Analysis of Crop Yield Based on Temperature at Particular Region

Shivakumar Valpady

***Abstract*—India is predominantly an Agro-based economy. However, farmers cannot estimate whether their crop in a season will fetch him invested amount. Basically, a farmer estimates the price his crop can fetch through the monsoon that is occurring in his locality and his soil conditions alone. The traditional method of dividing the crop growth based on Kharif and Rabi season provides only a rough estimate but not accurate results. The farmer is unable to predetermine what his crop can fetch for a season mainly staple grains such as Wheat. This project aims to build an application using prediction model in Machine Learning to aid the farmers in providing an estimate of the optimum temperature their Wheat crop should be grown to fetch the maximum yield. To predict both temperature and Yield using previous seasonal price data, a Time-Series Forecast method needs to be used. An important time-series forecast is Auto Regressive Integrated Moving Average (ARIMA) which predicts the crop price based on the seasonality and trends of the previous prices has been used to predict the temperature and Yield. The results of which are combined for three different states and analysis is generated.**

Keywords*—*Optimum Temperature, Maximum Yield, Predict, ARIMA, Wheat Crop etc.

I. Introduction

India is tropical country with Climate generally suitable for any kind crop as the seasons determine the crop to be grown. 58% of the land is used for cultivation with over 62% of the population dependent on the Agro-Production based industry. India is the second largest producer of wheat globally with the export quantity in 1000MT. The leading states of the wheat production are Uttar Pradesh, Madhya Pradesh, Haryana, Punjab, Rajasthan etc. However, wheat is a winter crop, more predominantly known as Rabi Crop. It is sown at the end of monsoon (Around October) and the yield is harvested around February. The Few states like Assam produce both Spring and Winter Wheat as their temperature is maintained constant and is below 10 degrees centigrade throughout the year.

The wheat crop yield is dependent on various meteorological factors. Major of which are Local Temperature, Soil Moisture Retention Rate, Humidity, Soil composition, Precipitation etc. Determining the relation between the optimum temperature to obtain the maximum yield is the main objective of this project.

II. Data Flow Diagram

The methodology to obtain the optimal temperature requires the prediction of both temperature and yield data for the period between 2011 and 2018 as the values predicted will be incorporating global weather changes included in last few years as trend. Once the temperature and yield are predicted, the optimal temperature corresponding to maximum yield is taken as output.

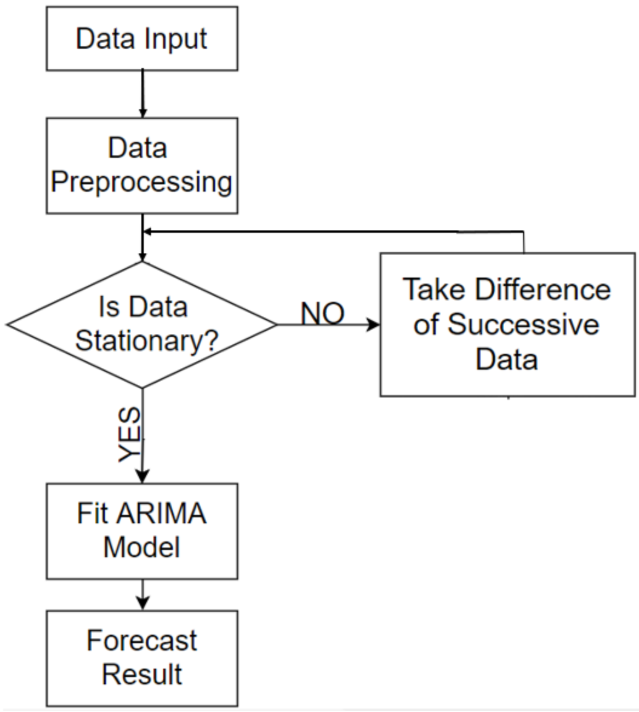


Fig. 1. Data Flow Diagram

A. *Data Collection and Pre-Processing*

Two Different Data Sets regarding the annual temperature status for the given states (Uttar Pradesh, Assam and Rajasthan) were obtained from 1960-2018 in Comma Separated Values (CSV) format. The data set size was around 2 MB in total and were stored as individual files.

