Yield Forecast of Sugarcane Using Two Different Techniques in Discriminant Function Analysis

Ali J. Ramadhan[1*], *S. R.* Krishna Priya[2], *R* Keerti Balambiga[2], *Ali* J. Othman[3], *Shikha* Yadav[4], *Pradeep* Mishra[5], *Mostafa* Abotaleb[6], *Hussein* Alkattan[6] and *Zainalabideen* Albadran[1]

[1]University of Alkafeel, Najaf, Iraq
[2]PSG College of Arts and Science, Tamil Nadu, India
[3]Plekhanov Russian University of Economics, Moscow, Russia
[4]University of Delhi, New Delhi, India
[5]JNKVV College of Agriculture, Rewa, India
[6]South Ural State University, Chelyabinsk, Russia

Abstract. The present study aims to develop yield forecast models for the Sugarcane crop of the Coimbatore district in Tamilnadu using two different techniques namely Variables and Months in Discriminant function analysis. For this, the Sugarcane yield data for 57 years along with the monthly data on seven weather variables have been taken. For applying discriminant analysis, the yield data of sugarcane has been divided into two categories namely two groups and three groups. The discriminant scores from the two and three-group discriminant functions were employed as independent variables in the development of yield forecast models. The yield forecast models for both strategies were created utilizing scores and trend values as independent variables. The first 52 years of yield data (1960-2012) were used to create the model, and the last five years of data (2012-2016) were used for validation. The comparison has been made between two and three groups for both techniques. The results indicate the technique using the variable-wise method gives better results based on goodness of fit. Among the two categories in the variable-wise method, three groups performed better.

1 Introduction

In India, Sugarcane is one of the important cash crops which takes lesscropped area but contributes more in production. It plays a substantial role in Indian economy and is one of the important crops in earning foreign exchange.It affords raw materials to the sugar industry (like paper, chemicals and cattle feed) as well as provides employment to lakhs of people over the country.Presently sugarcane has been cultivated around 21 percent of the total area and it has 23 percent of the total production in the world.

* Corresponding author: ali.j.r@alkafeel.edu.iq

The demand for sugarcane production is constantly increasing. An enormous international market place for sugarcane maintains the enterprise booming. So, it is vital to develop a yield forecast models to manage social and environmental risks of sugarcane for the crop growers, processors and also food industries due to regularity pressures as well as the expectations of consumers and shareholders for sustainably produced goods.

The forecast models for various crops have been developed by many researchers using different techniques [1] Aditya and Das (2012) developed yield forecast models for wheat using discriminant analysis which states that the model using discriminant scores performed better.[2] Agrawal et al., (2012) aimed to forecast wheat yield using discriminant function analysis and the model gave reliable results two months before harvest.[4] Deepankar et al., (2020) have developed a model using step wise regression for rice yield in Haryana.

[7] Kumari et al., (2016) compared ordinal logistic regression methodwith discriminant function analysis based on discriminant scores and posterior probabilities. [8] Kumar et al., (2021) aimed to predict two varieties rice yield using principal component analysis. [9] Priya et al., (2016) tested yield data of sugarcane by fitting linear, quadratic and cubic models through discriminant analysis and it stated that quadratic model performed better.[10] Sanjeev et al., (2018) developed a forecast model by computing eigen values and eigen vectors using principal component analysis.[11]Pandey et al., (2017) developed rice yield forecast models through discriminant scores using discriminant function analysis.[12]Yadav et al., (2016) found reliable yield forecast models for pigeon pea using weather indices and discriminant scores. [13] Poonam et al., (2017) conducted a study on wheat using yield data and seven weather parameters and the yield forecast models were developed by stepwise regression method. [14] Priya and Suresh (2009) conducted a study on yield data along with daily weather data on sugarcane using weather indices approach and the developed model was in good accuracy. [15] Kandiannan et al., (2002) performed multiple regression to forecast yield of turmeric.[16] Kartika et al., (2016) developed a model using artificial neural network for oil palm. The study reveals that the model had less error value. [17] Ajithkumar et al., (2021) used discriminant scores and probabilities to develop a yield forecast models in discriminant analysis. [18] Garde et al., (2015) developed yield forecast model for wheat at Varanasi district of Eastern Uttar Pradesh and [19] Kumari et al., (2019) developed model to forecast wheat yield using Bayesian discriminant analysis in Kanpur, Uttar Pradesh. [20] Nain et al., (2021) compared principal component analysis and discriminant analysis of rice yield using yield data and it was found that model based on discriminant function performed better.

