International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2023; 8(6): 118-123

https://www.mathsjournal.com
Received: 16-07-2023
Accepted: 20-08-2023

Monika Devi
Department of Mathematics & Statistics, CCS Haryana Agriculture University, Hisar, Haryana, India

Joginder
Department of Mathematics & Statistics, CCS Haryana Agriculture University, Hisar, Haryana, India

Dalip Kumar Bishnoi
Department of Agricultural Economics, CCS Haryana Agriculture University, Hisar, Haryana, India

Corresponding Author:
Monika Devi
Department of Mathematics & Statistics, CCS Haryana Agriculture University, Hisar, Haryana, India

An application of explanatory variables to model and forecast sugarcane yield

Monika Devi, Joginder and Dalip Kumar Bishnoi

Abstract
This study delves into the intricate relationship between climatic variables and sugarcane productivity in India, offering valuable insights into the factors affecting crop yield. This paper's major goal is to make estimates of how climatic factors affect sugarcane productivity. Pre- harvest models; i.e., Principal component analysis, discriminant function analysis and Post-harvest models; i.e., ARIMA and ARIMAX models are all used to examine the consistency of empirical results. The data set includes data spanning 40 years, from 1980 to 2019. All of these models have productivity of the sugarcane in Yamuna Nagar district as a dependent variable. Accuracy results revealed that univariate models have lesser accuracy as compared to the models with weather parameters. Discriminant function analysis has the higher level of accuracy in sugarcane yield forecasting and found best among all tried models. Also, selected model was found significant along-with individual scores. In discriminant function analysis 20th fortnight (16th Oct-31th Oct) is the best time for forecasting the sugarcane yield. Hence, use of weather parameters was found contributing positively towards the yield forecasting of sugarcane crop.

Keywords: Sugarcane yield, principal component analysis, discriminant function analysis, ARIMA, ARIMAX

Introduction
Sugarcane holds a pivotal role in India's agricultural economy as one of the most significant cash crops. Cultivated by approximately six million farmers, its growth also provides employment for a vast number of agricultural workers. With a history spanning over 4,000 years, the cultivation of sugarcane for refined sugar production has been deeply rooted in India. The rich composition of sugarcane juice, abundant in potassium, calcium, magnesium, iron, zinc, thiamine, riboflavin, and amino acids, makes it a valuable resource for medicinal purposes, addressing various ailments.

Sugar derived from sugarcane finds extensive usage in the confectionery industry, followed by bakers and cereal manufacturers. It serves as a common sweetener in various food and beverage products. Notably, sugarcane juice contains 113.43 calories, 0.20 g of protein, 0.66 g of fat, and 25.40 g of carbohydrates. Beyond its nutritional value, sugarcane is also a rich source of polyphenols and flavonoids, contributing to its antioxidant properties that reduce oxidative stress and enhance overall health.

Globally, sugarcane accounts for approximately 80% of the world's sugar supply, with India standing as the second-largest producer after Brazil. In the 2022–2023 season, Maharashtra emerged as India's leading sugarcane producer, yielding over 138 lakh tonnes of this versatile crop. Sugarcane finds its utility in an array of products, including sugar, jaggery, khansari, molasses, and even paper production.

With ambitious plans to produce 35 million tonnes of sweeteners by 2030, India stands as the world's second-largest producer and consumer of sugar. Sugarcane, a prominent Kharif crop, occupies a significant position among the various crops cultivated in India, providing livelihood to nearly 60% of the nation's population. The conducive climate in India supports year-round sugarcane plantations. The country is renowned as the world's leading producer, consumer, and second-largest exporter of sugar. The Indian Sugar Mills Association (ISMA) reported a 3.69% rise in sugar production, totaling 12.07 million tonnes during the October-December quarter of 2022.

Maharashtra takes the lead in sugarcane production, contributing 138 lakh tonnes in 2022–2023, with Uttar Pradesh, Karnataka, and Haryana collectively contributing 80% of the nation's sugarcane output. Haryana, known for its significant contributions to India's food grain supply, also produces substantial quantities of rice, jawar, bajra, and maize, alongside sugarcane cultivation. This versatile crop occupies more than 1.3 million hectares of land, yielding over 8 million tonnes. India's sugar industry stands as the second-largest agro-based industry in the country, following cotton production. To thrive, sugarcane requires specific conditions, including temperatures ranging between 21-27 °C, a hot and humid climate, 75-100 cm of annual rainfall, and deep, rich loamy soil.

