Forecasting By Discriminant Function Weather Based Analysis

KK Pandey*, BVS Sisodia and VN Rai

1Department of Agricultural Statistics, VCSG College of Horticulture, India
2Department of Agricultural Statistics, Narendra Dev University of Agriculture and Technology, India

Submission: August 26, 2017; Published: September 01, 2017

*Corresponding author: KK Pandey, Department of Agricultural Statistics, VCSG college of Horticulture, Uttarakhand University of Horticulture and Forestry, Bharsar, Pauri-Garhwal, India, Email: kkpandeystat@gmail.com

Abstract

In this present study, an attempt to be made on an application of discriminant function analysis of meteorological parameters for developing suitable statistical model to forecast rice yield for Faizabad district of eastern Uttar Pradesh has been demonstrated. Time series data on rice yield for 21 years (1990 to 2010) have been divided into three groups, viz. congenial, normal and adverse based on de-trended yield distribution. Considering three groups as three populations, Discriminant function analysis using weekly data of crop season on seven meteorological parameters has been carried out. The Discriminant scores obtained from this have been used as regress or variables along with time trend in development of statistical model. In this procedure using weekly weather data have been proposed. For the development of the model the data (1990 to 2008) has been used and rest two year (2009 and 2010) data has been used for the validation of model. It has been found that the model provide reliable forecast of the rice yield about two months before the harvest. This model is newly proposed model. However, the model has been most suitable on the basis of Radj2 (87.5%) and RMSE (0.4688).

Keywords: Forecast model; Meteorological parameters; Rice yield; Radj; RMSE

Introduction

Weather is one of the most important factors influencing crop growth. It may influence production directly through affecting the growth structural characteristics of crop such as plant population, number of tillers leaf area etc., and indirectly through its effect on incidence of pests and diseases. The effect of weather parameters at different stages of growth of crop may help in understanding their response in term of final yield and also provide a forecast of crop yields in advance before the harvest. Changes in the timing of phonological events are among the most important indicators of global warming Parmesan and Yohe [1]. The extent of weather influence on crop yields depends not only on magnitude of weather parameters but also on their frequency distribution. Menzel and Fabian [2] reported on phonological change due to increasing of temperature.

The alternation in global warming has dramatically affected agriculture and its productivity. The increase in temperature has significantly led to change in the agricultural zones and shift in the growing season. Fisher [3] has been used by biologists to solve the classificatory problems involving multiple measures in different contexts. Models based on weather parameters can provide reliable forecast of crop yield in advance of harvest. Agrawal and Mehta [4]. The forecasting equations have also been developed for wheat yield in Kanpur district U.P. Agrawal [5], Rai, Chandrasah [6] made use Discriminant function of weather variables to develop statistical models for pre-harvest forecasting of rice yield in Raipur district of Chhattisgarh. The model on the basis of weather variable have been done by Agrawal [7,8]. A lot of works have been done for the development of the model with the weather variables but no work has been done in this direction for the eastern Uttar Pradesh for rice crop. In the present paper, an attempt has been made to develop suitable statistical model for forecasting of pre-harvest rice yield in Faizabad district using Discriminant scores from Discriminant functions obtained from the weekly data on weather variables with a few modifications.

Materials and Methods

The study has been conducted for Faizabad district of Eastern Uttar Pradesh, which is situated between 26° 47’ N latitude and 82° 12’ E longitudes. It lies in the Eastern Plain Zone (EPZ) of Uttar Pradesh. It has an annual rainfall of about 1002 mm. Time series data pertaining to yield of rice crop for Faizabad district of Uttar Pradesh for 21 years (1990 to 2010) has been procured.
Then the vector of coefficients $\hat{\beta}_j$ that maximize the ratio

$$
\frac{\hat{\beta}_j'\tilde{Y}_j}{\hat{\beta}_j'\hat{\beta}_j} = \frac{\hat{\beta}_j'\sum_{i=1}^g (\bar{x}_i - \bar{x}) (\bar{x}_i - \bar{x})' \hat{\beta}_j}{\hat{\beta}_j'\sum_{i=1}^g (\bar{x}_i - \bar{x}) (\bar{x}_i - \bar{x})' \hat{\beta}_j} \frac{\hat{\beta}_j'\sum_{i=1}^g (\bar{x}_i - \bar{x}) (\bar{x}_i - \bar{x})' \hat{\beta}_j}{\hat{\beta}_j'\sum_{i=1}^g (\bar{x}_i - \bar{x}) (\bar{x}_i - \bar{x})' \hat{\beta}_j}
$$

is given by $\hat{\beta}_1 = \hat{c}_1$. The linear combination $\hat{I}_1'X$ is called the simple first discriminant. The choice $\hat{I}_2' = \hat{c}_2$ produce the sample second discriminant, $\hat{I}_3'X$ continuing $\hat{I}_k' = \hat{c}_kX$ is the sample $k^{th}$ discriminant function, $k \leq s$. For example, if $g=3$ and $p=4$, the number of discriminant function will be 2.

