

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2024; 9(5): 59-63
© 2024 Stats & Maths
<https://www.mathsjournal.com>
Received: 01-08-2024
Accepted: 05-09-2024

Anita Yadav
Research Scholar, Department of
Mathematics, Statistics and
Computer Science, College of
Basic Sciences & Humanities,
GBPUA & T, Pantnagar,
Udham Singh Nagar,
Uttarakhand, India

Dr. AK Shukla
Professor, Department of
Mathematics, Statistics and
Computer Science, College of
Basic Sciences & Humanities,
GBPUA & T, Pantnagar,
Udham Singh Nagar,
Uttarakhand, India

Corresponding Author:
Anita Yadav
Research Scholar, Department of
Mathematics, Statistics and
Computer Science, College of
Basic Sciences & Humanities,
GBPUA & T, Pantnagar,
Udham Singh Nagar,
Uttarakhand, India

Maize crop yield forecast using discriminant function analysis of weather parameters for Udham Singh Nagar district of Uttarakhand

Anita Yadav and Dr. AK Shukla

DOI: <https://dx.doi.org/10.22271/math.2024.v9.i5a.1811>

Abstract

A discriminant function analysis was employed to forecast maize yield in Udham Singh Nagar district of Uttarakhand. Maize crop yield data for a period of 21 years (2001-2021) were drawn from the Dacnet website and the weather data were collected from the Meteorological Observatory, Department of Agrometeorology, College of Agriculture, G.B. Pant University of Agriculture and Technology Pantnagar, Uttarakhand. Maximum temperature, Minimum temperature, Relative Humidity A.M, Relative Humidity P.M, Total rainfall, Sunshine hours, Wind velocity and Evapotranspiration were the weather parameters considered for the study. Out of the 21-year data, 17-year data were used for training the model while remaining 4 years data were used for testing the model. Weekly data on weather variables was used to create weather indices (Agrawal *et al.*, 1983). Crop yield has been classified into three groups: congenial, normal and adverse. Taking these three groups as three populations, discriminant function analysis has been carried out. The resulting discriminant scores were used as regressors, along with time trend, to develop statistical models. Five models were generated using various strategies for incorporating weather indices. Model 4, which included all calculated weighted weather indices, emerged as the best-suited model, with the highest adjusted R-squared value and lowest root mean square error (RMSE). Therefore, this model is recommended for forecasting maize yield in the studied region.

Keywords: Discriminant function analysis, discriminant score, forecast models, maize crop yield, parameters

Introduction

Maize is a vital crop for food security and sustainable agriculture, with over 90% of global production used for food, feed, and biofuel (FAO, 2020) ^[1]. Accurate yield prediction is crucial for farmers, policymakers, and stakeholders to make informed decisions. Weather variables significantly impact maize yield, but analyzing multiple weather variables can be challenging.

Climate change is one of the most pressing issues of our time, with far-reaching consequences for global food security, sustainable agriculture, and ecosystem resilience. Maize, a staple crop for over 50% of the global population, is particularly vulnerable to climate change impacts. Rising temperatures, changing precipitation patterns, and increased frequency of extreme weather events can significantly affect maize yields, leading to reduced food availability, economic losses, and social instability. Weather variables significantly impact maize yield, but analyzing multiple weather variables can be challenging.

Understanding the relationships between climate change and maize yields is crucial for developing effective mitigation and adaptation measures. This research aims to contribute to the existing body of knowledge on climate change and agriculture, providing insights for policymakers, researchers, and practitioners to ensure food security and sustainable agriculture in the face of climate change.

Researchers have previously attempted to develop statistical models for predicting agricultural yield prior to harvest, utilizing time series data on crop yield and meteorological variables.