Meteorological Data obtained was of only Year and respective annual Temperature for that specific state. The major source of data was Indian Meteorological Department (IMD) and AgMark. The Data Set so obtained contained null values, duplicate values and Not Applicable values such as negative values in Yield section.

Pre-Processing of data included the cleaning of data with removal of Null values, duplicate values and Not applicable values. The pre-processed data is saved back into CSV file and stored in a different file. After pre-processing the size of the data set was about 1.5MB.

B. *Data Processing*

The Data Set had only two columns: Year and Annual Temperature or Year and Yield. Both the column values are strings. The ARIMA Model is time series forecast model and hence the index is expected to be in the format of DateTimeIndex. Hence the first column named "Year" is converted into the form of DateTimeIndex and set as the index value for the next step in processing.

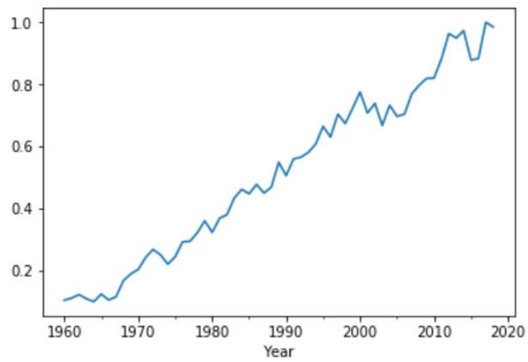


Fig. 2. Plot of Yield Against Year

III. ARIMA METHODOLOGY

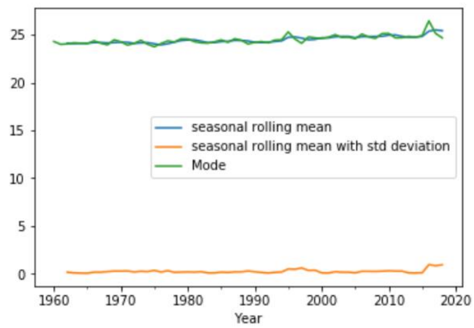
Before using ARIMA model, there is a prerequisite that the Data Set being used should be stationary. A stationary dataset is one in which there is no presence of Seasonality, Cyclicity, Trend and extreme randomness. In the above cases, ARIMA is not required as a standard function or mathematical equation can be used instead for prediction of values. Hence, a stationary dataset cannot be used as input for ARIMA model.

A. Augmented Dickey Fuller test

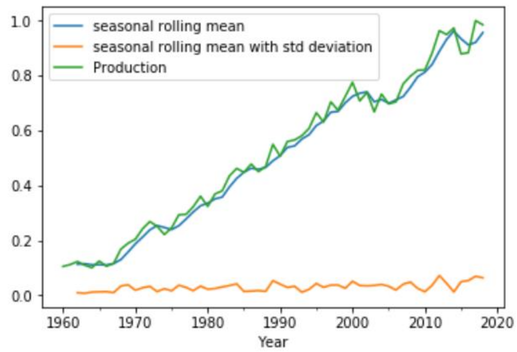
Augmented Dickey-Fuller test (ADF) is a stationarity test that uses null hypothesis to reject that there exists a unit root to the time series model. Hence, the static version is used which is a negative value and the more negative that value is the more stationary the dataset is.

B. Rolling mean plot

- Another method to determine the stationarity of a data set is by plotting the seasonal rolling mean of a data set with its standard deviation. If there is a trend, seasonality or cyclicity then the data set is said to be stationary.
- For our data set we plot a seasonal rolling mean considering 3 years as one sowing season. Another graph is plotted to for 10 years to obtain yearly rolling mean. From both it is observed that stationarity exists.



