FORECAST OF CROP YIELD USING DISCRIMINANT FUNCTION ANALYSIS OF WEATHER PARAMETERS

Ria Biswas and Banjul [1]Bhattacharyya

Department of Agricultural Statistics, Bidhan Chandra Krishi Viswavidyalaya,
[1] Faculty of Agriculture, Mohanpur, Nadia, West Bengal, India, PIN – 741252

ABSTRACT

Forecasts of crop yield can provide vital information about commodity markets and are frequently used by growers, industry and government to make decisions. In this study, suitable statistical models have been developed through discriminant function analysis to forecast wet-season rice yield for Nadia district of West Bengal, India. Wet-season rice yield data and weekly weather data have been taken as study materials for twenty one years (1990-2010). Based on yield distribution, time series data on wet-season rice yield have been divided into three sets, viz. congenial, normal and adverse. Considering these three groups as three populations, discriminant function analysis (using weekly meteorological data of crop season) has been carried out. The discriminant scores, obtained from discriminant function analysis have been used as regressor along with time trend. To develop the models first nineteen years data (1990-2008) have been utilized to and last two years (2009-2010) data have been used to forecast wet-season rice yield about two months before harvest. Most of the models give reliable forecast of yield. Among all developed models model-3 has been found to be most suitable.

Keywords: Discriminant function analysis, weather variables, wet-season rice, yield forecast model.

1. Introduction

Agriculture is extremely dependent on climate and variability of crop yield is affected by year-to-year climatic inconsistency (Hoogenboom, 2000; Ogallo *et al.,* 2000). For these uncertainties during crop growing seasons could have significance effects on crop production (Jones et al., 2000). Forecasts of crop yields can provide very important information about product markets and are frequently used by farmers, industry and government for decisions making (Vogel and Bange 1999). For instance, farmers may use forecasts to plan their harvest, storage and distribution strategies in advance. Similarly, industries involved in handling and trading commodities often use information on future harvests to make logistical decisions (Hammer et al. 2001). Various research works have made attempts in past to develop statistical models using weather variables and crop yield data to forecast crop yield (Agrawal *et al.,* 1983; Agrawal *et al.,* 1986). Discriminant function analysis was used to determine yield forecast (Rai and Chandrahas, 2000; Agrawal *et al.,* 2012; Sisodia *et al.,* 2014).

In this paper an attempt has been made to develop suitable statistical model to forecast wet-season rice yield at district level using discriminant function analysis of weekly data on weather variables.

2. Materials and methods

Study area

The study was conducted for Nadia District of West Bengal, Eastern India, which falls under New Alluvial Agro-climatic Zone. In this district, wet-season rice is the main crop and the water requirement for this crop is met mainly from rainfall, although farmers have to depend upon irrigation due to erratic rainfall distribution. Sub-tropical climate with an average annual rainfall of 1500 mm ranging from 1200 to 1690 mm is very often found. The maximum rainfall is received during the months of June to September from South-West monsoon.

Materials

District level wet-season rice yield data for twenty one years (1990-2010) have been collected from District Statistical Handbook (Nadia), published by the Bureau of Applied Economics and Statistics, Government of West Bengal. The weekly rainfall data for active monsoon period was collected for the period (1990-2010) from the India Meteorological Department, Pune. Some other meteorological parameters namely, maximum relative humidity and maximum temperature of Nadia district were collected from Regional Meteorological station, Ali pore, West Bengal for the above said period.

Weather during pre-sowing period is important for establishment of crop; data starting from two weeks before transplanting were included in model development. Further, as the forecast is required well in advance of crop harvest, weather data about one and half month before harvesting was considered. Thus, data on three weather variables namely, rainfall, maximum relative humidity and maximum temperature during 16

E.mail:ria.biswas6@gmail.com

weeks [23rd standard meteorological week (smw) to 36th smw of a year year] were used in the analysis.

Yield forecasting has been done by using regression analysis. First nineteen years' data (1990-2008) were utilized to develop the models and last two years (2009-2010) data were used for validation of the models.

Methodology

Discriminant function analysis is a multivariate technique (Anderson, 1984; Hair *et al.,* 1995, Sharma 1996; Johnson and Wichern, 2006). To identify 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 this technique was used.

