Smart Agricultural Technology 3 (2023) 100076

Contents lists available at ScienceDirect

Smart Agricultural Technology

journal homepage: www.journals.elsevier.com/smart-agricultural-technology

Sugarcane yield forecast using weather based discriminant analysis

S R Krishan Priya [a,*], R Keerti Balambiga [b], Pradeep Mishra [c], Soumitra Sankar Das [d]

[a] Assistant Professor, Department of Statistics, PSG College of Arts & Science,Coimbatore-14, India
[b] Research Scholar, Department of Statistics, PSG College of Arts & Science,Coimbatore-14, India
[c] Assistant Professor (Statistics), College of Agriculture, Powarkheda,J.N.K.V.V. Madhya Pradesh, India
[d] Assistant Professor(Statistics), Birsa Agricultural University, Kanke, Ranchi, Jharkhand, India

ARTICLE INFO

Keywords:
Discriminant function analysis
Discriminant scores
Forecast models

ABSTRACT

Discriminant function analysis has been used for forecasting of Sugarcane yield of Coimbatore district in Tamilnadu. Crop yield has been classified into two and three groups. Using this crop yield and monthly data on weather variables,discriminant function analysis has been carried out. The scores calculated from function this along with the trend have used as regressors in developing yield forecast models. The forecast models based on two groups and groups have been compared. The results show that the forecast models using three groupsare considered to be better.

Introduction

Sugarcane is amajor commercial crop in the agricultural map of India. Sugar Industryforms the basisof incomefor a huge sector of population. Sugarcane being a prominent cash crop earns a considerable quantity of foreign revenue.Forecasts of such major commercial crop areof great significance to trade and industry. Sugarcane yield forecasts will help policy makers in making decisions onprice fixation, distribution, and for storage and marketing. For development of yield prediction models, researchers' have adopted different techniques such as models based on weather indices [1–3], principal component analysis [4,5,30, 31], neural networks [6,7], logistic regression [8, 9], time series models (Yonar et.al, [10] in wheat and Mishra et.al, [28,29], in Sugarcane and pulses).

Apart from these techniques,yield forecasting based on discriminant functionwas done by various authors [11–19],[28,29,31]. Kocisova and Misankova [20] have used discriminant analysis to predict the company's financial status. Garde et al., [21] found that multiple linear regression with additional variables was found better than discriminant function analysis for forecasting wheat yield. Goyal and Verma [22] have combined spectral data along with weather parameters for forecasting wheat yield of Haryana. Goyal [23] has compared different multivariate techniques like multiple linear regression, discriminant, and principal component analysis for estimating the wheat yields of Haryana. Kumari et al. [24] have compared Bayesian discriminant analysis with score-based discriminant function approach.Studies in the past using discriminant analysis were carried using weather indices. In the present study efforts has been made to develop a suitablemodel for forecasting sugarcane yield of Coimbatore district.The methodology suggested by Pandey et al. [17] has been carried out along with suitable modifications.

Materials and methods

Description of the data

The present study has been conducted in Coimbatore district of Tamilnadu. The weather variables used in the study are X_1 – Maximum temperature, X_2 – Minimum temperature, X_3 – Relative Humidity 7 hrs, X_4 - Relative Humidity, X_5 – Evaporation, X_6 – Rainfall, X_7 – Sunshine hours.Sugarcane yield data (tonnes / hectare) for a period of 57years (1960–2016) have been used as dependant variable in developing the model.

The sugarcane yield figures have been collected from different volumes of annual Season and Crop report issued by State Government of Tamilnadu. Monthly Data on weather variable have been collected from agro-climate research center, Tamil Nadu Agricultural University, Coimbatore. Data for the period of 52 years (1960–2011) have been used in model building and the remaining five years (2012–2016) data has been used for validation.

* Corresponding author.
E-mail address: krishnapriya@psgcas.ac.in (S.R.K. Priya).

https://doi.org/10.1016/j.atech.2022.100076

Available online 3 June 2022

Methodology

The theory of discriminant analysis is available in standard books on multivariate analysis [25, 26].This technique is used for identifying appropriate functions that discriminate best between set of observations from two or more groups and classifying future observations into one of the previously defined groups. Suppose observations are to be classified into k groups on the basis of p variables. The technique involves identifying linear/quadratic function(s) where the coefficients are determined in such a way that the variation between the groups gets maximized relative to the variation within the groups. The maximum number of discriminant functions that can be obtained is equal to minimum of k-1 and p. These functions are used to calculate discriminant scores, which are used to classify the observations into different groups.