The present study has been carried out to forecast Sugarcane yield of Coimbatore district using discriminantscores calculated using two different techniques namely variable wise method and month wise method. Yield forecast models were developed using these discriminant scores as independent variables and yield as dependent variable.

2 Materials and Methods

The present study deals with yield data (tonnes / hectare) of 57 years (1960-2016) which have been collected from various volumes of annual Season and Crop report published by State Government of Tamilnadu and monthly data on seven weather variables have been collected from Agro-Climatic Research Centre, Tamilnadu Agricultural University, Coimbatore and the corresponding weather variables are X_1 – Maximum temperature, X_2 – Minimum temperature, X_3 – Relative Humidity at 7hrs, X_4 - Relative Humidity at 14hrs, X_5 – Evaporation, X_6 – Rainfall, X_7 – Sunshine hours.

In this study, the yield data for 52 years from 1960-2011 have been used for model development and remaining five years data from 2012-2016 have been used for validation purpose.

3 Methodology

3.1 Discriminant Function Analysis in case of Two Groups

Consider a linear function of the form

$$Z = \sum_{i=1}^{p} l_i\, X_i \qquad (1)$$

Here Z is the discriminant score, X_i is the i^{th}weather variable and l_i is thediscriminant coefficient of corresponding weather variable, $p$is weather variables.

Now,l_i is calculated as the variation between two groups is maximized relative to the variation within the groups. Maximizing the ratio getsfollowing simultaneous equations,

$$l_1 \sum X_1^2 + \ l_2 \sum X_1\, X_2 + \ldots\ldots\ldots\ldots + l_p \sum X_1 X_p = d_1 \qquad (2)$$

$$l_1 \sum X_1 X_p + l_2 \sum X_2 X_p + \ldots\ldots\ldots\ldots + l_p \sum X_p^2 = d_p \qquad (3)$$

where $X_1, X_2 \ldots\ldots X_p$ are deviations varied from its corresponding group means. It can be written in matrix form as,

$S\, l = d.$

where the discriminant dispersion matrix S = $\begin{pmatrix} \sum x_1^2 & \sum x_1\, x_2 & \ldots\ldots\ldots & \sum x_1\, x_p \\ \ldots\ldots & \sum x_2^2 & \ldots\ldots\ldots & \sum x_1\, x_p \\ \ldots\ldots & \ldots\ldots. & \ldots\ldots\ldots & \ldots\ldots\ldots \\ \ldots\ldots & \ldots\ldots. & \ldots\ldots\ldots & \sum x_p^2 \end{pmatrix}$

= (S_{ij})

where S_{ij}=value respected to the i^{th} row and j^{th}column of the matrix S. So, the value *l*can be written as,

$l = \frac{1}{S} d.$

2.2.2Discriminant function analysis in the case of three groups

Consider a linear function which is of the form

$$Z = \sum_{i=1}^{p} l_i\, X_i = l'\mathrm{X} \qquad (4)$$

where Z is the discriminant function, $l' = (l_1,\ l_2, \ldots\ldots l_p)$, $x' = (x_{1,}\, x_{2,} \ldots\ldots x_p)$, x_i is the i^{th}weather variable, l_i is the discriminant coefficient of corresponding weather variable and p is the no. of. weather variables taken for the study.The mean of j^{th} group for i^{th} variable is $\overline{x_{ij}} = \frac{1}{n_j} \sum_{m=1}^{n_j} x_{ijm}$ and overall average for i^{th} variable is,

$$\overline{x_i} = \frac{\sum_{j=1}^{k} n_j \overline{x_{ij}}}{\sum_{j=1}^{k} n_j} = \frac{\sum_{j=1}^{k} \sum_{m=1}^{n_j} x_{ijm}}{\sum_{j=1}^{k} n_j} \qquad (5)$$

