RESEARCH ARTICLE OPEN ACCESS

Crop Yield Prediction using Temperature and Rainfall parameters prediction

S.Jagadeesan M.Sc(CS)., MCA., M.Phil(CS)., ME(CSE).,*, G.Sunil Sharma**
* Assistant Professor, Department of MCA, Nandha Engineering College (Autonomous), Erode, Tamilnadu, India.
Email: jagadeesan12398@gmail.com
** Final MCA, Department of MCA, Nandha Engineering College (Autonomous), Erode, Tamilnadu, India.
Email: vimalsathya42@gmail.com

****

Abstract:

Agriculture plays a significant role in the economy of India. This makes crop yield prediction an important task to help boost India's growth. Crops are sensitive to varied weather phenomena like temperature and rainfall. Therefore, it becomes crucial to include these features when predicting the yield of a crop. Weather forecasting is a complicated process. In this work, three methods are wont to forecast- ARMA (Auto Regressive Moving Average), SARIMA (Seasonal Auto Regressive Integrated Moving Average) and ARMAX (ARMA with exogenous variables). The performance of the three is compared and therefore the best model is employed to predict rainfall and temperature which are successively wont to predict the crop yield based on a fuzzy logic model.

Keywords — **ARMA; SARIMA; ARMAX; Temperature prediction; Rainfall prediction; Fuzzy Logic; Crop Yield Prediction .**

****

I. INTRODUCTION

Agriculture is one of the most important economic sectors in India. It plays an important role in rural development and sustainability. The level of agriculture may decrease thanks to factors like unpredicted rainfall, global climate change , use of excessive pesticides etc. The main aim of this study is to provide a methodology for crop yield production based on the historical climatic and production data. Crop yield prediction supported the previous years of temperature and rainfall can help farmers take necessary steps to enhance crop yield within the coming season. Understanding crop yield can help ensure food security and reduce impacts of global climate change . We have tried to develop a method such that the crop yield can be predicted beforehand using only temperature and rainfall of previous years. Accurate rainfall prediction may be a difficult task because rainfall depends on various features like cloudiness , evapotranspiration, and lots of other.

climatic factors but we wanted to extract useful information about crop yield using only two features i.e. temperature and rainfall. The proposed method uses Auto Regressive Moving Average and Seasonal ARIMA models to predict temperature. Since we would like to predict temperature from past values, these models suit our needs. A time-series created from the dataset is fed into the model to predict temperature. Similarly, ARMA and ARMA with exogenous variables (ARMAX) models are wont to predict rainfall. We decided to use the ARMAX model in case of rainfall so that other factors such as cloud cover, temperature and evapotranspiration can also be taken into account.

We used a fuzzy logic system to predict yield. The fuzzy model takes within the predicted values from the model with least errors and provides the yield for that season.

II. RELATED WORK

Several models have been used in the past to predict temperature and rainfall. Jesleena Rodrigues et al. have considered the ARIMA and Multiple Linear Regression to predict rainfall for all states of India. In MLR equation, parameters are taken from the dataset and variables are extracted from the dataset by correlation. ARIMA is employed for modelling statistic and rainfall prediction. [1]. S.Meenakshi Sundaram and M.Laxami have used Seasonal Auto Regressive Integrated Moving Average(SARIMA) model which includes iterative identification, estimation, analysis and forecasting stages to predict the monthly rainfall. Mean Absolute Percentage Error (MAPE) is used to calculate the accuracy of this research work [9]. Sandeep Kumar Mohapatra el al. have considered using regression. They used fixed sampling and K-fold cross validation techniques. The proposed work collected seven meteorological data parameters for predicting rainfall of Bengaluru, India Those parameters were rainfall, maximum temperature,