The entire Fisher’s discriminant function can be expressed as Anderson’s classification function (statistic) for the purpose of discriminant scores as follows:

$$
\mathbf{d}_i = a_i + \hat{I}_{1i} x_1 + \hat{I}_{2i} x_2 + \ldots + \hat{I}_{pi} x_p , \; i=1,2,\ldots,k \; (ss)
$$

**Development of forecast models**

The crop years have been developed into three groups namely, congenial normal and adverse on the basis of crop yield, which is adjusted for trend effect. Here, only the first 19 year data from 1990 to 2008 have been utilized for the model fitting and remaining two years were left for the validation of the model. Weekly data on weather variables corresponding to three pre defined groups have been used for the development of forecast models. In the present study the number of groups is three and number of weather variable is seven. Therefore only two scores will be obtained. Discriminant analysis approach predicts the future observations qualitatively in different groups. For quantitative forecasting, regression models are fitted by taking the scores and the trend variable as and crop yield as the entire 19 weeks data from 23rd to 36th (Standard meteorological week) have been utilized for development of the model.

**Development of the Model**

In this procedure, function analysis have been carried out using the data on the first weather variables spread over 12 weeks using 23rd to 36th SMW. Using two scores obtained function of the data on the first weather variable and 14 week data on second variable, function analysis has been again performed and two sets of scores are obtained (here the discriminating variables will now become 16). Using these two sets of scores and 14 week data of third variable have been again used to analysis and subsequently two sets of scores have been obtained up to seventh weather variables, and ultimately we get two set of scores. These two sets of scores and the trend as the variable and crop yield as were utilized to develop forecast model by fitting the following model:
\[ y = \beta_0 + \beta_1 \Delta d_1 + \beta_2 \Delta d_2 + \beta_3 T + e \]

Where \( y \) is detrended crop yield, \( \beta_i \)'s \( (i = 1, 2, 3) \) are model parameters, \( \Delta d_1 \) and \( \Delta d_2 \) are two sets of discriminant scores, \( T \) is the trend variable and \( e \) is error term assumed to follow \( N(0, \sigma^2) \). This model utilized the complete data over 14 weeks and also considers relative importance of weather variables in different weeks.

**Comparison and validation of forecast models**

Different procedures have been used in the present study for the comparison and the validation of the developed models. These procedures are given below.

a) \( R_{adj}^2 \): The significance of the model can evaluate on the basis of adjusted coefficient of determination \( (R_{adj}^2) \) which is as follows:
\[
R_{adj}^2 = 1 - \frac{S_{res}/(n-p)}{S_{tot}/(n-1)}
\]

\( S_{res} \) is the residual mean square and \( S_{tot} \) is the total mean square.

b) The percent deviation of the forecast yield from actual yield have been computed by the following formula:
\[
\text{Percentage deviation} = \left( \frac{\text{Actual yield} - \text{Forecasted yield}}{\text{Actual yield}} \right) \times 100
\]

c) Root Mean Square Error (RMSE):

It is also a measure for comparing two models. The formula of RMSE is given below
\[
RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (O_i - E_i)^2}
\]

where \( O_i \) and the \( E_i \) are the observed and forecasted value of the crop yield respectively and \( n \) is the number of years for which forecasting has been done.

d) Percent Standard error of forecast:

Let \( \hat{Y}_f \) be forecast value of crop yield and \( X_0 \) be the column vector of \( P \) independent variable at which \( y \) is forecasted then variance \( \hat{Y}_f \) is given by (Draper and Smith, 1998) is obtained as
\[
V(\hat{Y}_f) = \hat{\sigma}^2 X_0' (X'X)^{-1} X_0
\]

where \( X'X \) is the dispersion matrix of the sum of square and cross products of regressors (independent variables) and is the estimated residual variance of the model. Therefore the Percent Standard Error (C.V.) of forecast value is given by
\[
\text{SE}(\hat{Y}_f) = \sqrt{\frac{V(\hat{Y}_f)}{\hat{\sigma}^2}}
\]

Then Percent S.E. = \( \frac{\text{SE}(\hat{Y}_f)}{\text{Forecast value}} \times 100 \)

*Table 1*: Rice Yield Forecast Model.