Let $\overline{x'} = (\ \overline{x_1}, \overline{x_2}, \ldots\ldots \overline{x_p}\ )$, $\overline{x_j'} = (\ \overline{x_{1j}}, \overline{x_{2j}}, \ldots\ldots \overline{x_{pj}}\ )$ and $\overline{x_i'} = (\ \overline{x_{i1}}, \overline{x_{i2}}, \ldots\ldots \overline{x_{ik}}\ )$

Let $M = \sum_{j=1}^{k} (\bar{x}_j - \bar{x})\,(\bar{x}_j - \bar{x})'$ be between a group matrix of the sum of squares and a cross products and let $N = \sum_{j=1}^{k} \sum_{m=1}^{n_j} (\bar{x}_{jm} - \bar{x}_j)(\bar{x}_{jm} - \bar{x}_j)'$ be the pooled matrix of corrected sum of squares and the sum of products. The *l*can be calculatedas $\frac{l'Bl}{l'Al}$ and substitute the value of *l* in equation (2.2.2) the scores are obtained.

3.2 Formation of Two and Three Groups

The yield data (1960-2011) of Sugarcane crop has been divided into two categories namely two and three groups. In two groups, the yield data of 52 years has been categorized into groups namely low (1) and high (2). For formation of two groups, regression analysis has been carried out between yield as dependent variable and year as independent variable. Among the residuals, the negative values were taken as years with low yield and positive residuals were taken as years with high yield Similarly, in three groups the yield data has been formed into categories viz. low (1), medium (2) and high (3).For formation of three groups, the residual values have been arranged in ascending order and were divided into three equal groups namely low, medium and high. The categorical yield data in two/three groups have been used a dependent variable in carrying out discriminant analysis.

3.3 Development of the Model

3.3.1 Technique 1: Variable WiseMethod

For applying discriminant function analysis iterative procedure has been followed using categorical yield data and monthly data on seven weather variables. For two group discriminant analysis,considering the first weather variable,maximum temperature for twelve months period, the analysis has been carried out and the discriminant score of first weather variable has been generated. For two group discriminant analysis one discriminant score is generated. Now, taking the minimum temperature (second weather variable)for twelve months into consideration along with discriminant score of first weather variable(here the number of discriminating variables is thirteen),discriminant analysis has been carried outand the discriminant score for second variable has been generated. Again, taking third weather variable (relative humidity at 7 hours) and score of a second weather variable the procedure has been carried out and score for the corresponding weather variable has been found. The iterative procedure has been repeated till the seventh weather variable which is sunshine hours.For two groups discriminant analysis the total number of scores can be generated is one. For the final weather variable one discriminant score has been found.Weather has its influence in different stages of the crop. Monthly data on seven weather variables for twelve months makes the number of explanatory variables 84 for discriminant analysis. In order to overcome the problem of using number of variable more than the number of data points, variable wise iterative procedure has been followed.

The yield forecast model has been developed using dependent variable as yield and discriminant score along with trend values as independent variables by regression analysis. The same procedure has been followed for three groups discriminant function analysis. In three group discriminant analysis, the total number of scores generated is two. The regression model for yield forecast has been developed using the yield as dependent variableand two discriminant scores along with trend values as independent variables.