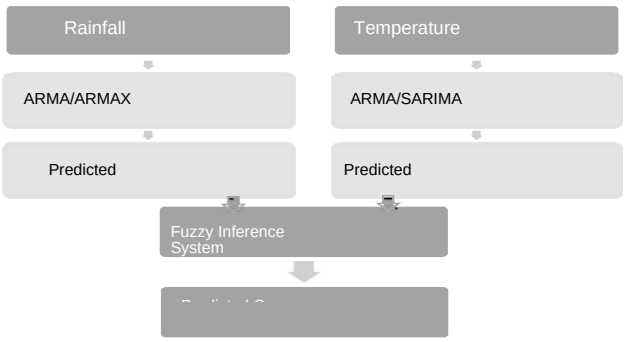
precipitation, wet day frequency, mean temperature, ratio , total cloud amount and wind speed. For making the model, they studied the collected data for different parameters and selected two features to predict the rainfall. The feature selection was supported the trend and variation of the info . [8]. S Prabakaran has considered applying modified rectilinear regression on parameters like average temperature, cloudiness to predict rainfall [7]. Inderjeet Kaushik predicted rainfall and temperature for Mirzapur using seasonal ARIMA model. Unexpected and missing values within the data reduced the accuracy of predictions [5]. Studies have also used data mining techniques to predict crop yield [6], where the information gain for each attribute is calculated to obtain a ranking of the attributes such as rainfall, potential evapotranspiration, maximum and minimum temperature, cloudiness and wet day frequency to pick the attributes. Pankaj Kumar has used Adaptive Neuro Fuzzy Inference System to develop a forecasting model supported statistic agricultural data of rice crops supported weather parameters. ANFIS utilizes information from symbolic logic also learning of an ANN [2].

III. DATASET

We acquired our data for rainfall and temperature prediction from indiawaterportal.org [4]. The database contains record of the quantity of rainfall/temperature in various districts of India of the previous 100 years. We take the rainfall and temperature data for Allahabad and Bengaluru. Each row in the dataset represented a year with the average temperature/rainfall of every month distributed in 12 columns. We restructured the data such that each row represented the average temperature/rainfall for each month of every year. This modelling allowed us to form a time series data required in the models used. This dataset is divided into training and testing set. The values for years 1900-80 are used as training set and the values from 1981-2002 are used to prepare the testing set. We also acquired the dataset for crop yield prediction from Kaggle [3] which contained the quantity of crop produced supported the Kharif and Rabi season of 18 years to study the change in crop yield in response to the change in temperature and rainfall.

IV.METHODOLOGY

The Auto Regressive Moving Average (ARMA) and Auto Regressive Moving Average with exogenous variables (ARMAX) models are applied to the remodelled rainfall dataset. Similarly, ARMA and Seasonal Auto Regressive Integrated Moving Average (SARIMA) models are applied to the temperature data to predict rainfall and temperature. Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) are calculated to verify model accuracy. These predicted values are then supplied to the fuzzy inference system to predict crop yield based on the two parameters.



A. ARMA

Model ARMA model is stands for Autoregressive Moving Average. It is a technique for modelling time series data such as temperature. Basically, it is the combination of autoregressive and moving average models. The autoregressive part of this model, in which response variable depends on its previous values can be mathematically defined as:

$$x(t) = \varphi(1)x(t-1) + \varphi(2)x(t-2) + \ldots + \varphi(p)x(t-p) + e(t)$$

where, e(t) is the error term.

Whereas the moving average part indicates that the regression error is a linear combination of error terms, mathematically-

$$x(t) = e(t) + \theta(1)e(t-1) + \ldots + \theta(q)e(t-q).$$

Combining the two, an ARMA (p, q) model can be defined as-

$$x(t) = \varphi(1)x(t-1) + \varphi(2)x(t-2) + \ldots + \varphi(p)x(t-p) + e(t) + \theta(1)e(t-1) + \ldots + \theta(q)e(t-q)$$

where p and q are the orders of the AR and MA parts respectively.

• Model Preparation

An ARMA model is chosen in this step to model the rainfall and temperature data. This model is separately applied to rainfall and temperature data.

• Prediction

The rainfall and temperature values are given to the model to predict future values of rainfall and temperature data.

B. SARIMA Model

Replacing original data values with the difference between their values and former values adds an 'Integrated' factor to the previously used ARMA model which is then called as an ARIMA model. In certain types of data, we would want the data values to depend on the last season data rather than the previous values. For example, the temperature value for the month of January in 2001 would be dependent on the temperature for January 2000 rather than December 2000. Therefore, we would want to add a 'Seasonal' component to our model. The Seasonal Autoregressive Integrated Moving Average (SARIMA) model does exactly that- it models our data to depend on the previous season values rather than previous values.

• Model Preparation

A SARIMA model is carefully chosen to be applied to the temperature data. We set the seasonal factor to 12, meaning that our data is dependent on previous year values.

• Prediction

The temperature values are given to the model for prediction.

C. ARMAX Model

In the ARMA and SARIMA models, the variable is regressed only on its own previous values which proves successful for temperature prediction but rainfall maybe dependent on various other weather phenomena like cloudiness , temperature, vapor pressure , etc. We include exogenous variables in our ARMA model to convert it into an ARMAX model.

• Exogenous Variable Selection

The independent variables are taken under consideration but only those variables are chosen that absolutely the correlation values with rainfall are greater than 0.8.