Early attempts by Fisher (1924) ^[2], Hendricks and Scholl (1943) ^[3], Agrawal *et al.* (1980) ^[4], Agrawal *et al.* (1983) ^[5], Agrawal *et al.* (1986) ^[6], and Agrawal *et al.* (2001) ^[7], and Jain and Singh (1980) ^[8] employed regression models. Later, Rai and Chandras (2000) ^[9], Agrawal *et al.* (2012) ^[10], Sisodia *et al.* (2014) ^[11], Azfar *et al.* (2014) ^[12], Azfar *et al.* (2015) ^[13], and Yadav *et al.* (2016) ^[14] explored the use of discriminant function analysis, incorporating weather indices and weekly weather data, to forecast agricultural yield. Notably, the application of discriminant function analysis has yielded promising results, demonstrating its potential as a valuable tool for agricultural yield prediction.

In this research, we utilized discriminant function analysis (DFA) to investigate the relationship between weather indices and maize yields in the Udham Singh Nagar district of Uttarakhand. By applying DFA to various weather variables, our objective was to identify the most critical weather factors that impact maize yield and develop robust statistical models for predicting yields in this region, thereby enabling data-driven decision-making for agricultural planning and management.

Materials and Methods

Description of the study area

The study area for constructing Yield forecast models encompasses the Udham Singh Nagar district in Uttarakhand state, particularly focusing on kharif maize crop yield data (kg/ha) and corresponding weather data. The study area located in the Kumaon Division's Terai region, this district spans across 78° 45' E to 80° 08' E longitude and 28° 53' N to 29° 23' N latitude. The region experiences a sub-tropical to sub-humid climate, characterized by three distinct seasons: summer, monsoon, and winter. The soil profile is predominantly shallow, with a texture ranging from sandy to loamy.

Description of the data

Time series data of maize crop yield spanning 21-year (2001-2021) was obtained from the Dacnet website, while weather data for the same period was sourced from the Meteorological Observatory, Department of Agrometeorology, College of Agriculture, G.B. Pant University of Agriculture and Technology, Pantnagar, Uttarakhand. The study considered eight key weather parameters: Maximum temperature, Minimum temperature, Morning relative humidity (A.M.), Afternoon relative humidity (P.M.), Total rainfall, Sunshine hours, Wind velocity and Evapotranspiration.

Software used

The analysis and modeling in this study were carried out using two software tool: SPSS (Statistical Package for the social Sciences) and Microsoft Excel.

Statistical procedure

Out of the 21-year data, 17-year data were used for training the model while remaining 4 years data were used for testing the model. Weekly data on weather variables was used to create weather indices (Agrawal *et al.*, 1983) ^[5]. In this study five models were created with discriminant function analysis as independent variables which also include time trend and maize yield as dependent variable.

Discriminant analysis is a multivariate technique concerned with separating distinct sets of objects (or sets of observations) and allocating new objects (or observations) to the previously defined groups. The technique of discriminant

function analysis is used to identify an appropriate function that discriminates best between set of observations from two or more groups and classifying the function observations into one of the previously defined groups. Observations are classified into k non-overlapping groups on the basis p-variables.

Development of weather indices using correlation coefficient as weight

This is based on the method given by (Agrawal *et al.*, 1983) ^[5] for developing forecast using weather indices. In this procedure, the entire 15 weeks data have been utilized for constructing weighted and un-weighted weather indices of weather variables along with their interactions. In all, 72 indices (36 weighted and 36 unweighted) consisting of 8 weighted weather indices and 28 weighted interaction indices; 8 un-weighted indices and 28 un-weighted interaction indices have been obtained. These weather indices and interaction indices have been computed by using the following formula:

$$Z_{ij} = \sum_{w=1}^n r_{iw}^j X_{iw} \text{ and } Z_{ii'j} = \sum_{w=1}^n r_{ii'w}^j X_{iw} X_{i'w}$$

Where,

j = 0, 1 (where, '0' represents unweighted indices and '1' represents weighted indices)

n = Number of weeks considered in developing the indices

r_{iw} = Correlation coefficient between de-trend crop yield and i^{th} weather variable in w^{th} week

