International Journal of Chemical Studies

P-ISSN: 2349–8528
E-ISSN: 2321–4902
IJCS 2019; 7(4): 2902-2906

Received: 16-05-2019
Accepted: 18-06-2019

HG Singh
Department, of Agri. Statistics, ND. University of Agri and Tech., Kumarganj, Faizabad, Uttar Pradesh, India

VN Rai
Department, of Agri. Statistics, ND. University of Agri and Tech., Kumarganj, Faizabad, Uttar Pradesh, India

BVS Sissodi
Department, of Agri. Statistics, ND. University of Agri and Tech., Kumarganj, Faizabad, Uttar Pradesh, India

V Nand
Department, of ILFC, N. D. University of Agri and Tech., Kumarganj, Faizabad, Uttar Pradesh, India

HY Singh
Department, of Horticulture, Janta College Bakewar Etawah, Pradesh, India

Sushil Kumar
Directorate of Extension, C.S. Azad University of Agri. & Tech., Kanpur, Pradesh, India

Correspondence
V Nand
Department, of ILFC, ND. University of Agri and Tech., Kumarganj, Faizabad, Uttar Pradesh, India

Forecasting of pre-harvest wheat crop yield using discriminant function analysis (DFA) of meteorological parameters

HG Singh, VN Rai, BVS Sissodi, V Nand, HY Singh and Sushil Kumar

Abstract
The present paper entitled "A study on crop yield weather relationship and yield forecast models of wheat for Barabanki district of Eastern U.P." Time series data on yield of wheat and weekly data from 44th SMW of the previous year to 6th SMW of the following year on eight weather variables *viz.,* Minimum Temperature, Maximum Temperature, Rainfall, Rainy days, Relative humidity 08.30hrs, Relative humidity 17.30hrs, and Wind-Velocity 08.30 Wind-Velocity 17.30hrs covering the period from 1990-91 to 2011-12 have been utilized for development of pre-harvest forecast model. Statistical methodologies using multiple regressions, discriminant functions analysis for developing pre-harvest forecast model have been described. In all, 13 models (one based on regression, six from discriminant and six from composite forecast) have been developed. The Model-A is based on weather indices and D1 to D6 are based on discriminant functions analysis and 6 composite forecast models have been developed on the basis of adjR^2, RMSE and %SE, the best two models obtained by the application of discriminant and composite forecast analysis of weekly weather data are D3 & D4. However, the composite forecast models CF-1 & CF-3 have further reduced the percentage standard error of the forecast yield to some extent. These models can be used to get the reliable forecast of wheat yield two and half months before the harvest.

Keywords: Wheat, yield, discriminant function, model and forecasting

Introduction

Wheat (*Triticum aestivum* L.) is important cereal crop in India occupying second place, next to rice in production and plays a critical role in food security and has wide adaptability. It can be grown not only in the tropical and sub-tropical zones but also in the temperate zone and the cold tracts of the far north, beyond even 60 degree latitude. The best wheat are produced in areas favored with cool, moist weather during the major portion of the growing period followed by dry, warm weather to enable the grain to ripe properly. India raises almost exclusively winter wheat.

Forecasting opens menu window on to future. It plays a vital role in most of our activities and in all we do concern about future. Establishing a functional crop yield, production forecasting system is an extremely important component of a national agricultural statistical system. Crop forecasts are used for a number of policy decisions such as those relating to procurement, trade, storage, transportation, distribution, import and export of food grains in general, and for implementing food security programme in particular. Crop production forecasts are not only meant to serve the interests of governments but other stakeholders in the agricultural sector also find use for crop forecast data in their day-to-day decision functions. The value of the various policy and business decisions could be enhanced if these are supported by a strong system of food crop production forecasting. The variability both within and between season is the second and uncontrollable sources of variability in yields. Pre-harvest forecast statistical models based on weather variables, input factors and biometrical character have been developed by using regression and discriminant function analysis by present workers in wheat crop Sisodia *et al*. (2014) [12]. V Cannor, W.P. (1999) [6], Carter and Eisner (1997) [7] Bharati, R. C. Khatri, T. J. Parikh, R. K. (1999) [5]. Recently developed forecast models for wheat yield in Kanpur district (U.P.) using discriminant functions analysis of weakly data on weather variables. Since the discriminant function analysis discriminates best between sets of observations from two or more groups and classify the future observations into one of the previously defined groups Agrawal *et al*. (2012) [1], an attempt has been made in the present paper to develop suitable statistical models for forecasting of pre-harvest wheat yield in Barabanki district of Uttar Pradesh using discriminant functions analysis of weekly data on weather variables.