Formation of two/three groups

For the formation of two groups,linear regression equation has been fitted between sugarcane yield and year.Residuals were calculated through the regression equation. Crop yield has been classified into two groups' namely low yield (0) and high yield (1), taking negative residuals as valuesfor low yield and positive residuals as high yield. For three groups, Crop years were grouped into three namely low (0), average (1) and high (2). Residuals were formed into three groups after arranging them in ascending order.Using weather variables in these two/three groups, scores were obtained by discriminant function analysis. Thesediscriminant scores along with trend as regressors have been used for developing forecast models.

Development of the model

The discriminant function analysis has been performed making use of the data on the first weather variable spread over 12 months for different years.For the second weather variable discriminant analysis has been again carried out againalong with the scores obtained from the discriminant function for the first weather variable. Discriminant analysis has been again repeated using the above scores a monthly data of the third weather variable. Discriminant scores have been obtained up to the final weather variable using the iterative procedure. It is, however, known that weather variables affect the crop differently during different phases of crop development. Therefore, using monthly weather data as such in developing the model poses a problem as no. of independent variables in the regression model would increase enormously. In order toover come the difficulty, the iterative procedure suggested above has been used.

Forecast models have been developed using the discriminant scores of the final weather variable along with trend as regressors and yield as dependent variable.Yield forecast models are built using these discriminant scores as independent variable in place of weather variables thus reducing number of variables to avoid multicollinearity.

Forecast model

Discriminant function analysis is only capable of predicting the yieldquantitatively. For quantitative yield prediction, regression models have been fitted by taking discriminant scores along with trend (year) as regressors.

For two groups

$$Yield = \beta_0 + \beta_1 ds + \beta_2 T + e \ldots\ldots\ldots\ldots$$

Where β_0 is the Intercept.

β_i's are the regression coefficient. ds denotes discriminant score.

T denotes the trend

Table 1
Fitted regression equation for two groups.

Equation	Adj R^2
$Y = 96.350 + 3.890ds^{**} + 0.281T^{**}$	0.638

** Significant at 1% level.

Table 2
Results of regression equation for three groups.

Equation	Adj R^2
$Y = 95.045 + 1.834ds_1^{**} - 3.593ds_2^{**} + 0.326T^{**}$	0.746

** Significant at 1% level.

e *is the error* $\sim N\ (0, \sigma^2)$

For three groups

$$Yield = \beta_0 + \beta_1 ds_1 + \beta_2 ds_2 + \beta_3 T + e \ldots\ldots\ldots\ldots$$

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.

The use of weather variables as such involves multi-collinearity among the variables which would inflate the variances of regression coefficients. Multi-collinearity in the data causes serious problems in estimation, prediction and interpretation. Further the estimated regression coefficient may be unrealistic in magnitude or sign. To overcome the above drawbacks in the present study, the yield forecast model has been developed using discriminant scores in the place of weather variables in order to avoid multi-collinearity (Devi et.al, [27]).

Results & discussion

modeling with two groups

Sugarcane yield data has been classified into two groups based on crop yield taking negative residuals as low yield (0) and positive residuals as high yield (1). From the 57 years of yield data low yield values are 28 and the high yield values are 29. Monthly data on weather variables corresponding to the two predefined groups have been used for the development of the scores for each year through discriminant function analysis. The discriminant scores have been developed through iterative procedure. The forecast model for two groups has been fitted using yield as the dependent variable. The discriminant score (ds) obtained from the two-group model and trend have used as regressors. The results are presented in Table 1 along with the adjusted R^2 values.

Forecast model with three groups

In this case residuals of 57 years of yield data have been divided into three groups namely low (0), average (1) and high (2) and each group consists of 19 years. Among the weather variable in three groups, the same iterative procedure is carried out but the only difference is that we get two scores in each analysis. The two final discriminant scores (ds_1, ds_2) along with trend are used for fitting the forecast model. The forecast equation for two three group model is presented in Table 2. From Table 1, the result of coefficient of determination indicates that the independent variables are able to explain 63.8% variations in the yield. Both the regression coefficients discriminant score and trend are significant at 1% percent level which indicate that both the variables significantly contribute for the yield.The Table 2 indicates R^2 value is 0.746 which indicates that 74.6% of total variance in the dependent variable is influenced by the independent variable and all the regression

Table 3
Goodness of Fit values.