3.3.2 Technique 2: Month Wise Method

Considering 52 years of yield data and monthly data on seven weather variables an iterative procedure has been carried out for discriminant analysis. Taking data on seven weather variables of January into consideration the discriminant analysis has been carried out. The one-discriminant score has been found for the month of January. Now, taking seven weather variables of February and discriminant score of January (total number of independent variables are eight) the analysis has been performed and score has been generated. Again, the procedure is repeated for the month of March taking weather data of March along with discriminant score of February and the score has been generated. The procedure has been repeated up to month of December and the discriminant scores has been calculated for the month of December.For developing the yield forecast model, yield has been used as dependent variable and discriminant score as well as trend values as independent variables. The same procedure has been carried out for three groups. In three groups, the discriminant scores have been generated are two.

3.3.3 Yield forecast model

Discriminant analysis is used to forecast the yield qualitatively. For quantitative forecasting of the yield, regression analysis is performed using yield as dependent variable and the independent variables arediscriminant scores obtained from two/ three group and trend values.

For two groups

$$Yield = \beta_0 + \beta_1 ds + \beta_2 T + e \quad (6)$$

Where β_0 is the Intercept.
β_i's are the regression coefficient.
ds denotes discriminant score.
T denotes the trend
e $is\ the\ error \sim N$ $(0, \sigma^2)$
For three groups

$$Yield = \beta_0 + \beta_1 ds_1 + \beta_2 ds_2 + \beta_3 T + e \quad (7)$$

Where β_0 is the Intercept.
β_i's are regression coefficient.
ds_1, ds_2 denotes the two sets of discriminant scores
other terms are same as defined earlier.

When the independent variables are significantly linked to each other, the introduction of meteorological variables as such causes multicollinearity in the yield forecast model. The presence of multicollinearity affects the accuracy of the predicted coefficients, hence weakening the regression model's power. To address the aforementioned shortcomings, the yield forecast model was created by substituting discriminant scores obtained from a discriminant function for meteorological data.

3.4 Validation of the Model

The validation has been done for the subsequent years from 2012-2016. The formula for the validation of the model

$$Percentage\ Deviation = \frac{Actual\ Yield - Predicted\ Yield}{Actual\ Yield} \times 100 \quad (8)$$

3.5 Goodness of Fit

The comparison has been made between two and three groups for both the techniques and the formula for the comparison are listed below

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right| \times 10 \quad (9)$$

where,
MAPE = Mean Absolute Percentage Error
n = number of yield data
A_t = Actual yield
F_t = Forecast yield

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(A_t - F_t)} \quad (10)$$

where,
RMSE = Root Mean Square Error
n = number of yield data
A_t = Actual yield
F_t = Forecast yield

$$MSE = \frac{1}{n}\sum_{t=1}^{n}(A_t - F_t)^2 \quad (11)$$

where,
MSE = Mean Squared Error
n = number of yield data
A_t = Actual yield
F_t = Forecast yield

$$MAD = \frac{1}{n}\sum_{t=1}^{n}|A_t - F_t| \quad (12)$$

where,
MAD = Mean Absolute Deviation

n = number of yield data
A_t = Actual yield
F_t = Forecast yield

4 Results and Discussion

The yield forecast models for sugarcane crop have been fitted using variable wise and month wise methods which are presented in Table 1. The Adj R^2 of all four model are above 50% and among the models fitted, model using variable wise method for three groups has highR^2 value (74.65) which shows that the model is able to explain 74% variations in Sugarcane yield. According to F statistic, which shows the adequacy of the fitted regression model,the modelsusing both the techniques are found to be significant at 1% level.

Table 1. Fitted Regression Equation on Sugarcane Using Variable Wise & Month Wise Method

No	Tech.	Groups	Equation	Adj R^2	F	Sig
1	1	Two	Y= 94.722 +3.502 ds + 0.364 T	0.638	45.879**	.000
2	1	Three	Y= 94.201+1.80 ds_1 -3.226 ds_2 +0.373 T	0.746	51.029**	.000
3	2	Two	Y= 94.086 +3.228 ds + 0.395 T	0.557	33.050**	.000
4	2	Three	Y= 94.498-3.565 ds_1 +1.266 ds_2 +0.363 T	0.671	35.691**	.000

** Significant at 1% level. * Significant at 5% level.

From Table 2 the regression coefficients of all four models are significant at 1% level except discriminant score 2 of model based on month wise method of three groups (significant at 5% level). The high significance of the regression coefficients indicates that the regression coefficients significantly contribute to the Sugarcane yield.