(a)

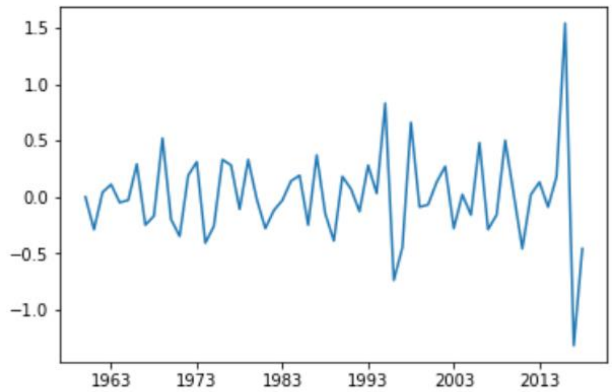


(b)

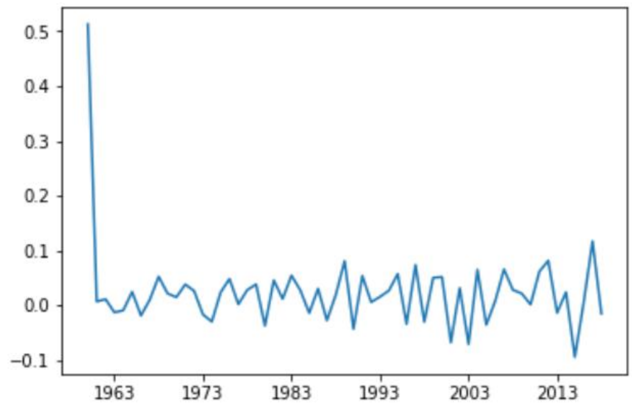
Fig. 3. Seasonal Rolling Mean plot (a) Temperature (b) Yield

C. First Difference and Seasonal Decompose

To convert the non-stationary data set into stationary data set, the seasonality, cyclicity and trend should be removed. This can be removed by taking the first difference of the data set. The ADF test shows that the resulting data set is stationary.



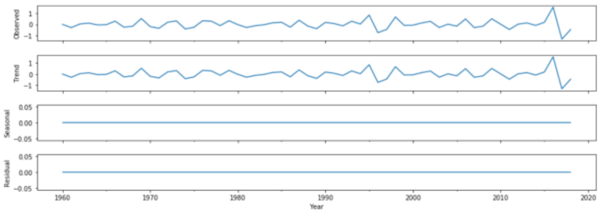
(a)



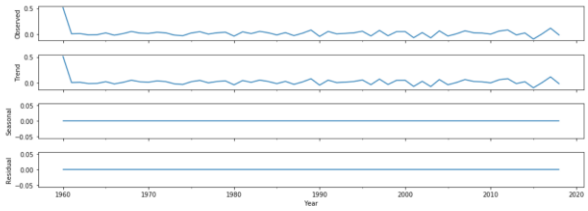
(b)

Fig. 4. Plot After First Difference (a) Temperature (b) Yield.

The Seasonal decompose of the data set is taken through the direct usage of statsmodel.tsa library. The graph so obtained shows that there is no seasonality remaining the data set.



(a)



(b)

Fig. 5. Decomposition Graph (a) Temperature (b) Yield

D. *Autocorrelation Plot and Partial Autocorrelation Plot*

The ARIMA model contains three input parameters. They are (p,q,d). They each stand for:

- **p** is the number of autoregressive terms,
- **d** is the number of non-seasonal differences needed for stationarity, and
- **q** is the number of lagged forecast errors in the prediction equation.

The autocorrelation plot determines the q value. The initial point where the plot line crosses the safe zone, i.e, the dotted line area is taken as the q value.

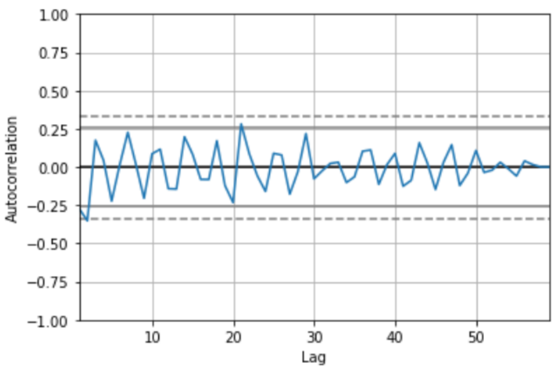


Fig. 6. Autocorrelation Graph for Temperature.

