DEVELOPMENT OF AGRO-METEOROLOGICAL WHEAT YIELD MODEL USING PRINCIPAL COMPONENT TECHNIQUE FOR WESTERN ZONE IN HARYANA

Megha Goyal* and Urmil Verma
* Teaching Associate
Department of Mathematics, Statistics and Physics
CCS Haryana Agricultural University, Hisar, India

Abstract-An attempt has been made in this paper to estimate the yield of wheat crop using principal component technique. Fortnightly weather parameters viz., average maximum temperature, average minimum temperature and accumulated rainfall calculated over critical growth phases of wheat crop and trend based yield were used for the model building on agroclimatic zone basis in western zone comprises districts Hisar, Bhiwani, Sirsa and Fatehbad (Haryana). Zonal yield models provided considerable improvement in district-level yield prediction by showing good agreement with Department of Agriculture (DOA) yields. The crop yield estimates on the basis of developed models may be obtained 4-5 weeks in advance of the harvest time. The percent deviations of fitted model for all districts with in western zone are lies within acceptable limits.

Key words: Eigen values, Linear time trend, percent deviation and principal component score.

Introduction

Timely and effective forecasts of the crops are vital for an agrarian economy. Crop yield forecasts are important for advance planning, formulation and implementation of policies related to the crop procurement, distribution, price structure and import export decisions etc. These are also useful to farmers to decide in advance their future prospects and course of action. Thus reliable and timely pre-harvest forecasting of crop yield is very important. Crop yield is a complicated process which is governed by number of factors. The main factors affecting crop yield are weather variables and agricultural input. Use of these factors form one class of models for forecasting crop yields. Weather plays a dominant role in crop growth and a development and hence can be conveniently used as indicator of change in crop yield modeling. The relationship between crop yield and weather parameters is generally carried out with help of multiple regression models. These models are generally employed for making quantitative crop yield forecast on operational basis. Bal (2004), Bazgeer *et al*. (2007), Esfandiary *et al*. (2009), Gill *et al*. (2015) and Walker (1989) have used a series of weather predictors for crop yield forecasting.. Azfar et al. (2015) and Jain *et al*. (1984) used principal component analysis for mustard and sorghum yield forecast models. Goyal and Verma (2015) have used agromet/spectral indices in context of pre-harvest yield forecasting of different crops in Haryana. Keeping in view the importance of the subject matter, an attempt has been invariably made to develop trend-agromet wheat yield models on agroclimatic zone basis in Hisar (Haryana) using principal component technique. The emphasis has been given to see the forecasting performance of the zonal yield models for district-level yield estimation during the period (s) of model development and model testing.

Data description

The Department of Agriculture (DOA) wheat yield data published by Bureau of Economics and Statistics were compiled for period 1980-81 to 2013-14 for Hisar, Sirsa and Bhiwani districts and 1997-98 to 2013-14 of Fatehbad and used for computing the trend based yield. i.e. Tr = a+br, where Tr = Trend yield(q/ha), a = Intercept, b = Slope and r = Year. The weather data for 33 years were collected from IMD and different meteorological observatories in Haryana. The weather data of maximum temperature, minimum temperature and rainfall were used for the purpose. Weather data starting from 1st fortnight of November to 1 month before harvest have been utilized for the model building (crop growth period: 1st November to 15th April). The average weather value calculated over 1st to 15th November gives 1st fortnight weather parameter(s), average value calculated over 16th to 30th November provides 2nd fortnight weather parameter(s) and proceeding in the next year, the average weather value calculated over 16th to 31st March and 1st to 15th April gives the 10th and 11th fortnight weather. The various fortnightly meteorological parameters have been computed as follows:

$$\text{Average Maximum Temperature (TMX)} = \frac{\sum_{i=1}^{15} \text{TMXi}}{15}$$

$$\text{Average Minimum Temperature (TMN)} = \frac{\sum_{j=1}^{15} \text{TMNj}}{15}$$