$r_{ii'w}$ = Correlation coefficient between de-trend crop yield and the product of i and i'^{th} weather variable in w^{th} week

X_{iw} and $X_{i'w}$ are the i and i'^{th} weather variable in w^{th} week respectively

Determination of number of discriminant scores

Consider that observations are classified into k (=3) non-overlapping groups on the basis p (=8) variables. The technique identifies linear functions where the coefficients of the variables 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 function that can be obtained is equal to minimum of (k-1) and p (Sharma, 1996) ^[15].

Development of the forecast model using discriminant function analysis

The crop years have been divided into three groups based on crop yield adjusted for trend effect in order to apply discriminant function analysis for modeling yield using weather variables. These three groups' meteorological data were utilized to create linear discriminant functions, and each year's discriminant scores were calculated. The forecast models are fitted by taking the discriminant scores and the trend variable as the regressors and crop yield as the regressand. The models are developed using stepwise regression procedure. Since there are three groups and eight meteorological variables in the current study, only two discriminant functions are necessary to classify a crop year into one of the three categories. These are the three groups of crop years, namely adverse, normal, and congenial: Let y and s represent the average and range of agricultural production after adjusting for n years. Adjusted crop yields less or equal to y-s would be in the adverse group, those between y-s and y+s would be in the normal group, and those above or equal to y+s would be in the congenial group.

Model 1

In this approach, all 72 indices (weighted and unweighted including interaction indices) served as discriminant variables in a discriminant function analysis. This yielded two sets of discriminant scores, which were subsequently utilized as regressor variables in a forecasting model. The model was fitted with untrended yield as the dependent variable (regressand) and the two sets of discriminant scores, along with the trend variable (T), as independent variables (regressors). The resultant model takes the form as follows:

$$Y = \beta_0 + \beta_1 ds_1 + \beta_2 ds_2 + \beta_3 T + \varepsilon$$

Where,

y = crop yield

β_0 = intercept of the model

β_i 's (i=0, 1, 2, 3) = the regression coefficients

ds_1 and ds_2 are the two discriminant scores. T is trend variable (T=1,2,3,...,n) and ε is error term assumed to follow independently normal distribution with mean 0 and variance σ^2 .

Model 2

In this approach, 16 weather indices (8 weighted and 8 unweighted) derived from 8 weather variables served as discriminant variables in a discriminant function analysis. This yielded two sets of discriminant scores, which were subsequently utilized as regressor variables in a forecasting model. The model was fitted with untrended yield as the dependent variable (regressand) and the two sets of discriminant scores, along with the trend variable (T), as independent variables (regressors). The resultant model takes the form as follows:

$$Y = \beta_0 + \beta_1 ds_1 + \beta_2 ds_2 + \beta_3 T + \varepsilon$$

Where the notations represent the same variables as defined in Model 1.

Model 3

In this approach, 36 unweighted weather indices (8 primary indices and 28 interaction terms) served as discriminant variables in a discriminant function analysis. This yielded two sets of discriminant scores, which were subsequently utilized as regressor variables in a forecasting model. The model was fitted with untrended yield as the dependent variable (regressand) and the two sets of discriminant scores, along with the trend variable (T), as independent variables (regressors). The resultant model takes the form as follows:

$$Y = \beta_0 + \beta_1 ds_1 + \beta_2 ds_2 + \beta_3 T + \varepsilon$$

Where the notations represent the same variables as defined in Model 1.

Model 4

This approach utilizes 36 weighted weather indices (8 primary indices and 28 interaction terms) as discriminant variables in a discriminant function analysis.