- The Partial Autocorrelation graph is used to determine the p value.
- Generally the p value will be greater than or equal to the q value which in turn will be greater than or equal to d value. Thus, d value is determined.

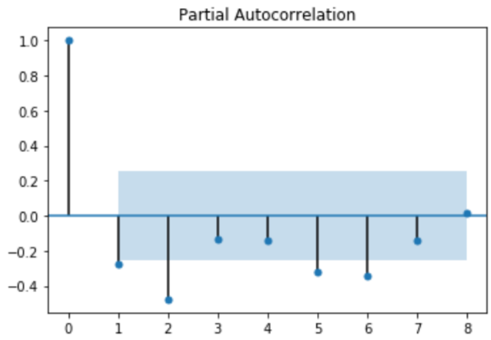
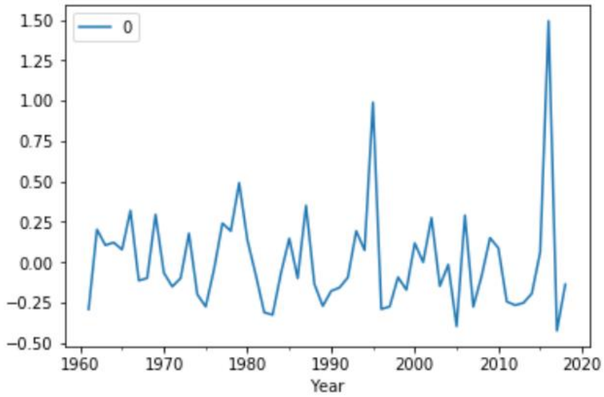
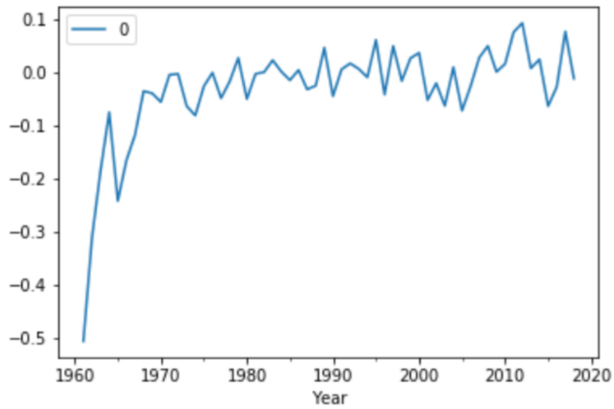


Fig. 7. Partial Autocorrelation Graph for Temperature.

- To verify whether the model will fit with the corresponding p,q,d values, a residuals graph is plotted.



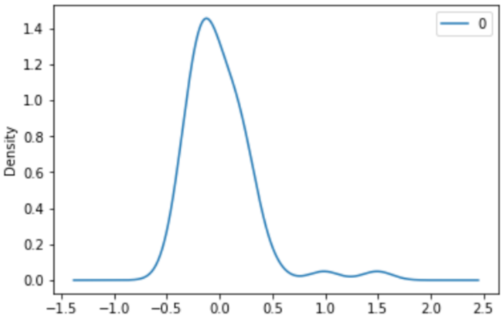
(a)



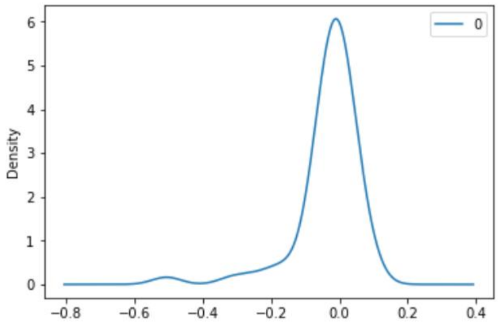
(b)

Fig. 8. Residuals Graph for (a) Temperature (b) Yield.

- A graph to determine the gaussian resulting bias is plotted to check whether the prediction values are biased positively or negatively. If the result is closer to zero, then the predicted values are not biased.