• Model Preparation

We choose the same ARMA model but include certain exogenous variables to include in forecasting.

• Prediction The rainfall values are given to the model for prediction.

V. TEMPERATURE PREDICTION

The temperature of Allahabad for the months from January 1981 to December 2002 is predicted using ARMA and SARIMA models.

A. Using ARMA model:

A few samples of predicted values using ARMA are shown in Table I. Fig I shows a graph of the actual values in blue overlapped with the predicted values using ARMA model in red so that the two can be compared.

TABLE I: Sample Predicted values of temperature for Allahabad

Month/Year	Actual Temperature	Predicted Temperature
07/2001	28.409	28.516
08/2001	28.695	27.476
09/2001	28.676	28.472
10/2001	27.045	27.852
11/2001	22.529	26.094
12/2001	17.426	21.249

B. Using SARIMA model:

Some predicted values using SARIMA are shown in Table II. Fig II shows a graph of the actual values in blue overlapped with the predicted values using SARIMA model in red so that the two can be compared visually.

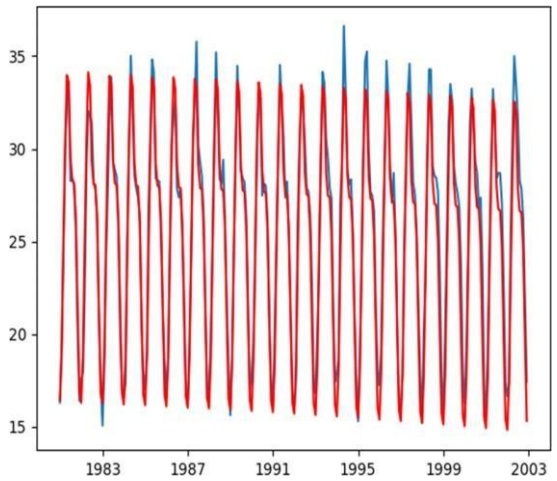
TABLE II: Predicted values of temperature for Allahabad district

Month/Year	Actual Temperature	Predicted Temperature
07/2001	28.409	27.966
08/2001	28.695	26.743
09/2001	28.676	26.653
10/2001	27.045	24.696
11/2001	22.529	19.705
12/2001	17.426	15.412

Table III shows the MAE and RMSE vales for the two models used for temperature prediction.

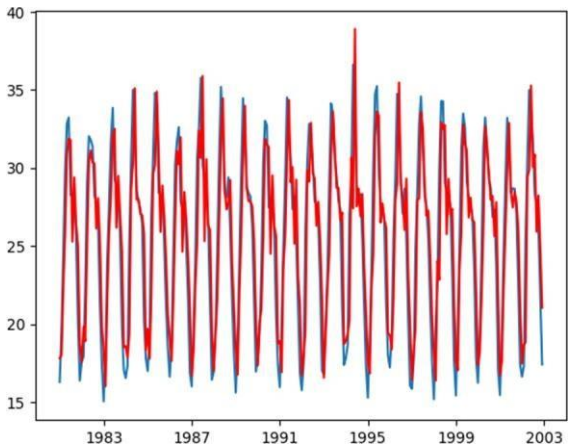
TABLE III: ERROR

	MAE	RMSE
Temperature- ARMA	8.11	1.52
Temperature- SARIMA	1.032	1.299



----- Predicted ----------Actual

Fig. 1: Allahabad Temperature Graph – ARMA



----- Predicted ----------Actual

Fig. 2: Allahabad Temperature Graph - SARIMA

VI. RAINFALL PREDICTION

The rainfall of Allahabad and Bengaluru for the months from January 1981 to December 2002 is predicted using ARMAX and ARMA models. The rainfall for each month of a year is then combined

to get the average yearly rainfall for that year which would be used in crop yield prediction.

A. Using ARMAX model:

Out of temperature, maximum temperature, minimum temperature, cloud cover and, evapotranspiration, only cloud cover and evapotranspiration have absolute correlation values > 0.8 hence,

Exogenous variables:

- Cloud Cover
- Evapotranspiration

Some predicted values of rainfall for Allahabad using ARMAX are shown in Table IV. Similarly, Table V shows some predicted values of rainfall for Bengaluru.

TABLE IV: Predicted values of rainfall for Allahabad

Year	Actual Rainfall	Predicted Rainfall
1999	89.505	101.880
2000	71.718	87.526
2001	66.848	99.183

TABLE V: Predicted values of rainfall for Bengaluru

Year	Actual Rainfall	Predicted Rainfall
1999	76.902	66.747
2000	83.397	67.634
2001	60.986	71.461

B. Using ARMA model:

Table VI gives the anticipated values of rainfall for Allahabad for a couple of years using the ARMA model and therefore the predicted values for Bengaluru are given in Table VII.