Despite sugarcane's vital role in India's agricultural landscape, there remains a notable gap in research exploring the connection between climate change and sugarcane productivity. Various past studies have examined how climate change could lead to a decline in the production of essential food and income crops, raising concerns about the impact on sugarcane productivity. Yet, it remains uncertain whether climate change will ultimately boost or hinder sugarcane productivity.

In light of these uncertainties, several crucial questions emerge. What role do climatic variables play in influencing sugarcane productivity in India? How do annual variations influenced by climate change impact sugarcane productivity? What are the broader implications of these climate-induced fluctuations on the sugarcane industry? Addressing these queries becomes imperative for shaping effective agricultural policies.

Several studies have delved into the intricacies of forecasting and modeling sugarcane production, often considering the influence of climatic variables. Singarasa (2015) [8] conducted a study on forecasting sugarcane production in India, while Deenapanray and Goburdhun (2012) [2] employed ARIMA models for predicting sugarcane yields. Time series analysis was utilized by Singh and Sinha (2013) [10] for modeling and forecasting sugarcane yields. Additionally, the work of Srivastava and Rai (2012) [11] encompassed a comprehensive review of crop yield forecasting methods. Furthermore, Zhang, Li, and Xu (2015) [13] applied grey prediction techniques to forecast short-term sugarcane yields, and Suryavanshi, Deosarkar, and Rairakhwada (2017) [12] explored the application of machine learning for sugarcane yield prediction. Moreover, general agricultural literature by Lobell and Burke (2010) [5], Hatfield and Prueger (2015) [3], Jones and Thornton (2003) [4], and Olsen *et al*. (2011) [7] provided insights into modeling crop yield responses to climate change and the impacts of temperature extremes. Lastly, Singh and Boote (1986) [9] contributed to understanding sugarcane's response to temperature. These references collectively provide a foundational understanding of the factors affecting crop yield, particularly in the context of sugarcane productivity and climate influences.

To bridge the existing research gap, this current study aspires to unravel the complex relationship between climate and sugarcane productivity in India. Its primary objective is to assess the impact of climatic and non-climatic factors on sugarcane productivity across India's diverse seasons—monsoon, winter, and summer. Given that sugarcane is an annual crop with a growth cycle spanning 12 to 18 months, this research promises to provide invaluable insights into the influence of various climate factors on distinct sugarcane growing seasons. This study's significance lies in its endeavor to decipher the multifaceted impact of climate conditions on sugarcane productivity.

Data and methodology

The locale of the experiment was the Yamunanagar district of Haryana state. The selected crop for investigation was sugarcane, and the study encompassed data spanning from 1981 to 2019. This research was conducted using secondary data collected from various published sources, as data on sugarcane yield were gathered from statistical abstract of Haryana, alongside the collection of weather parameters, including maximum temperature, minimum temperature, and rainfall. From Indian meteorological department (IMD).

Box-Jenkins Autoregressive Integrated Moving Average methodology

ARIMA is the most comprehensive time series data forecasting model. ARIMA forecasts only use the variable's past values. They are specifically designed for short-term forecasting and do not rely on any other data series. This method can be applied to both, regularly spaced time intervals continuous and discrete data. Furthermore, developing an ARIMA model requires a minimum sample size of 50 observations by Pankraz (1991) and only works with stationary data.

Stationarity of a time series

A time series is considered to be stationary if mean, variance and autocorrelation remain mostly constant across time. An ARIMA model is identified by the notation ARIMA (p, d, q), where p, d, and q stand for the orders of autoregressive, differencing and moving average respectively. A firstorder autoregressive model of ARIMA for a certain time series Y_t (1, 1, 0) is simply AR (1)

$$Y_t = \mu + \phi_1 Y_{t-1} + e_t$$

A first order moving average model represented by ARIMA (0,0,1) is simply MA (1) given by

$$Y_t = \mu + \theta_1 e_{t-1} + e_t$$

Alternatively, the model may also be a mixture of AR and MA of higher orders as well.

$$Y_t = \mu + \phi_1 Y_{t-1} + \phi_1 Y_{t-2} \ldots\ldots \phi_1 Y_{t-p} - \theta_1 e_{t-1} - \theta_1 e_{t-2} \ldots\ldots \theta_1 e_{t-q} + e_t$$