Weather variables affect the crop at different stages of crop development. Thus extent of weather influence on crop yield depends not only on magnitude of weather variable but also on the distribution pattern of weather over the crop season which, as such calls for the necessity of dividing the whole crop season in to finer intervals. However doing so will increase number of variable in the models and in turn a large number of observations may not be available for precise estimation of these parameters. Suitable dimension technique such as like transformation say forming indices or multivariate technique such as principle component, discriminate function etc. can be employed to tackle such problem. Pre-harvest forecasting is extremely useful in formulation of polices regarding stock, distributions and supply of agricultural produce to various part of the country. Pest and diseases are one of the major causes of reduction in crop yield thus there is need to develop forewarning model which provide advance information for outbreak of pest and diseases attack so that production measures can be implemented before the actual onset of the damage. Timely application of control measures may reduce the yield losses. This information would be obtained through modeling qualitative data (Agrawal et al. 2007) [2].

Materials and Methods

Area and crop covered

The studies were conducted for Barabanki district of Eastern Uttar Pradesh, India which is situated at 26.92^0 N latitude 81.20^0 E longitudes. It lies in the Eastern plain zone of Uttar Pradesh. It has an annual rainfall of about 1002 mm and is liberally sourced by the Gomti River and its tributaries. Soils are deep alluvial, medium to medium heavy textured but are easily plough able. The favorable climate, soil and the availability of ample irrigation facilities make growing of rice and wheat a natural choice for the area. The objective is to develop pre-harvest forecast model for wheat yield.

Data

Time series data on yield for wheat crop of Barabanki district of Eastern Uttar Pradesh for 22 years (1990-91-2011-12) were collected data from the bulletins of Directorate of Agricultural Statistics and Crop Insurance, Govt. of Uttar Pradesh.

Weekly weather data for the period (1990-91-2011-12) on the weather variables of Barabanki district of Uttar Pradesh during the different growth phases of wheat crop were obtained from the Department of Agro meteorology, Amousi Airport Lucknow. The data were collected up to the first 15 weeks of the crop cultivation which included 44th standard meteorological week (SMW) to 52nd SMW of a year and 1st SMW to 6th SMW of the next year. The data on five weather variables viz. minimum temperature, maximum temperature, rainfall, rainy day, relative humidity 8.30, relative humidity 17.30, wind-velocity 8.30 and wind-velocity 17.30 were used in the study.

Statistical Methodology

The primary objective of this study is to develop suitable statistical models for forecasting pre-harvest yield of wheat crop using of discriminant function analysis.

Discriminant function analysis

We are going to describe the technique of discriminant function analysis.

The discriminant function analysis is a multivariate technique already discussed in many standard books, by Anderson (1984), Hair *et al*. (1995), Sharma (1996) [12], Johnson & Wichran (2006), etc. Here we will discuss only at discriminant function analysis of Johnson & Wichran (2006).

$$DF=\sum_{i=l}^{p} li\, X_i \quad \text{.... (1)}$$

where, X_i is the ith variable used to discriminate the groups and l_i is the corresponding discriminant coefficient, p being the number of variables. DF is a scalar quantity which is generally designated as discriminant score based on X_i's and l_i's.