This yielded two sets of discriminant scores, which were subsequently utilized as regressor variables in a forecasting model. The model was fitted with untrended yield as the dependent variable (regressand) and the two sets of discriminant scores, along with the trend variable (T), as

independent variables (regressors). The resultant model takes the form as follows:

$$Y = \beta_0 + \beta_1 ds_1 + \beta_2 ds_2 + \beta_3 T + \varepsilon$$

Where the notations represent the same variables as defined in Model 1.

Model 5

This approach utilizes 28 weighted weather interaction indices and 28 unweighted weather interaction indices as discriminant variables in a discriminant function analysis. This yielded two sets of discriminant scores, which were subsequently utilized as regressor variables in a forecasting model. The model was fitted with untrended yield as the dependent variable (regressand) and the two sets of discriminant scores, along with the trend variable (T), as independent variables (regressors). The resultant model takes the form as follows:

$$Y = \beta_0 + \beta_1 ds_1 + \beta_2 ds_2 + \beta_3 T + \varepsilon$$

Where the notations represent the same variables as defined in Model 1.

All the aforesaid models have been fitted with the data pertaining to the years 2001 to 2017 and the data pertaining to the year 2018 to 2021 were used for validation of the forecast models.

Testing the Performance of the Model

Finally the performance of the developed models was evaluated on the basis of Coefficient of determination (R^2), Adjusted Coefficient of determination ($Adj.R^2$), Root Mean Square Error (RMSE) and Percentage Error (PE).

R^2 & $Adj.R^2$ towards 1 and RMSE towards 0 indicate better performance of the developed models. Also least range of PE indicates better performance.

Results and Discussion

Yield forecast models were constructed using maize crop yield data (kg/ha) and weather data of maize for a period of 21 years (2001-2021). 17 years data for training the model, from 2001 to 2017 were used and 4 years data for testing the model, from 2018 to 2021 were used. Training set establishes the relationship between predictors and dependent variable. Testing dataset determine the prediction accuracy of forecasted models. In this study five models were created with discriminant scores as independent variables which also include time trend and maize crop yield as dependent variable.

Table 1: Maize yield forecast models

Model	Forecast equation	R^2 (%)	$Adj. R^2$ (%)
1.	$Y = 218.02 + 9.35ds_1 + 21.85ds_2 - 6.01T$	76.57	67.52
2.	$Y = 125.02 - 12.01ds_1 + 68.11ds_2 - 8.66T$	51.65	44.13
3.	$Y = 78.16 - 51.28ds_1 - 69.24ds_2 - 14.02T$	68.31	59.52
4.	$Y = 69.23 + 11.36ds_1 + 11.08ds_2 + 12.66T$	81.28	74.05
5.	$Y = 93.57 + 38.91 + 48.85ds_2 - 7.34T$	78.72	70.55

Forecast models are presented in Table 1 along with their values of R^2 and $Adj. R^2$. The value of R^2 and $Adj. R^2$ has been found to be maximum of about 81.28% and 74.05% respectively in model 4 followed by model 5 with the value of R^2 (78.72%) and $Adj. R^2$ (70.55%). Model 2 shows the minimum value of R^2 (51.65%) and $Adj. R^2$ (44.13%).

Using these forecast models the forecast values of maize crop yield for the years 2018, 2019, 2020 and 2021 were obtained

and the results are presented in Table 2.

Table 2: Actual and Forecasted yield of maize crop

Year	Actual yield (kg/ha)	Forecasted yield (kg/ha)				
		Model 1	Model 2	Model 3	Model 4	Model 5
2018	1324.50	1201.30	1427.20	1185.33	1330.69	1296.13
2019	1338.24	1498.71	1178.71	1215.36	1289.32	1405.63
2020	2174.60	1967.01	2501.19	2492.13	2205.69	2081.50
2021	2688.20	2809.03	2490.56	2507.96	2718.11	2573.56
RMSE(Kg/ha)		387.85	601.26	494.71	251.14	301.69

Forecasted yield of maize crop are presented in Table 2 along with their values of Root Mean Square Error (RMSE). RMSE value was found to be 387.85 kg/ha for model 1, 601.26 kg/ha for model 2, 494.71 kg/ha for model 3, 251.14 kg/ha for model 4 and 301.69 kg/ha for model 5. Model 4 shows lower value of RMSE followed by model 5. On the other hand model 2 shows higher value of RMSE.