<table>
<thead>
<tr>
<th>Model</th>
<th>Forecast regression equation</th>
<th>( R_{adj}^2 ) (%)</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Yield = 20.164 + 0.009<strong>ds1+ 0.020</strong>ds2+ 0.088*T</td>
<td>87.5</td>
<td>0.4688</td>
</tr>
<tr>
<td></td>
<td>(0.007) (0.005) (0.024)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: figures in brackets denote Standard error of regression coefficient. **P < 0.01, *P < 0.05, +P<0.10

*Table 2*: Actual and Forecasted Yield of Rice (Q/ha).

<table>
<thead>
<tr>
<th>Year</th>
<th>Actual Yield</th>
<th>Model</th>
<th>( R_{adj}^2 ) (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2008-09</td>
<td>23.77</td>
<td>23.35 (1.77)</td>
<td></td>
</tr>
<tr>
<td>2009-10</td>
<td>23.01</td>
<td>22.45 (2.43)</td>
<td></td>
</tr>
<tr>
<td>2008-09</td>
<td>PSE (CV)</td>
<td>2.87</td>
<td></td>
</tr>
<tr>
<td>2009-10</td>
<td>PSE (CV)</td>
<td>2.99</td>
<td></td>
</tr>
</tbody>
</table>

Note: figures in brackets denote percent deviation of forecast. CV: Coefficient of variation.

**Results and Discussion**

The forecast models for the rice crop yield have been developed under this procedure along with \( R_{adj}^2 \) and RMSE are given in (Table 1). First Discriminant score \( ds1 \) has been found to be significant at one percent probability level of significance \( (p < 0.01) \) in the model and the second discriminant score has been found to be significant at one percent probability level of significance \( (p < 0.01) \) in model. Adjusted coefficient of determination \( (R_{adj}^2) \) has been found to be 87.5% in the model.

The root mean square error (RMSE) of forecast model has been found 0.4688 for model. Based on forecast model, the forecasted yield for the year 2008-09 and 2009-10 were obtained and the results are presented in (Table 2).

It is evident from result that percent deviation of forecast yield 1.77 in 2008-09 and 2.43 in 2009-10. The percent standard error (CV) for forecasted yields has been also computed for the model and i.e. presented in (Table 2). The percent standard error
Thus, it can be concluded that the proposed model is most suitable model to forecast rice yield in Faizabad district of Eastern Uttar Pradesh. Hence, a reliable forecast of rice yield about two months before the harvest can be obtained from the proposed model.

Minhajuddin [13] proposed a method to simulate the joint distribution which have equal to positive pair-wise correlations and the method was illustrated for the p-dimensional families of beta and gamma distributions. Sever [14] compared fisher’s discriminant analysis under normal and skewed curved normal distribution based on the apparent error rates, which were used as a measure of classification performance and found that fisher’s linear discriminant analysis to be highly robust under skewed curved normal distribution. Rausch, Kelley [15] compared different methods for discriminant analysis with respect to classification accuracy under non normality through Monte Carlo simulation. Pandey [16] compared different distribution as normal, lognormal, and pearson’s type on the basis of weather variable on wheat yield for Faizabad district of Eastern U.P. Raman [17] compared non-normal rice and maize yield with linear discriminant function analysis under multivariate analysis for New Delhi. Ito and Schull [18] discuss the robustness of T02 Statistics, when the conditions of equality of covariance matrices are not satisfied.

The Dirichlet distribution is a multivariate generalization of beta distribution Kotz [19]. Almost similar results, as observed in this study, have also been reported by Kandiannan [20] for Coimbatore in Tamil Nadu, where temperature, rainfall and radiation entered significantly in a stepwise prediction equation of rice yield. In Andhra Pradesh also, rainfall and temperature have been reported to affect rice yield significantly Barnwal and Kotani [21], Lal [22] also observed that maximum temperature, minimum temperature and moisture stress were crucial weather variables affecting soybean yield. Temperature, rainfall and relative humidity were found significantly correlated with sugarcane yield, Srivastava [23,24].

References