The optimal value for the weighing coefficient, l_i, is determined in such a way that the variation between the two groups gets maximized relative to the variation within the groups. Maximizing this ratio yields simultaneous equations of the following form

$$l_1\sum x_1^2 + l_2\sum x_1x_2 + \ldots\ldots\ldots + l_p\sum x_1x_p = d_1$$

$$l_1\sum x_1x_p + l_2\sum x_1x_p + \ldots\ldots\ldots + l_p\sum x_p^2 = d_p$$

where x_1, x_2, ……., x_p are deviations from their respective group means and $d_i = \bar{X}_{i1} - \bar{X}_{i2}$ for i= 1, 2, … p. $\bar{X}_{i1}$ and $\bar{X}_{i2}$ are means of ith variable in two population. In matrix notation

$S\,l = d$

where **S** is the discriminant dispersion matrix given by

$$S=\begin{pmatrix} \Sigma x_1^2 & \sum x_1x_2 \ldots\ldots\ldots \sum x_1x_p \\ & \sum x_2^2 \ldots\ldots\ldots \sum x_2x_p \\ \ldots\ldots \quad \ldots.. & \ldots\ldots \ldots\ldots \\ & \Sigma x_p^2 \end{pmatrix}$$

where S_{ij} is the sum of product of the ith and jth variables in matrix S.*l* and *d* are column vectors of discriminant coefficients and differences in group means respectively. The value of l can be obtained as

$l = S^{-1}d$

The adequacy of the function is tested with the help of 'F' test as

$$F_{(p,n_1+n_2-p-1)} = \frac{n_1n_2(n_1+n_2-p-1)}{(n_1+n_2)p} D^2 \quad \text{... (2)}$$

where, $D^2=\sum_{i=1}^{p} l_1 d_i$ (known as Mahalanobis D^2 statistic), d_i is the difference in means i.e. $\bar{X}_{i1} - \bar{X}_{i2}$ for the ith variable, n_1,

n_2 are the number of units in the two groups and $\overline{X}_{i1}$ and $\overline{X}_{i2}$ are means of the i[th] variable for the two groups.

The discriminatory point S_c for classifying individual years into two groups is calculated as

$$S_c = \frac{(n_1\overline{S}_1 + n_2\overline{S}_2)}{(n_1+n_2)} \quad \ldots (3)$$

where $S_1 = \sum_{i=11}^{p} l_i \, X_{i1}$ and $S_2 = \sum_{i=1}^{p} l_i X_{i2}$ and $\overline{S}_1 \,\&\, \overline{S}_2$ are the corresponding group centroids. 1 and 2 denote the groups.

Development of forecast Model

The crop years were divided into three groups namely adverse, normal and congenial has been obtained as follows: Let *y* and s be the mean and standard deviation of the adjusted crop yields of n years. The adjusted crop yields less than or equal to – s would form adverse group, the adjusted crop yields between -s and +s would from normal group and adjusted crop yields above or equal to +s would from congenial group. It is, however, known that weather variables affect the crop differently during different phases of crop development. Its effect depends not only on its magnitude but also on its distribution pattern over the crop season. Therefore, using weekly weather data as such in developing the model poses a problem as number of independent variables in the regression model would increase enormously. To solve this problem, following weather indices have been developed using the procedure of (Agrawal *et al.* 1983, 1986). Where Z_{ij} is un-weighted (for j=0) and weighted (for j=1) weather indices for i[th] weather variable and is the un-weighted (for j=0) and weighted (for j=1) weather indices for interaction between i[th] and i[th] weather variables. $\boldsymbol{X_{iw}}$ is the value of the i[th] weather variable in w[th] week, *riw/rii'w* is correlation coefficient of yield adjusted for trend effect with i[th] weather variable/product of i[th] and i'[th] weather variable in w[th] week, n is the number of weeks considered in developing the indices and p is number of weather variables. Here, p=5 and n=15.