Table 3: Percentage Error of developed models

Year	Error (%)				
	Model 1	Model 2	Model 3	Model 4	Model 5
2018	9.30	-7.75	10.49	-0.47	2.14
2019	-11.99	11.92	9.18	3.65	-5.03
2020	9.55	-15.02	-14.60	-1.43	4.28
2021	-4.49	7.35	6.70	-1.11	4.45

Table 3 revealed the percentage error of five developed models. Percentage error ranged from -11.99% to 9.55% for model 1, ranged from -15.02% to 11.92% for model 2, ranged from -14.60% to 10.49% for model 3, ranged from -1.43% to 3.65% for model 4 and ranged from -5.03% to 4.45% for model 5. Model 4 shows least range of percentage error followed by model 5. On the other hand model 2 shows highest range of percentage error.

The model 4 is found to be with lower RMSE (251.14 kg/ha), high adj R^2 (74.05%) and least range of percentage error (-1.43 to 3.65) followed by model 5. On the other hand, model 2 performs poorly in comparison with other models. The Tables (1, 2 and 3) indicate that the model 4 is the most appropriate one followed by the model 5 for the forecast of the maize crop yield during kharif season in Udham Singh Nagar district of Uttarakhand. Hence, the model 4 and model 5 are recommended for maize crop yield forecasting of the studied region.

Conclusion

Discriminant function analysis is a powerful statistical technique for predicting outcomes based on multiple variables. In this case, it's been used to develop forecast models for maize crop yield in Udham Singh Nagar district of Uttarakhand using weekly data on eight weather variables. One of the five developed models, model 4 which includes all the 36 weighted weather indices (8 primary indices and 28 interaction terms) has been identified as the best performers based on various evaluation criteria such as coefficient of determination, Root Mean Square Error (RMSE) and percentage error. This model can provide a reliable forecast of maize crop yield which is incredibly valuable for farmers, agricultural planners, and policymakers. This kind of predictive insight can help inform decision-making related to crop management, resource allocation, and market preparation.

References

1. Food and Agriculture Organization of the United Nations (FAO). Maize: facts and trends. Rome; c2020.
2. Fisher RA. The influence of rainfall on the yield of wheat at Rothamsted. Philosophical Transactions of the Royal Society of London. Series B. 1924;213:89-142.
3. Hendricks WA, Scholl GC. Technique in measuring joint relationship: The joint effects of temperature and precipitation on crop yield. North Carolina Agricultural Experiment Station Technical Bulletin; c1943, p. 74.
4. Agrawal R, Jain RC, Jha MP, Singh D. Forecasting of rice yield using climatic variables. Indian Journal of Agricultural Sciences. 1980;50(9):680-684.
5. Agrawal R, Jain RC, Jha MP. Joint effects of weather variables on rice yields. Mausam. 1983;34(2):189-194.
6. Agrawal R, Jain RC, Jha MP. Models for studying rice crop weather relationship. Mausam. 1986;37(1):67-70.
7. Agrawal R, Jain RC, Mehta SC. Yield forecast based on weather variables and agricultural input on agro-climatic zone basis. Indian Journal of Agricultural Sciences. 2001;71(7):487-490.
8. Jain RC, Singh D. Forecasting rainfall over Puerto Rico. Annual Report, Department of Meteorology. The Florida State University; c1980.
9. Rai T, Chandras. Use of discriminant function of weather parameters for developing forecast model of rice crop. Publication of Indian Agricultural Statistics Research Institute, New Delhi; c2000.
10. Agrawal R, Chandras, Aditya K. Use of discriminant function analysis for forecasting crop yield. Mausam. 2012;63(3):455-458.
11. Sisodia BVS, Yadav RR, Kumar S, Sharma MK. Forecasting of pre-harvest crop yield using discriminant function analysis of meteorological parameters. Journal of Agrometeorology. 2014;16(1):121-5.
12. Mohd Azfar, Sisodia BVS, Rai VN, Devi M. Pre-harvest forecast of rapeseed and mustard yield based on weather variables: An application of discriminant function analysis. International Journal of Agricultural and Statistical Sciences. 2014;10(2):497-502.
13. Azfar, Sisodia BVS, Rai VN, Kumar S. A study on impact of changes in climate variables on rice production in Faizabad district of Uttar Pradesh. Mausam. 2015;66(2):281-286.
14. Yadav RR, Sisodia BVS, Kumar S. Pre-harvest forecast of pigeon-pea yield using discriminant function analysis of weather variables. Mausam. 2016;67(3):577-582.
15. Sharma S. Applied multivariate techniques. New York: John Wiley & Sons; c1996.
16. Agrawal R, Ramakrishna YS, Rao KAVR, Kumar A, Bhar L, Mohan M, Saksena A. Modeling for forecasting of crop yield using weather parameters and agricultural