S. No	Method	Two Group	Three Group
1	R^2	0.652	0.761
2	MAD	5.5230	4.4888
3	MSE	47.6115	32.6471
4	RMSE	6.9001	5.7138
5	MAPE	5.2641	4.2486

Table 4
Predicted Values.

Year	Actual	Predicted
2012	113	115.35
2013	104	111.56
2014	115.6	111.92
2015	101	107.14
2016	87	103.83

coefficients namely First discriminant score (ds_1), Second discriminant score (ds_2) and the trend variable are highly significant.

Comparison of the model

The best model for yield forecast of sugarcane has been identified based on the high R^2 and low error values namely Mean Absolute Deviation (MAD), Mean Square Error (MSE), Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE). The goodness of fit measures are presented in the table below.From the Table 3, the R^2 value is high for three group discriminant model compared with thetwo group model. For the three group model the error values MAD, MSE, RMSE and MAPE are low compared to the two-group forecast model. Based on the above values forecast model based on three groups is most suitable for forecasting sugarcane yield. The predicted values are calculated based on three group and presented in the table below.

Conclusion

Sugar factoriesare a source of jobs for persons living within the area. Production of commercial crops (like sugarcane) forms the raw materialfor the industries at local and national level. Crop yield forecasting helps the farmers to know the potential of farming system and to prevent future losses. From the forecast model using discriminant scores for two groups and three groupsit is concluded that the performance of three group model is better.And it is recommended for forecasting the sugarcane yield in Coimbatore district of Tamilnadu.

Declarations

Funding

Not applicable

Availability of data and material

Sugar Yield data has been collected from various volumes of annual Season and Crop Report issued by the Department of Economics and Statistics, State Government of Tamilnadu. Data on weather variable has been collected from Agro climate research center, Tamil Nadu Agricultural University, Coimbatore.

Code availability

Not applicable

Ethics approval

Not applicable

Consent to publication

We the authors of this paper entitled" Sugarcane Yield Forecast using Weather based Discriminant Analysis" give consent for publishing this article. Also, we hereby declare that the content of this article has not been published in any form.