Table 2. Results of t Test

No	Tech.	Groups	Coefficients	T	Sig
1	1	Two	Constant	46.582**	.000
			T	5.399**	.000
			Discriminant Score	6.530**	.000
2	1	Three	Constant	53.986**	.000
			T	6.412**	.000
			Discriminant Score 1	5.087**	.000
			Discriminant Score 2	-7.371**	.000
3	2	Two	Constant	42.029**	.000
			T	5.352**	.000
			Discriminant Score	5.093**	.000
4	2	Three	Constant	48.069**	.000
			T	5.515**	.000
			Discriminant Score 1	-7.029**	.000
			Discriminant Score 2	2.133*	.038

** Significant at 1% level. * Significant at 5% level.

Table 3 shows the goodness of fit values for both the techniques using two and three groups. Among the values, the R^2 value of model based on three groups using variable wise method is high (76.1%) as well as RMSE is less (5.7138).

Table 3. Goodness of Fit

No	Method	Variable Wise Method		Month Wise Method	
		Two Groups	Three Groups	Two Groups	Three Groups
1	R^2	0.652	0.761	0.574	0.690

No	Method	Variable Wise Method		Month Wise Method	
2	MSE	47.6115	32.6471	58.2249	42.3342
3	RMSE	6.9001	5.7138	7.6305	6.5065
4	MAPE	5.2641	4.2486	5.1660	4.3655
5	MAD	5.5230	4.4888	5.4384	4.6081

From Table 4, shows the percentage of deviation values. The results indicate thatvalues lie between 1 to 25 for two groups and ranges between 2 to 20 for three groups using variable wise method.

Table 4. Percent Deviation of Variable Wise Method

Year	Actual Yield	Predicted Yield		% Deviation	
		Two Groups	Three Groups	Two Groups	Three Groups
2012	113	117.84	115.35	4.28	2.08
2013	104	109.80	111.56	5.58	7.27
2014	115.6	116.80	111.92	1.04	3.18
2015	101	111.83	107.14	10.72	6.08
2016	87	108.41	103.83	24.61	19.34

The values deviated between ± 28 for two groups and ± 19 for three groups based on month wise method are presented in Table 5.

Table 5. Percent Deviation of Month Wise Method

Year	Actual Yield	Predicted Yield		% Deviation	
		Two Groups	Three Groups	Two Groups	Three Groups
2012	113	120.06	112.83	-6.25	0.15
2013	104	113.85	116.08	-9.47	-11.62
2014	115.6	120.08	112.19	-3.88	2.95
2015	101	115.32	106.26	-14.18	-5.21
2016	87	110.72	103.73	-27.26	-19.23

5 Conclusion

Sugarcane is one of the important cashcrops in Indian economy and it provides large amount of employment to the people which helps to improve the living standard of the people. Timely, forecasting of crop yield helps farmers to make timely decision to avoid losses as well as improve the production. From the fitted models, it is concluded that based on Adj R^2, RMSE and percent deviation, the model for three groups using variable wise method has performed better and is recommended for forecasting sugarcane yield in Coimbatore district, Tamilnadu.

References

1. K. Aditya, S. Das, Crop Yield Forecasting using Discriminant Function Analysis, LAP LAMBERT Academic Publishing.
2. R. Agrawal, Chandrahas, K. Aditya, Use of discriminant function analysis for forecasting crop yield, Mausam **63**(3) (2012)455-458.
3. Al-Mahdawi, H. K., Albadran, Z., Alkattan, H., Abotaleb, M., Alakkari, K., & Ramadhan, A. J. (2023, December). Using the inverse Cauchy problem of the Laplace equation for wave propagation to implement a numerical regularization homotopy method. AIP Conference Proceedings (Vol. **2977**, No. 1). AIP Publishing.
4. Deepankar, D.R. Aneja, A. Kumar, Impact of weather parameters on rice yield in Karnal district, Journal of Pharmacognosyand Phytochemistry, **9**(6) (2020) 633-636.