Assume that observations are to be classified into k non-overlapping groups on the basis of p variables. The technique identifies linear functions where the variables 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 function 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.

In order to apply discriminant function analysis to model crop yield using weather variables, crop years have been divided into three groups namely congenial, normal and adverse on the basis of crop yield. Data on weather variables were used to develop linear discriminant functions and the discriminant scores were obtained for each year. These scores were used alongwith year as regressor in developing the forecast model. In the present study the number of groups was three and number of weather variables was three, therefore only two discriminant functions were sufficient for discriminating a crop year.

Model 1

Total time starting from two weeks before transplanting to the time of harvest (*i.e.*, 16 weeks starting from 23rd swm) was divided into four phases where each phase consists of different numbers of weeks. For each phase and each weather variable simple average of the weather data in the different weeks within the phase was obtained. This way for each phase three average weather variables were obtained. Taking these three average weather variables, phase wise discriminant function analysis was carried out and entire weather variables data were converted to three discriminant scores for each phase in each year. Backward regression technique was carried out using these twelve discriminant scores and time trend as regressors and the model was fitted as follows:

$$Y = \alpha + \sum_{l=1}^{3}\sum_{m=1}^{4} \beta_{1m} ds_{lm} + \beta_t T + \varepsilon \qquad (1)$$

where, Y is the crop Yield

α is the intercept of the model

β_{lm}'s (l=1,2,3; m= 1,2,…, 4) and β_t are the regression coefficient,

ds_{lm} is the ith discriminant scores in mth phase.

T= the trend variable (year) and

ε = error~ $N(0, \sigma^2)$

This model was based on significant discriminant scores in some phases selected by backward regression and as such suffers from the drawback that it was not based on complete data. Further, it gave equal weightage to the weather variables in all the weeks within the phases as it used phase-wise simple averages. But, weather variables affect the crop differently during different phases of development. Thus weightage of different weather variables was needed.

Model 2:

In this model an index was developed as weightage total for each weather variable over all the weeks up to the time of forecast. Weights being the correlation coefficients between detrended yield and respective weather variable in different weeks. The total data of three weather variables in sixteen weeks were converted into three weather indices. Discriminant function analysis was carried out on these three weather indices and two discriminant scores were obtained for each year. To develop forecast model through regression technique these two discriminant scores and time trend were utilized. The form of model considered is as follows:

$$Y = \alpha + \beta_1 ds_1 + \beta_2 ds_2 + \beta_3 T + \varepsilon \qquad (2)$$

where, symbols α, β_i's, ds_1, ds_2, T and ε was as defined earlier.

This model utilizes the complete data over all 16 weeks and also considers relative importance of weather variables in different weeks as against model-1 in which equal importance in different weeks was assigned.

Model 3

In this model, three weighted and three un-weighted weather indices of three weather variables were used as discriminating variables. From two discriminant functions two set of scores were obtained. Taking the yield as dependent variable and two sets of scores (*ds1* and *ds2*), trend as the regressor forecasting model was fitted. The form of this model was same as model 2.

Comparison and validation of forecast models

The three models were compared on the basis of adjusted coefficient of determination (R^2_{adj}) which is as follows:

$$R_{adj}{}^2 = 1 - \frac{SS_{res}/(n-p)}{SS_t/(n-1)} \qquad (3)$$

where, $SS_{res}/(n-p)$ is the residual mean square and $SS_t/(n-1)$ is the total mean square.

From the fitted models, wet-season rice yield forecasts for the years 2009 to 2010 were obtaind and forecasts were compared on the basis of Root Mean Square Error (RMSE).

$$RMSE = \left[\frac{1}{n}\sum_{i=1}^{n}(O_i - E_i)^2\right]^{\frac{1}{2}} \qquad (4)$$

where, O_i and E_i are the observed and forecast values of wet-season rice yield respectively and n is the number of years for which forecasting had been done.

3. Results and discussions

The forecast models developed under the three approaches along with R^2 and R^2_{adj} are presented in table 1. In all the models, the time trend variable *T* was found significant. Adjusted coefficient of determination (R^2_{adj}) varied between 0.828 to 0.943 in different models, and was found to be maximum of 94 percent (0.943) being in model3.