This is called a mixed auto regressive moving average model of order (p, q). Alternatively, an ARIMA (p,d,q) may be written in polynomial form as :

$\phi_p(B)\Delta^d Y_t = C + \theta_q(B)\ a_t$

Y_t= Variable under forecasting
B = Lag operator
a_t= Error term ($Y_t - \hat{Y}_t$,
where
$\hat{Y}t$ is the estimated value of Y_t)
$\phi_p(B)$= Non-seasonal AR
$(1-B)^d$= Non-seasonal difference
$\theta_q(B)$ = Non-seasonal MA
The above model contains p+q parameters, which need to be estimated.

Principal component analysis

Principal Component Analysis (PCA) is a powerful technique employed for dimensionality reduction in datasets that involve a large number of interrelated variables, all while retaining a significant portion of the original variation. This dimensionality reduction is achieved by transforming the original variables into a new set of variables known as principal components (PCs). These PCs are carefully designed to be uncorrelated with one another and are ordered in such a way that the first few components capture the majority of the variation present in the data.

These principal components are linear combinations of the original variables, and they adhere to specific principles:

- The first principal component (PC1) is constructed to retain the maximum possible variance.
- Subsequent principal components are also optimized to capture as much variance as possible while being orthogonal (uncorrelated) to the preceding components.

In our research study, we have conducted PCA on all the weekly weather variables to extract the most crucial and meaningful variables. Only the principal components (PCs) with eigenvalues exceeding 1 were considered, as proposed by Brejda *et al*. (2000). The application of PCA serves the purpose of mitigating issues related to multicollinearity and overfitting, both of which are common challenges when dealing with high-dimensional datasets.

Subsequently, stepwise regression analysis was performed using PC scores to formulate crop yield models. This approach allows for a more streamlined and effective analysis of the relationship between weather variables and crop yield, helping to identify and emphasize the most significant factors influencing the outcome.

RMSE (Root Mean Square Error) and MAD (Mean Absolute Deviation) are both metrics used to measure the accuracy of forecasts or predictions. Here are the formulas for each:

1. Root Mean Square Error (RMSE): RMSE is a measure of the average error or the square root of the average of the squared differences between predicted and actual values.

2. Mean Absolute Deviation (MAD):

- MAD measures the average absolute difference between predicted and actual values.

These metrics are often used in various fields, including statistics, machine learning, and time series analysis, to evaluate the accuracy of forecasting models. Smaller RMSE and MAD values indicate better accuracy, as they represent smaller errors between predictions and actual observations.

Results

The following results have been obtained based on secondary data of sugarcane yield and weather variables for the period 1980-2018 of Yamunanagar district of Haryana:

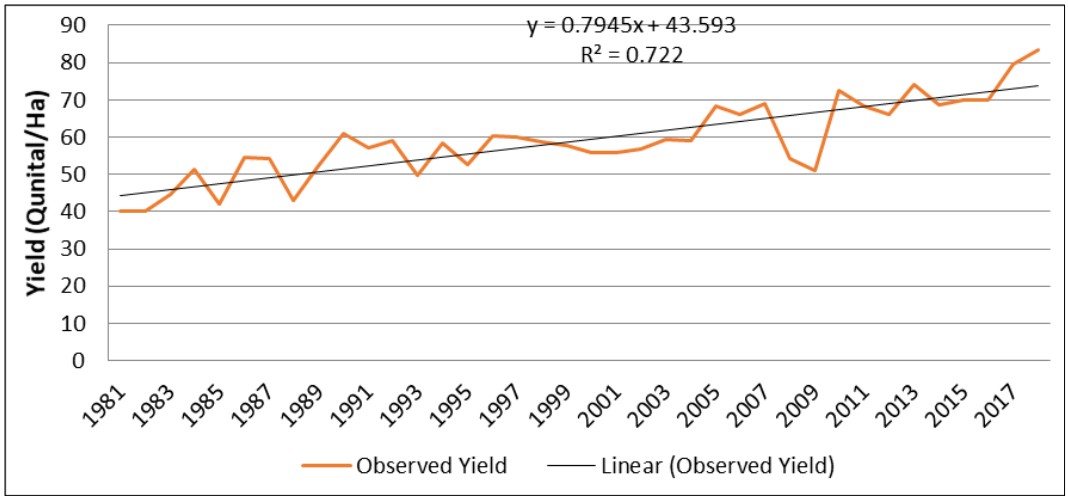

Fig 1: Trend line of sugarcane yield in Yamunanagar district

Table 1: Summary statistics of sugarcane yield of Yamuna Nagar:

Statistic	Mean	Standard Deviation	Range	Minimum	Maximum	CV
Yield (Quintal/Ha)	59.08	10.39	43.51	40.02	83.53	17.59

Fig. 1 shows the trend line for yield of sugarcane crop of Yamuna Nagar district. The trend of sugarcane yield was found almost linear with 0.722 value of R^2.