Declaration of Competing Interest

Not applicable

References

[1] N. Kumar, R.R. Pisal, S.P. Shukla, K.K. Pandey, Crop Yield forecasting of Paddy and Sugarcane through modified Hendrick and Scholl technique for South Gujarat, Mausam 67 (2) (2016) 405–410.
[2] K.B. Banakara, H.R. Pandy, Y.A. Garde, S. Ojha, Pre-harvest modeling of Kharif rice using weather parameters in Valsad district of south Gujarat, J. Agrometeorol. 19 (2017) 196–199.
[3] M. Rajavel, P. Khare, J.R. Prasad, K.K. Singh, H.V. Puranik, G.K. Das, Development of rice yield forecast in mid-season using weather indices based agrometerological model in Chhattisgarh, Vayumandal 44 (1) (2018) 38–45.
[4] R.R. Yadav, B.V.S. Sisodia, S. Kumar, Application of principal component analysis in developing statistical models to forecast crop yield using weather variables, Mausam 65 (3) (2014) 357–360.
[5] S.R.K. Priya, K.K. Suresh, P.K. Bajpai, Multivaraite technique in sugarcane yield forecast, Indian J. Sugarcane Technol. 31 (1) (2016) 29–35.
[6] R.R. Laxmi, A. Kumar, Weather based forecasting model for crop yield using neural network approach, Stat. Appl. 9 (1&2) (2011) 55–69.
[7] G. Niedbala, R.J. Kozlowski, Application of artificial neural networks for multi-criteria yield prediction of winter wheat, J. Agric. Sci. Technol. 21 (2019) 51–61.
[8] V. Kumari, A. Kumar, Forecasting of wheat (Triticum aestivum) yield using ordinal logistic regression, Indian J. Agric. Sci. 84 (6) (2014) 691–694.
[9] V. Kumari, R. Agrawal, A Kumar, Use of ordinal logistic regression in crop yield forecasting, Mausam 67 (4) (2016) 913–918.
[10] A. Yonar, H. Yonar, P. Mishra, B. Kumari, M. Abotaleb, Amr Badr, Modeling and forecasting of wheat of South Asian region countries and role in food security, Adv. Computational Intell. 1 (2021). Doi:0.1007/s43674-021-00027.
[11] T. Rai, Chandrahas, Use of Discriminant Function of Weather Parameters For Developing Forecast Model of Rice Crop, IASRI Publication, New Delhi, 2000.
[12] R. Agrawal, Chandrahas, K Aditya, Use of discriminant function analysis for forecasting crop yield, Mausam 63 (3) (2012) 455–458.
[13] S.R.K. Priya, P.K. Bajpai, K.K. Suresh, Use of data reduction technique for sugarcane yield forecast, Indian J. Sugarcane Technol. 29 (2) (2014) 77–80.
[14] B.V.S. Sisodia, R.R. Yadav, S. Kumar, M.K Sharma, Forecasting of pre-harvest crop yield using discriminant function analysis of meteorological parameters, J. Agrometeorol. 16 (1) (2014) 121–125.
[15] M. Goyal, U. Verma, Zonal trend-agro meteorological models for wheat yield estimation in Haryana, J. Appl. Nat. Sci. 8 (3) (2016) 1485–1492.
[16] R.R. Yadav, B.V.S. Sisodia, S. Kumar, Pre-harvest forecast of pigeon-pea yield using discriminant function analysis of weather variables, Mausam 67 (3) (2016) 577–582.
[17] K.K. Pandey, B.V.S. Sisodia, V.N. Rai, Forecasting by discriminant function weather based analysis, Int. J. Environ. Sci. Nat. Res. 4 (4) (2017) 1–4.
[18] H.G. Singh, V.N. Rai, B.V.S. Sisodia, V. Nand, H.Y Singh, Forecasting of pre-harvest crop yield using discriminant function analysis of meteorological parameters, Res. Environ. Life Sci. 10 (1) (2017) 11–14.
[19] S. Kumar, V.N. Rai, K.K. Mourya, Annu Gupta, Forecasting of pre-harvest rapeseed and mustard yield using discriminant function analysis of meteorological parameters, Int. J. Chem. Stud. 7 (3) (2019) 1897–1900.
[20] K. Kocisova, M. Misankova, Discriminant analysis as a tool for forecasting company's financial health, Procedia-Social and Behavioral Sciences 110 (2014) 1148–1157.
[21] Y.A. Garde, B.S. Dhekale, S. Singh, Different approaches on pre harvest forecasting of wheat yield, J. Appl. Nat. Sci. 7 (2) (2015) 839–843.
[22] M. Goyal, U. Verma, Spectral-weather –crop yield forecasting: discriminant function analysis, J. Appl. Probab. Stat. 10 (1) (2015) 1–14.
[23] M. Goyal, Use of different multivariate techniques for pre-harvest wheat yield estimation in Hisar (Haryana), Int. J. Comput. Mathe. Sci. 5 (12) (2016) 36–40.
[24] V. Kumari, K. Aditya, Baycsian discriminant function analysis based forecasting of crop yield in Kanpur district of Utter Pradesh, J. Agrometeorol. 21 (4) (2019) 462–467.
[25] T.W. Anderson, An Introduction to Multivariate Statistical Analysis, John Wiley & Sons, USA, 1998.
[26] Johnson R.A., Wichern, D.W., (2006). Applied Multivariate Statistical Analysis. Pearson Education.

[27] M. Devi, J. Kumar, D.P. Malik, P. Mishra, Forecasting of wheat production in Haryana using hybrid time series model, J. Agric. Food Res. (2021), 100175.
[28] P. Mishra, A. Yonar, H. Yonar, B. Kumari, M. Abotaleb, S.S. Das, S.G. Patil, State of the art in total pulse production in major states of India using ARIMA techniques, Curr. Res. Food Sci. 4 (2021) 800–806, https://doi.org/10.1016/j.crfs.2021.10.009.
[29] P. Mishra, M.G. Al Khatib, I Sardar, et al., Modeling and forecasting of sugarcane production in India, Sugar Tech. (2021), https://doi.org/10.1007/s12355-021-01004-3.
[30] Sharma, S., (1999). Applied Multivariate Techniques. John Wiley & Sons.
[31] Annu Sisodia, An application of principal component analysis for pre-harvest forecast model for wheat crop based on biometrical characters, Int. Res. J. Agric. Econ. Stat. 8 (1) (2017) 83–87.