(a)



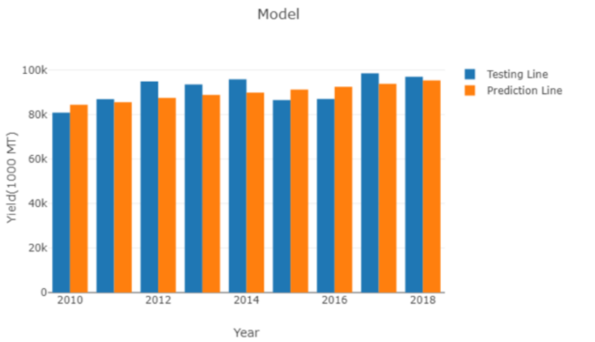
(b)

Fig. 9. Bias Plot for (a) Temperature (b) Yield.

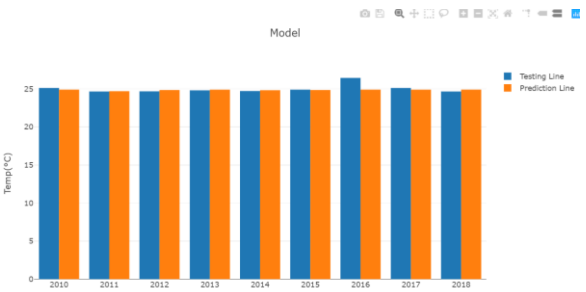
IV. Prediction and Results

The available data set is split into train data and test data, with 90% of the data used for training. The test data is passed through the fitted ARIMA model and the predicted values are printed against the expected values.

A graph is plotted for the predicted values against the testing points for both yield and temperature.



(a)



(b)

Fig. 10. Prediction Graph for (a) Temperature (b) Yield.

A. Error Calculations

To calculate the accuracy of the prediction, the Root-Mean Squared error is calculated for the predicted values against the expected values. Also, error percentage is calculated.

Each individual state has different error values and error percentage values. This is because other meteorological factors have not been taken into consideration.

B. Merging Two Predications

The resulting prediction values of temperature and yield and stored into separate Comma Separated Value (CSV) files. Then they are merged based on index and stored into a new CSV file. From this file the maximum value for yield is obtained. Then the corresponding temperature is the optimum temperature for the growth of wheat. This temperature value is extracted from the file and the results are printed.

C. Analysis

The prediction was carried out three different states: Uttar Pradesh, Assam and Rajasthan. These three places where considered specifically due to the availability of data and the variation in temperature zone.

Uttar Pradesh generally has a tropical climate of hot summers and cold winters. The soil fertility and the soil temperature with humidity is moderate in Uttar Pradesh. However, Rajasthan is an arid state with less soil fertility and higher temperature. Meanwhile, Assam is a highly fertile land with consistent cold climate throughout the year.

Only Assam out of the three states produces both Spring Wheat and Winter wheat due to its low temperature consistency. The maximum yield produced in the above three states with their respective optimum temperature is displayed in the table below.

TABLE 1: Analysis Table

Region	Maximum Yield (1000 MT)	Temperature (ºC)
Uttar Pradesh	97426.066793	24.913573
Assam	134209.097334	7.328798
Rajasthan	9026.626268	23.491856

The Maximum yield is calculated in terms of 1000 Metric Tonnes and the temperature is in terms of degree centigrade.

Even though Assam is the highest producer based on the prediction, it is only because Assam produces both spring wheat and winter wheat. The annual mean temperature of Assam is consistently low compared to other states and leading to higher production of Wheat. However, due to the rich fertility of the soil available and other conducive agrometeorological factors such as precipitation, humidity and soil moisture retention rate, the land is cultivated with other commercial crops such as tea and coffee.

Another factor that theoretically provides higher yield is due to the altitude of the location from sea level. This leads to marking that yield is only partially dependent on global temperature. However, Assam has a very small fluctuation in temperature year-round and is barely affected by increasing global temperature, while both Uttar Pradesh and Rajasthan have fluctuating temperature throughout the year. Hence it is necessary to reduce the effect global warming by creating a controlled environment for growth of Wheat with consist lower temperature throughout the year.

REFERENCES

[1] G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. *(references)*

[2] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[3] I. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[4] K. Elissa, "Title of paper if known," unpublished.

[5] R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.

[6] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[7] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.