The summary statistics for sugarcane yield in Yamuna Nagar reveal important insights into the crop's performance in the region. On average, the sugarcane yield stands at approximately 59.08 Quintals per Hectare, with a relatively low standard deviation of 10.39, indicating a relatively moderate level of variability around the mean yield. The observed range in yield is substantial, spanning from a minimum of 40.02 Quintals per Hectare to a maximum of 83.53 Quintals per Hectare. This significant range signifies the potential for both challenges and opportunities in sugarcane production within Yamunanagar. The coefficient of variation (CV), a measure of relative variability, is calculated at 17.59%, demonstrating that the yield's variation is moderately high compared to the mean, and further highlighting the importance of understanding and managing factors affecting sugarcane productivity in this region.

Table 2: Relation between the weather parameters and the sugarcane yield (Yamuna Nagar) at different growth phases of the crop:

Variables		Germination (0-45 days)	Tillering (45-120) days	Grand Growth (120-240 days)	Ripening (Rest)
Temperature Maximum	Correlation	0.218	0.026	-.104	-0.183
	Sig.	0.189	0.875	.536	0.272
Temperature Minimum	Correlation	0.536**	0.299	.314	0.326*
	Sig.	0.001	0.068	.055	0.046
Rainfall	Correlation	-.228	0.000	.098	-0.259
	Sig.	.168	0.999	.558	0.116

Table 2 illustrates the relationships between various weather parameters and the sugarcane yield in Yamunanagar during different growth phases of the crop.

During the germination phase (0-45 days), the maximum temperature showed a positive correlation of 0.218, although it was not statistically significant (Sig. = 0.189). Conversely, the minimum temperature exhibited a stronger and statistically significant positive correlation of 0.536** (Sig. = 0.001). The negative correlation with rainfall (-0.228) was observed, but it was not statistically significant (Sig. = 0.168).

Moving on to the tillering phase (45-120 days), the maximum temperature displayed a minimal correlation of 0.026, which was not statistically significant (Sig. = 0.875). The minimum temperature showed a correlation of 0.299, indicating a positive relationship, although it was not statistically significant (Sig. = 0.068). Rainfall had no correlation with sugarcane yield during this phase (Correlation = 0.000), and the lack of significance was evident (Sig. = 0.999).

In the grand growth phase (120-240 days), the maximum temperature exhibited a negative correlation of -0.104, which was not statistically significant (Sig. = 0.536). Similarly, the minimum temperature (Correlation = 0.314) and rainfall (Correlation = 0.098) showed positive correlations, yet neither was statistically significant (Sig. = 0.055 and Sig. = 0.558, respectively).

During the ripening phase (rest), the maximum temperature displayed a negative correlation of -0.183, while the minimum temperature showed a slightly stronger negative correlation of -0.326*, both of which were not statistically significant (Sig. = 0.272 and Sig. = 0.046, respectively). Rainfall had a more substantial negative correlation of -0.259 but was not statistically significant (Sig. = 0.116).

These findings suggest that the relationships between weather parameters and sugarcane yield at various growth phases in Yamunanagar are characterized by both positive and negative correlations, though many of them do not reach statistical significance. It underscores the complex interplay between weather and crop productivity, and further investigation may be necessary to determine the specific impact of these weather parameters on sugarcane yield during different growth stages.