RMSE was computed on the basis of yield forecasts for the years 2009 to 2010 (Table 2). The RMSE varied from a minimum of 22.31 in model-3 to a maximum of 55.58 in model-1.

Table 1: Wet-season rice yield forecast models

Model		Forecast regression equation						R^2	R^2_{adj}
1 Yield	=	1485.89 –	$78.57ds_{14}$+	$169.84ds_{22}$-	$139.65ds_{24}$	$–244.13ds_{32}$ +	63.14T	0.854	0.828
		(0.049)	(0.132)	(0.192)	(0.036)			(0.060)	(0.003)
2 Yield	=	1855.20 +	$63.57ds_1$ +	$96.08ds_2$ +	29.01T			0.934	0.920
		(0.038)	(0.275)	(0.126)	(0.045)				
3 Yield	=	20779.73 -	$42.12ds_1$ +	$334.38ds_2$ +	51.13T			0.956	0.943
		(0.013)	(0.386)	(0.042)	(0.049)				

Note: Figures in brackets denote the level of significance of regression coefficients

Table 2: Actual and forecasts of wet-season rice yield (Kg/ha)

Crop year	Actual yield		Forecasted yield	
		Model-1	Model-2	Model-3
2009	2426	2533.02	2501.02	2479.08
2010	2272	2305.67	2289.39	2245.48
	RMSE	55.58	30.21	22.31

4. Conclusion

On the basis of the overall results of the Table 1 and 2, it can be concluded that the proposed model 3 is the most suitable among all the models (as RMSE value is minimum and R^2_{adj} is maximum for model 3) to forecast Wet-season rice yield in Nadia district. Hence, a reliable forecast of Wet-season rice yield about one and half months before the harvest can be obtained from proposed model 3.

References

Agarwal, R., Chandrahas and Aditya, K. (2012): Use of discriminant function analysis for forecasting crop yield. *Mausam*, **63(3)**: 455-58.

Agrawal, R., Jain, R.C. and Jha, M.P. (1983): Joint effects of weather variables on rice yield. *Mausam*, **34(2)**: 189-94.

Agrawal, R., Jain, R.C., Jha, M.P. (1986): Models for studying rice crop-weather relationship. *Mausam*, **37(1)** : 67-70.

Anderson, T.W. (1984): An introduction to applied Multivariate Statistical Analysis, John Wiley & Sons Inc. New York.

District Statistical Handbook (Nadia) The Bureau of Applied Economics and Statistics, Government of West Bengal.

Hair, Joseph F., Anderson, Rolph, E., Tatham, Ronald L., Black, and William C. (1995): Multivariate Data Analysis with Readings, Prentice Hall, Inc New Jersey.

Hammer, G.L., Hansen, J.W., Phillips, J.G. (2001): Advances in application of climate prediction in agriculture. *Agric System*, **70** : 53. DOI: 10.1016/S0308-521X(01)00058-0.

Hoogenboom, G. (2000): Contribution of agrometeorology to the simula-tion of crop production and its application. *Agric. Meteorol*, **103**: 137–57.

Johnson, A. Richard and Wichern, W. Dean. (2001): Applied Multivariate Statistical Analysis, 3rd edition, Prenctice-Hall of India Private Limited, New Delhi-110001, 314-336.

Jones, J. W., Hansem, J. W., Royce F. S. and Messina, C. D. (2000): Potential benefits of climate forecasting in Agriculture. *Agric. Ecosyst. Environ*., **82**: 169–84.

Rai, T. and Chandrahas (2000): Use of discriminant function of weather parameters for developing forecast model of rice crop, IASRI Publication, New Delhi.

Ogallo, L. A., Boulahya, M. S. and Keane, T. (2000): Applications of seasonal to interannual climate predictions in agricultural planning and operations. *Agric. Meteorol*. **103**: 159–166.

Sharma, S. (1996): Applied Multivariate Techniques, John Wiley & Sons Inc. New York.

Sisodia, B.V.S.; Yadav, R.R.; Kumar, S. and Sharma, M.K. (2014): Forecasting of pre-harvest crop yield using discriminant function analysis of meteorological parameters. Journal of Agrometeorology, 16(1): 121-125.

Vogel F, Bange G. (1999): Understanding crop statistics Available online at www.usda.gov/nass/nassinfo/pub1554.htm .