The entire 15 weeks data from 44[th] week to 52nd week of a year and 1[st] week to 6[th] week of the next year have been utilized for constructing weighted and un-weighted weather indices of weather variables along with their interactions. In all 30 indices (15 weighted and 15 un-weighted) consisting of 5 weighted weather indices and 10 weighted interaction indices, and 5 un-weighted and 10 un-weighted interaction indices have been constructed. Besides, some more suitable strategies have been suggested. In all, six possible models are attempted. First two are the best models reported by Agrawal *et al*. (2012) [1] and last four are newly proposed. Models are developed using regression analysis. Only the first 22 years data from 1990 to 2012 have been utilized for modeling the yield and remaining three years yield data of 2010, 2011 and 2012 have been used for validation of the models.

These weather indices have been computed by using the following formula.

$$y = a_0 + \sum_{i=1}^{p} \sum_{j=0}^{1} a_{ij} Z_{ij} + \sum_{i \neq i'-1}^{p} \sum_{j=0}^{l} a_{ii'j} Z_{II'J} + cT + \varepsilon$$

where $Z_{ij} = \sum_{w=1}^{n} r_{iw}^{j} X_{iw} \Big/ \sum_{w=1}^{n} r_{iw}^{j}$ j = 0,1

$$Z_{ii'j} = \sum_{w=n_1}^{n_2} r_{ii'w}^{j} X_{iw} X_{i'w} \Big/ \sum_{w=n_1}^{n_2} r_{ii'w}^{j}$$

y is the original crop yield, X_{iw} is the value of the i[th] weather variable in w[th] week, $r_{iw}/r_{ii'w}$ is correlation coefficient of yield adjusted for trend effect with i[th] weather variable/product of i[th] and i'[th] weather variable in w[th] week, n is the number of weeks considered in developing the weather indices, and p is number of weather variables used. a_0, a_{ij}, $a_{ii,j}$ and c are the model parameters. ε is error term assumed to follow N (0, σ^2).

The step-wise regression analysis was employed to develop the forecast model.

Model 1

This model is the 2[nd] model of Agrawal *et al.* (2012) [1]. Using five weighted weather indices of five weather variables, discriminant function analysis was carried out and two discriminant functions have been obtained. Two sets of discriminant scores for the years under consideration from these two discriminant functions were obtained. For developing forecast model, these two sets of discriminant scores along with the trend variable were utilized as the regressor and the yield as the regress and. The form of model considered is as follows: y = *βo+ β1ds1+ β2ds2+ β3T + e* where y is un-trended crop yield, *âi's (i =0,1,2,3)* are model parameter, ds1 and ds2 are two sets of discriminant scores, *T* is the trend variable and *e* is error term assumed to follow independently N (0, ó2). This model utilizes the complete data over 15 weeks and also considers relative importance of weather variables in different weeks.

Model 2

This model is 4th model of Agrawal *et al.* (2012) [1]. Two discriminants functions and there from two sets of discriminant scores were obtained using the first week data (44[th] SMW) on five weather variables. Two sets of discriminant scores obtained from first week data and data on five weather variables in the second week (45thSMW) were used as discriminating variables, so in all there were 7 discriminating variables, and based on these 7 discriminating variables the discriminant analysis has been done and, therefore, two sets of discriminant scores were obtained. This process was repeated up to the last week till the time of forecast (6th SMW or 15th week) and finally two sets of discriminant scores (ds1 and ds2) were obtained. Based on these two sets of scores obtained at the 15th week, the forecasting model taking yield as the

Regress and and the discriminant scores and the trend as the regressor variables has been fitted similar to that given in Model 1.

Model 3

In this model, five weighted and five un-weighted weather indices of five weather variables were used as discriminating variables in the discriminant function analysis. Two sets of scores from two discriminant functions were obtained. The forecasting model were fitted taking the yield as the regress and and the two sets of scores (ds1 and ds2) and the trend as the regressor.