Table 3: Forecast for next five years from selected models:

Model	Post-Harvest Forecast Models					Pre-Harvest Forecast Models	
Year	ARIMA (0,1,1)	ARIMA (0,1,1) tmn2	ARIMA (0,1,1) tmn8	ARIMA (0,1,1) rf6	ARIMA (0,1,1) rf6, tmn2	PCA*	DFA*
2014	71.11	71.11	76.35	70.74	70.38	70.35	70.27
2015	69.61	69.58	74.71	71.37	70.68	71.41	71.09
2016	71.33	71.33	74.04	69.77	69.40	74.15	71.56
2017	71.07	71.09	75.68	71.28	71.34	73.71	78.04
2018	70.58	70.61	74.83	73.58	74.14	74.34	78.52

*PCA-Principal component analysis, DFA-Discriminant function analysis

The table provides forecasts for the next five years from selected models, differentiating between post-harvest and pre-harvest forecast models. Various combinations of ARIMA have been tried and with lower AIC, BIC value ARIMA (0,1,1) was selected for forecasting sugarcane yield. ARIMA with weather parameter (exogenous parameters) (ARIMAX) was also used with several exogenous parameters combination and selected models along with their forecast values are presented in the table 3. Further Principal component analysis and discriminant function analysis was also performed. Table 3 gives the forecast figures for last five years obtained from various applied models.

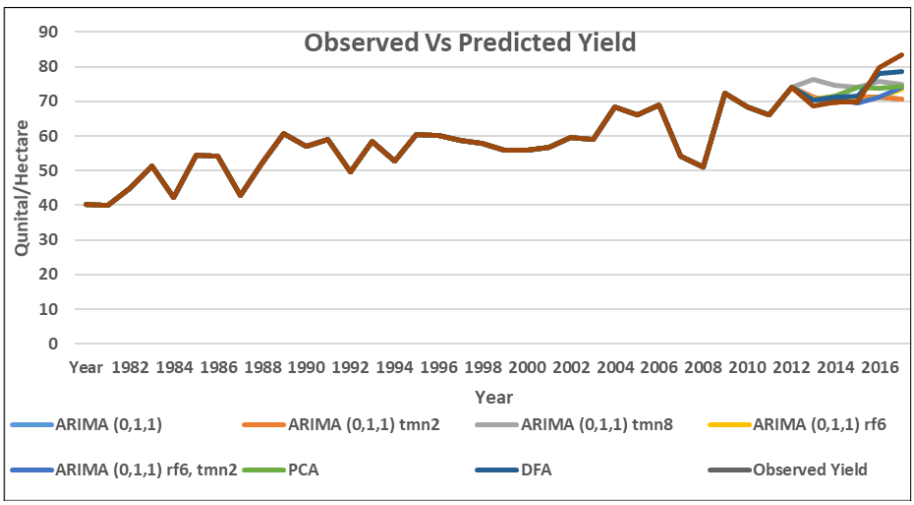

Fig 2: Observed vs predicted yield

Table 4: Accuracy measures of selected models

Model	ARIMA (0, 1, 1)	ARIMA (0, 1, 1) tmn2	ARIMA (0, 1, 1) tmn8	ARIMA (0, 1, 1) rf6	ARIMA (0, 1, 1) rf6, tmn2	PCA	DFA
RMSE	7.07	7.05	6.18	5.93	5.68	5.35	2.62
MAD	5.14	5.14	5.86	4.40	4.14	4.52	2.22

A comparison of accuracy measures of all models revealed that Root mean square error (RMSE) and Mean absolute deviation (MAD) were less in Discriminant function analysis (DFA). Also, RMSE and MAD were found least in all tried fortnights in Discriminant function analysis in comparison to other applied models. These accuracy measures are essential for assessing the reliability and performance of each model in predicting sugarcane yields. Lower RMSE and MAD values indicate a more accurate forecast, and the results suggest that the DFA model is the most accurate among the models considered.

Table 5: Fitted model using discriminant function analysis (DFA) for Sugarcane yield

Fortnight of forecast	Fitted Regression Model	P-value
20th fortnight (16th Oct-31th Oct)	Yield = -1441.653+0.751**T+0.510**Z_1-0.417*Z_2	p< 0.01

Note: *Significant at 5% level of significance, **Significant at 1% level of significance

Above table provides the model summary of selected model and overall model was found significant along-with individual scores. Thus, 20th fortnight is the best time for forecasting the yield. The comparison of observed and forecasted yield of subsequent years using selected model are shown in the figure 2 given below.

