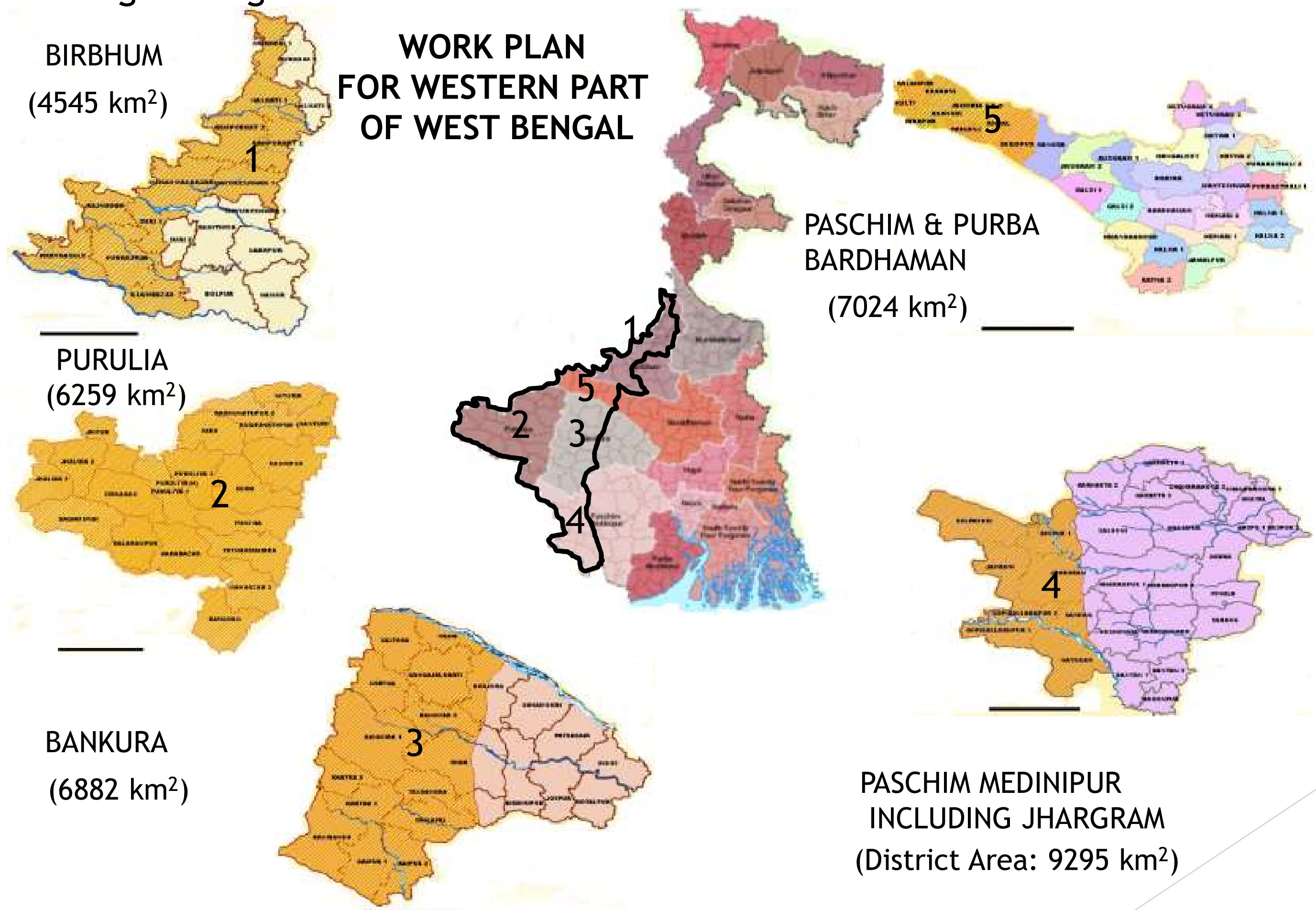
** The actual value is 2352 kg/ha

**Crop Yield Forecasting using Agromet Model: Indian experience by K K Singh (2015), India Meteorological Department, New Delhi, India,kksingh2022@gmail.com*

Back to Content

WORK PLAN

The immediate work plan is to develop prediction models for all the rainfed –drought prone blocks of the districts (approx. 17,000 sq km) : 1.Birbhum, 2.Purulia,3. Bankura, 4.West Medinipur and 5. West Bardhaman of West Bengal using block level weather and non-weather data and block level kharif rice yield



CONCLUSIONS

The study shows that rice yield is coupled to weather variables, some are strongly while some are weakly coupled

We are trying to establish the best combination of input weather variables

Within a district weather varies very widely

Yield predictions at block level with block level is to be attempted

ANN models have been developed for the 2 locations within a district

Three-layer ANN model is adequate for the purpose

Correlation ranges between 0.73 and 0.91 with MSE between 0.04 and 0.55

Back to Content

WAY FORWARD

- Define area and problem specific machine learning approach and further consolidate these into a robust machine learning model
- Improvise models with different components for the target area –rainfed hard rock part (17000 sq km approx.) of the western districts of West Bengal,
- Attempt crop yield predictions at block level,
- Identify and evaluate different ICTs for developing the appropriate framework, and
- Design and link the outcome to the ICT framework

THANK YOU

Back to Content

Back to Content

ANN Charts & Graphs