Fig III shows a graph of the actual values in blue overlapped with the predicted values of rainfall for Allahabad using ARMA model for visual comparison. Fig IV does the same for Bengaluru

TABLE VI: Predicted values of rainfall for Allahabad

Year	Actual Rainfall	Predicted Rainfall
1999	89.505	88.740
2000	71.718	79.526
2001	66.848	76.701

TABLE VII: Predicted values of rainfall for Bengaluru

Year	Actual Rainfall	Predicted Rainfall
1999	76.902	73.436
2000	83.397	74.450
2001	60.986	67.299

Table VIII shows the MAE and RMSE vales for the two models used for rainfall prediction of both Allahabad and Bengaluru

TABLE VIII: ERROR

	MAE	RMSE
Allahabad Rainfall- ARMAX	37.25	46.64
Allahabad Rainfall- ARMA	9.35	10.98
Bengaluru Rainfall- ARMAX	12.89	14.92
Bengaluru Rainfall- ARMA	8.09	9.74

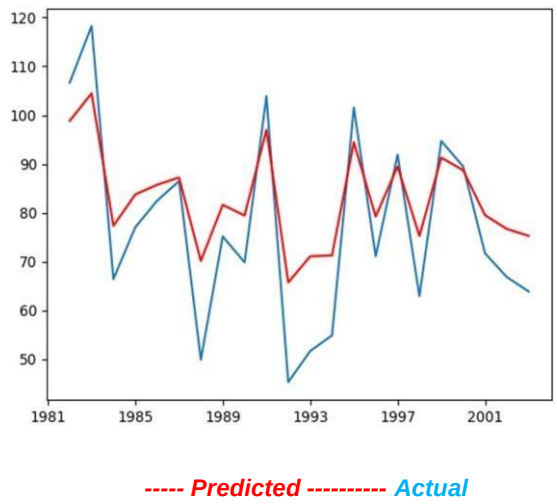


Fig. 3: Allahabad Rainfall Graph- ARMA

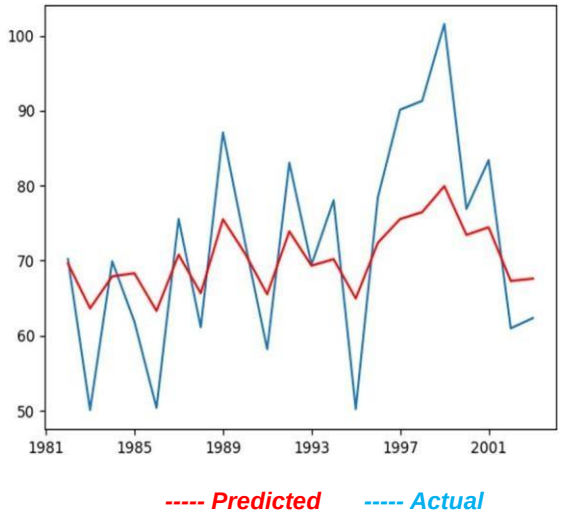


Fig. 4: Bengaluru Rainfall Graph- ARMA

VII. YIELD PREDICTION BY FUZZY LOGIC

After the prediction of Temperature and Rainfall with the help of SARIMA and ARMA models respectively, we have made some Fuzzy if- then rules to classify the yield in five categories – Very bad, Bad, Average, Good, Very Good. Rules are made by taking the average temperature and average rainfall suitable for the crop, for which the yield will be predicted. agropedia.iitk.ac.in [10] has been studied to form the principles for wheat given in Table IX. India has two major crop seasons- Kharif and Rabi. The Kharif season extends from June to October and the Rabi season begins in November and ends in March. Average yearly rainfall is calculated using the anticipated monthly rainfall and therefore the average temperature during the Rabi season is taken into account to predict crop yield for wheat, which may be a Rabi crop.