- inputs. Project Report under AP Cess Fund scheme of ICAR; c2005.
17. Anderson TW. An introduction to multivariate statistical analysis. New York: John Wiley & Sons; c1984.
 18. Aneja KG, Chandrahas. Preharvest crop yield forecast based on plant biometrical characters. In: Silver jubilee souvenir. Indian Agricultural Statistics Research Institute, New Delhi; c1984.
 19. Aravind K, Vashisth A, Krishanan P, Das B. Wheat yield prediction based on weather parameters using multiple linear, neural network and penalized regression models. *Journal of Agrometeorology*. 2022;24(1):18-25. <https://doi.org/10.54386/jam.v24i1.1002>
 20. Baier W. Crop-weather analysis model. Review and model development. *Journal of Applied Meteorology*. 1973;12(6):937-947.
 21. Baier W. Crop weather models and their use in yield assessments. Technical note no. 151. World Meteorological Organization, Geneva; c1977, p. 48.
 22. Chandrahas, Narain P. Pilot studies on pre-harvest forecasting of apple yield based on data on biometrical characters, weather factors and crop inputs in Shimla District (H. P.) during 1984-86. Indian Agricultural Statistics Research Institute, New Delhi; c1993.
 23. Deakin EB. A discriminant analysis of predictors of business failure. *Journal of Accounting Research*. 1972;10(1):167-179.
 24. Draper NR, Smith H. Applied regression analysis. 3rd Ed. New York: John Wiley & Sons Inc.; c1998.
 25. Goudie AW. Forecasting corporate failure: The use of discriminant analysis within a disaggregated model of the corporate sector. *Journal of the Royal Statistical Society, Series A (General)*. 1987;150(1):69-80.
 26. Hair JF, Anderson RE, Tatham RL, Black WC. Multivariate data analysis with readings. New Jersey: Prentice Hall; c1995.
 27. Jain RC, Jha MP, Agrawal R. Use of growth indices in yield forecast. *Biometrical Journal*. 1985;27(4):435-439.
 28. Johnson RA, Wichern DW. Applied multivariate statistical analysis. 3rd Ed. New Delhi: Prentice Hall of India; c2001.
 29. Johnson RA, Wichern DW. Applied multivariate statistical analysis. Pearson Education; c2006.
 30. Lee TS, Chih CC, Chi JL. Forecasting of credit scores using discriminant function method. Taipei: Department of Business Administration, Fu-Jen Catholic University; c2002.
 31. Taffler RJ. Forecasting company failure in the UK using discriminant analysis and financial ratio data. *Journal of the Royal Statistical Society, Series A*. 1982;145(3):342-358.