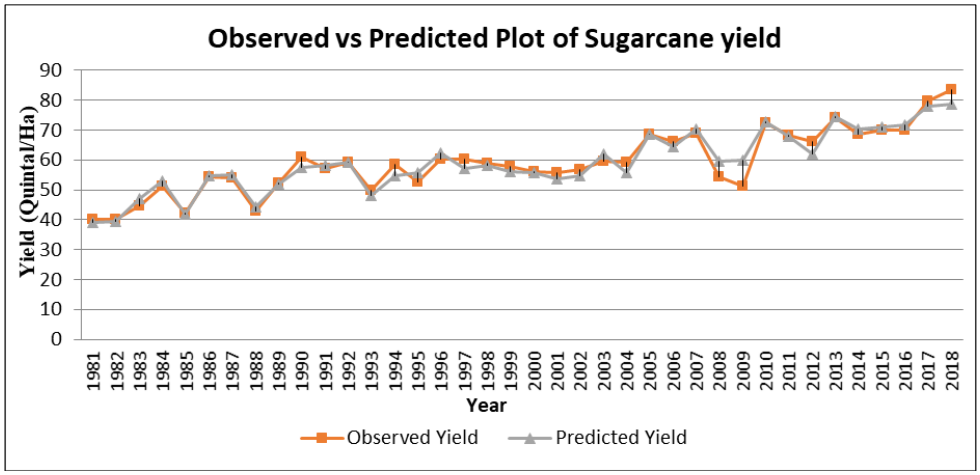

Fig 3: Comparison of observed and forecasted yield

Conclusion

The trend of sugarcane yield is found almost linear with 0.722 value of coefficient of determination (R^2). Also variability in sugarcane yield was about 17.59 per cent. Accuracy results revealed that univariate models have lesser accuracy as compared to the models with weather parameters. Discriminant function analysis has the higher level of accuracy in sugarcane yield forecasting and found best among all tried models. Also, selected model was found significant along-with individual scores. In discriminant function analysis 20th fortnight (16th Oct-31th Oct) is the best time for forecasting the sugarcane yield. Hence, use of weather parameters was found contributing positively towards the yield forecasting of sugarcane crop.

References

1. Akinseye FM, Ojeniyi SO. Crop-Weather Relationship and Climate Change Impacts on the Growth and Yield of Sugarcane in Nigeria. Journal of Agrometeorology. 2017;19(2):213-220.
2. Deenapanray PN, Goburdhun D. Modeling and Forecasting of Sugarcane Yield Using ARIMA Models. Proceedings of the Annual International Conference on Operations Research and Statistics. 2012;11(1):11-15.
3. Hatfield JL, Prueger JH. Temperature Extremes: Effect on Plant Growth and Development. Weather and Climate Extremes. 2015;10:4-10.
4. Jones PG, Thornton PK. The Potential Impacts of Climate Change on Maize Production in Africa and Latin America in 2055. Global Environmental Change. 2003;13(1):51-59.
5. Lobell DB, Burke MB. On the Use of Statistical Models to Predict Crop Yield Responses to Climate Change. Agricultural and Forest Meteorology. 2010;150(11):1443-1452.
6. Lohiya NK, Choudhary AK. A Comprehensive Review on Crop Yield Forecasting. International Journal of Agricultural and Biological Engineering. 2018;11(5):63-75.
7. Olsen JE, Trnka M, Kersebaum KC, Skjelvåg AO, Seguin B. Impacts and Adaptation of European Crop Production Systems to Climate Change. European Journal of Agronomy. 2011;34(2):96-112.
8. Singarasa V. A Study on Forecasting Sugarcane Production in India. International Journal of Research in Management, Science & Technology. 2015;3(2):142-149.
9. Singh RP, Boote KJ. Sugarcane Response to Temperature. Field Crops Research. 1986;14:177-190.
10. Singh R, Sinha R. A Study on Modeling and Forecasting of Sugarcane Yield Using Time Series Analysis.

International Journal of Engineering and Innovative Technology (IJEIT). 2013;2(8):7-11.
11. Srivastava A, Rai DC. A Review on Crop Yield Forecasting Methods. International Journal of Computer Applications. 2012;48(8):35-41.
12. Suryavanshi S, Deosarkar D, Rairakhwada D. Sugarcane Yield Prediction Using Machine Learning. 2017 2nd International Conference for Convergence in Technology (I2CT); c2017. p. 1064-1067.
13. Zhang X, Li M, Xu X. Short-Term Forecasting of Sugarcane Yield Based on Grey Prediction. 2015 4th International Conference on Agro-Geoinformatics; c2015. p. 1-5.