TABLE IX: Classification Rules

Classification	**Average Yearly Rain**	**Average Season Temperature**
Very Good	62-87 cm	20-25 C
Good	50-62 cm or 87-100cm	15-20 C
Average	25-50 cm or 100-150 cm	12-15 C or 25-33 C
Bad	10-25 cm or 150-200cm	8 -12 C or 33-40 C
Very Bad	All other values	All other values

IF (rain=='very good' and temp=='very good') or (rain=='very good' and temp=='good'):
THEN yield = 'very good'

IF (rain=='very good' and temp=='average') or (rain=='good' and temp=='very good') or (rain=='good' and temp=='good') or (rain=='good' and temp=='average'):
THEN yield = 'good'

IF (rain=='very good' and temp=='bad') or (rain=='very good' and temp=='very bad') or (rain=='good' and temp=='bad') or (rain=='good' and temp=='very bad') or (rain=='average' and temp=='very good') or (rain=='average' and temp=='good') or (rain=='average' and temp=='average'):
THEN yield = 'average'

IF (rain=='average' and temp=='bad') or (rain=='average' and temp=='very bad') or (rain=='bad' and temp=='very good') or (rain=='bad' and temp=='good') or (rain=='bad' and temp=='average') or (rain=='bad' and temp=='bad'):
THEN yield = 'bad'

IF (rain=='bad' and temp=='very bad') or (rain=='very bad' and temp=='very good') or (rain=='very bad' and temp=='good') or (rain=='very bad' and temp=='bad') or (rain=='very bad' and temp=='average') or (rain=='very bad' and temp=='bad') or (rain=='very bad' and temp=='very bad'):
THEN yield = 'very bad'

The temperature for Rabi season i.e., temperatures from the month of November to the March of next year is combined to get seasonal average temperature which is given to the fuzzy model along with the predicted rainfall to get results.

VIII. RESULTS

Table X gives the actual and predicted annual average rain, the actual and predicted average temperature of Rabi season and the actual and predicted crop yield of wheat for Allahabad from the year 1997 to 2001.

TABLE X: Crop yield of wheat for Allahabad

Year	Actual Rain	Predicted Rain	Actual Temp	Predicted Temp	Actual Yield	Predicted Yield
1997	62.95	75.26	19.41	18.52	Very Good	Very Good
1998	94.71	91.30	20.29	18.43	Good	Good
1999	89.51	88.74	20.32	18.34	Good	Good
2000	71.71	79.52	19.98	18.24	Very Good	Very Good
2001	66.84	76.70	19.87	18.14	Very Good	Very Good

IX. CONCLUSION

According to the results, temperature is best predicted by the SARIMA model and the accuracy of predictions made for rainfall by ARMA model is better than ARMAX model.

Rainfall, which is a crucial factor for the prediction of crop yield is difficult to estimate precisely. Climate factors may change thanks to other remaining variables which can influence the prediction of rainfall.

Also, the proposed work makes use of symbolic logic to estimate crop yield which works on a group range instead of discrete values, therefore, the error in predicted rainfall data does not cause problems as long because the difference between actual and estimated values isn't drastic. Our model can successfully predict crop yield for a given year when the rainfall and temperature values for the previous years is known.

REFERENCES

[1] J. Rodrigues, and A. Deshpande. "Prediction of Rainfall for all the States of India Using Auto-Regressive Integrated Moving Average Model and Multiple rectilinear regression ." 2017 International Conference on Computing, Communication, Control and Automation (ICCUBEA), pp. 1-4. IEEE, 2017.

[2] P.Kumar "Crop yield forecasting by adaptive neuro fuzzy inference system." Mathematical Theory and Modeling 1.3, pp. 1-7, 2011.

[3] https://www.kaggle.com/abhiseklewan/crop-production-statisticsfrom-1997-in-india.

[4] http://www.indiawaterportal.org/met_data.

[5] I. Kaushik, and S.M. Singh. "Seasonal ARIMA model for forecasting of monthly rainfall and temperature." Journal of Environmental Research and Development 3, no. 2, pp. 506-514, 2008.

[6] S. Veenadhari, B. Misra, and C. D. Singh. "Machine learning approach for forecasting crop yield based on climatic parameters." 2014 International Conference on Computer Communication and Informatics, pp. 1-5. IEEE, 2014.

[7] S.Prabakaran, P.N.Kumar, and P.S.M. Tarun. "Rainfall prediction using modified linear regression.", 2006.

[8] S.K. Mohapatra, A. Upadhyay, and C. Gola. "Rainfall prediction based on 100 years of meteorological data." 2017 International Conference on Computing and Communication Technologies for Smart Nation (IC3TSN), pp. 162-166. IEEE, 2017.

[9] Sundaram, S.Meenakshi, and M. Lakshmi. "Rainfall prediction using seasonal auto regressive integrated moving average model." Computer science 3, no. 4 (2014), pp. 58-60, 2014.

[10] http://agropedia.iitk.ac.in/content/climatic-requirement-wheattemperature
http://agropedia.iitk.ac.in/content/rainfall-climaticrequirement-wheat