Model 4

In this model, all 30 indices (weighted and un-weighted including interaction indices) were used as discriminating variables in discriminant analysis and two sets of discriminant scores from two discriminant functions were obtained. Forecasting model were fitted taking un-trended yield as the regress and variable and the two sets of discriminant scores (ds1 and ds2) and the trend as the regressor variables.

Model 5

In this model, discriminant function analysis was carried out using the data on the first weather variable spread over 15 weeks using (44^{th} SMW to 6th SMW of next year). Using two sets of discriminant scores obtained from two discriminant function of data on the first weather variable and 15 weeks data of the second variable, discriminant function analysis were again performed and two sets of discriminant scores were obtained (here the discriminating variables will now become 17). Using these two sets of discriminant scores and 15 week data of third variable were again used to carry out discriminant analysis and subsequently two sets of discriminant scores were obtained. This process was continued up to fifth weather variable, and ultimately we got two sets of discriminant scores ds1 and ds2 and forecast model was developed.

Model 6

In this model, discriminant function analysis were carried out using the un-weighted and weighted averages (weather indices) for the first weather variable (here discriminating factors will be only two). Using the two sets of discriminant scores obtained on the basis of first weather variable and un-weighted and weighted averages (weather indices) for the second weather variable, discriminant function analysis were further carried out (here, the discriminating factors will be four). This process was continued up to fifth weather variables, and finally we got two sets of discriminant scores ds1 and ds2 and forecast model was developed.

Comparison and validation of forecast models

Different procedures have been used in the present study for the comparison and the validation of the models developed. These procedures are given below. The six models were compared on the basis of adjusted coefficient of determination (R^2adj), the percent crop yield forecasting using discriminant function analysis

Results and Discussion

The Table 1 indicate that (i) models using correlations based on yield adjusted for trend effect were better than the ones using simple correlations; (ii) inclusion of quadratic term of weather variables did not improve the model except. We need to choose a most appropriate model across the weather variables for studying the crop-weather relationship. The Model-I, II, V and VI Table 1 have come out exactly the same and also performed consistently better than other models in terms of R^2-values.

Forecast Model-D1

$Y = 21.164 - 0.153ds_1 + 0.626ds_2 + 0.508T$

Forecast Model-D2

$Y = 21.681-0.002ds_1 + 0.144ds_2 + 0.421T$

Forecast Model-D3

$Y = 21.928 + 0.192ds_1 + 0.589ds_2 + 0.453T$

Forecast Model-D4

$Y =20.414 + 0.270ds_1 + 0.054ds_2 +0.563T$

Forecast Model-D5

$Y = 21.882 + 0.001ds_1 - 0.001ds_2 + 0.252T$

Forecast Model-D6

$Y = 21.596+0.408ds_1 - 0.365ds_2 + 0.465T$

Table 1: Comparison between actual and forecasted yield of different years of Barabanki District.

Model	Year	Actual yield	Forecast yield	Percent Deviation	RMSE	%SE	R^2	R^2adj.
D1	2009-10	30.79	29.57	3.94	5.69	1.260	85.7	82.8
	2010-11	32.52	31.31	3.70		2.015		
	2011-12	35.03	32.14	8.24		1.586		
D2	2009-10	30.79	30.75	0.11	11.69	2.205	86.3	83.6
	2010-11	32.52	28.89	11.15		2.567		
	2011-12	35.03	29.10	16.91		2.721		
D3	2009-10	30.79	32.78	6.48	5.55	2.053	87.3	84.7
	2010-11	32.52	34.81	7.05		2.629		
	2011-12	35.03	34.71	0.90		2.682		
D4	2009-10	30.79	32.21	4.62	4.94	3.795	81.8	80.8
	2010-11	32.52	34.80	7.02		9.328		
	2011-12	35.03	35.60	1.64		9.840		
D5	2009-10	30.79	31.37	1.91	10.98	2.796	82.9	81.5
	2010-11	32.52	29.34	9.76		6.144		
	2011-12	35.03	29.34	16.22		6.445		
D6	2009-10	30.79	31.84	3.41	3.34	2.330	87.8	85.3
	2010-11	32.52	32.76	0.74		4.560		
	2011-12	35.03	33.41	4.60		4.702		