5. Hair, F. Joseph, Jr. Anderson, E. Rolph, L. Ronld and C. William, Multivariate Data Analysis with Readings, PRENTICE HALL, New Jersey, 1995.
6. R.A. Johnson, D.W. Wichern, Applied multivariate statistical analysis, Pearson Education, 2006.
7. V. Kumari, R. Agrawal, A. Kumar, Use of ordinal logistic regression in crop yield forecasting, Mausam, **67**(4) (2016) 913- 918.
8. Al-Nuaimi, B. T., Al-Mahdawi, H. K., Albadran, Z., Alkattan, H., Abotaleb, M., & El-kenawy, E. S. M. (2023). Solving of the inverse boundary value problem for the heat conduction equation in two intervals of time. Algorithms, **16**(1), 33.
9. S.R.K. Priya, K.K. Suresh, P.K. Bajpai, Multivariate technique in sugarcane yield forecast, Indian Journal of Sugarcane Technology, **31**(1) (2016) 29-35.
10. Sanjeev, P. Verma and U. Verma, Development of Weather Based Wheat Yield Forecast Models in Haryana, Int.J.CurrMicrobiol.App.Sci., **7**(12) (2018) 2973-2978.
11. K.K. Pandey, B.V.S. Sisodia, V.N. Rai, Forecasting by Discriminant Function Weather Based Analysis, International Journal of Environmental Sciences & Natural Resources, **4**(4) (2017).
12. Akbari, E., Mollajafari, M., Al-Khafaji, H. M. R., Alkattan, H., Abotaleb, M., Eslami, M., & Palani, S. (2022). Improved salp swarm optimization algorithm for damping controller design for multimachine power system. IEEE Access, **10**, 82910-82922.
13. Poonam, D.R. Aneja, M.L. Khichar, Mamta, N. Tanwar, Statistical models for wheat yield forecasting using weather Parameters for Western agroclimatic zone of Haryana, Bull.Env.Pharmacol.Life.Sci, **6**(4) (2017) 422-427.
14. Ehsan khodadadi, S. K. Towfek, Hussein Alkattan. (2023). Brain Tumor Classification Using Convolutional Neural Network and Feature Extraction. Fusion: Practice and Applications, **13**(2), 34-41.
15. K. Kandiannan, K.K. Chandragiri, N. Sankaran, Balasubramanian, C. Kailasam, Crop-weather model for turmeric yield forecasting for Coimbatore district, Tamilnadu, India, Agricultural and Forest Meteorology, **112** (2002) 133-137.
16. N.D. Karthika, I.w.Astika, E. Santosa, Oil Palm Yield Forecasting Based on Weather Variables using Artificial Neural Network, Indonesian Journal of Electrical Engineering and Computer Science, **3**(3) (2016) 626-633.
17. B. Ajithkumar, V. Harithalekshmi, A. Vysakh, Rice yield forecasting using principal component regression and composite weather variables, Journal of Pharmacognosy and Phytochemistry, **10**(2) (2021) 595-600.
18. Y.A. Garde, B.S. Dhekale, S. Singh, Different approaches on pre-harvest for forecasting of wheat yield, Journal of Applied and Natural Science, **7**(2) (2015) 839-843.
19. V. Kumari, K. Aditya, H. Chandra, A. Kumar, Bayesian discriminant function analysis-based forecasting of crop yield in Kanpur district of Uttar Pradesh, Journal of Agrometeorology, **21**(4) (2019) 462-467.
20. G. Nain, N. Bhardwaj, P.K.M. Jaslam, C.H. Dagar, Anurag, Rice yield forecasting using agro-meteorological variables: A multivariate approach, Journal of Agrometeorology, **23**(1) (2021) 100-105.