The forecast models developed under the six procedures along with R^2 and R^2adj are given in Table 2. In all the models, the time trend variable T has been found to be significant at one percent probability level of significance ($P<0.01$). First discriminant score (ds1) has been found to be significant at $P<0.01$ in all the models except in model 5 where it is significant at $P<0.05$. The second discriminant (ds2) has been found to be significant at $P<0.01$ in only model 5. It is evident from the results of the Table-4.2.22 that adjusted coefficient of determination (R^2_{adj}) has been found to be 84.7% for the Model-D3 besides model-D1 followed by the Model-D4 80.8%. While rest of the models namely D2, D5 and D6 are not comparable with the above models as they have varying CV and more RMSE. On the basis of the overall results of the Table- 4.2.22 it can be concluded that the Model-D3 and Model-D1 are the most suitable models among all the models to forecast wheat yield in Barabanki district of Eastern Uttar Pradesh. Hence, a reliable forecast of wheat yield about two and half months before the harvest can be obtained from the Model-D3 and Model-D1. The similarly trend was found Garde *et al*. (2012) [9]

Acknowledgement

I would like to given a lot of thanks to Dr. Vishuddha Nand, Assistant Professor ND University of Agriculture & Technology for very nice cooperation with me during experiment and making the paper.

References

1. Agrawal R, Chandrahas, Kaustav A. Use of discriminant function analysis for forecasting crop yield, Mausam. 2012; 63(3):455-458.
2. Agrawal R Jain RC, MP Singh D. Forecasting of rice yield using climate variables. Ind. J Agri. Sci. 1980; 50(9):680-684.
3. Agrawal R, Jain RC, Jha MP. Models for studying rice crop weather relationship. Mausam. 1986; 37(1):67-70.
4. Agrawal R, Jain Rc, Jha MP. Joint effects of weather variables on rice yields. Mausam. 1983; 32(2):177-81.
5. Bharati RC, Khatri TJ Parikh RK. Pre-harvest forecast models for mango (var. Kesar) yield using weather variables. Gujarat Agricultural University Research Journal. 1999; 24(2):79-83.
6. Cannor WP. Forecasting survival and passage of migratory Juvenil Salmonids. North American Journal of Fishries Management. 1999; 20:651-660.
7. Carter MM, Eisner JB. A statistical methods for forecasting Rainfall over Puerto Rico. Weather and Forecasting, Department of Meteorology, the Florida State University. 1997; 12:515-525.
8. Fisher RA. The influences of rainfall on the yield of wheat at Rothamsted. Philosophical Transaction of Royal Society of London, Series B. 1924; 213: 89-142.
9. Garde YA, Shukla AK, Singh S. Pre-harvest forecasting of rice yield using weather indices in Pantnagar, (Uttarakhand), Int. J Agril. Statistical Sci. 2012; 8(1):233-241.
10. Hendricks WA, Scholl GC. Technique in measuring joint relationship: The joint effects of temperature and precipitation on crop yield. N. Carolina Agric. Exp. Stat. Tech. Bull, 1943, 74.
11. Rai T, Chandrahas, Use of discriminant function of weather parameters for developing forecast model of rice crop. (IASRI Publication), 2000.
12. Sisodia BVS, Yadav RR, Kumar S, Sharma MK. Forecasting of Pre- harvest crop yield using discriminant function analysis of meteorological parameter. Journal of Agrometeorology. 2014; 16(1):121-125.