$$\text{Accumulated Rainfall (ARF)} = \sum_{k=1}^{15} \text{ARF}_K$$

where TMX_i = i^{th} day maximum temperature

TMN_j = j^{th} day minimum temperature

ARF_K = k^{th} day rainfall

i,j,k = daily meteorological data

This helps in identifying critical growth phases and meteorological parameters influencing the final wheat yield. Since climatic data from adequate number of stations were not available, districts having equable climatic conditions were grouped into different agro-climatic zones based on their agro-climatic conditions i.e. for western-zone: Sirsa, Fatehbad, Hisar & Bhiwani districts were grouped. The fortnightly parameter(s) form 1^{st} November to 1 month before harvest during the crop growth period viz., Tmx_1, Tmx_2, ...,Tmx_9, Tmn_1, Tmn_2,...,Tmn_9 and Arf_1, Arf_2, ..., Arf_9 have been used to develop the zonal yield models.

Statistical procedure

Regression analysis is one of the most widely used statistical techniques for modelling multifactor data. The inference drawn from a multiple regression model often depends on the estimates of the individual regression coefficients. However in some situations, the problem of multicollinearity exists when there are near linear dependencies among the regressors. Principal component technique offers considerable improvement over ordinary least squares estimates in the presence of multicollinearity.

PC technique has been carried out to develop the zonal yield model. The procedure consists of finding the eigen roots and eigen vectors of the correlation matrix of explanatory variables. Principal components Pi (i=1,2,….) were obtained as P = kX , where P and X are the column vectors of transformed and the original variables, respectively and k is the matrix with rows as the characteristic vectors of the correlation matrix R. The variance of Pi is the ith characteristic root λi of the correlation matrix R; λs were obtained by solving the equation |R- λI| =0. For each λ, the corresponding characteristic vector k was obtained by solving |R-λI| k=0

Principal component method has been used for extraction of factors which consists of finding the eigen values and eigen vectors. The most frequently used convention is to retain the components whose eigen values are greater than one. Kaiser(1958) also suggested the dropping of components having eigen roots less than 1. The principal component scores can be used as new regressors in multiple regression analysis for selecting the suitable yield models.

Results and Discussion

Zonal wheat yield models based on weather data of 1980-81 to 2009-10 have been developed within the framework of principal component analyses. The validity of the models have been checked for the post sample period 2010-11 to 2013-14 for eastern zone of Haryana. Under this study, first 10 eigen values (shown below along with their percent of variation explained) of correlation matrix of explanatory variables (weather parameters) suggested 10 factor solution. However, the remaining components accounted for a smaller amount of total variation. Hence, those components were not considered to be of much practical significance. Eigen vectors being the weights were used to compute PC scores.

Eigen values and variance (%) explained by different principal components

Eigen value	4.52	3.45	2.77	2.59	2.09	1.79	1.62	1.34	1.10	1.07
Percent variance explained	16.75 (18.90)	12.79 (13.85)	10.25 (11.19)	9.61 (10.39)	7.74 (07.67)	6.63 (07.02)	5.99 (05.48)	4.96 (04.15)	4.08 (03.92)	3.96 (03.32)

For quantitative forecasting, regression models *via* step-wise regression (Draper and Smith, 2003) were fitted, considering PC scores as regressors and total yield (Y) as dependent variable. The models finalized were used to obtain the yield estimates of the wheat under consideration. The selected zonal yield model along with their standard error and R^2 is given below:

$$Yield_{est} = -1.78 + 1.05 * Tr - 1.10 * PC_6 + 1.02 * PC_{10}$$

R^2= 0.827 adj.R^2 =0.822 SE= 2.93

where $Yield_{est}$ - Model predicted yield

T_r - Linear time trend based yield

PC_i - i^{th} principal component score (i = 1,2,…,10)

SE - Standard error of yield estimate

R^2 - Coefficient of determination

The model predicted yields along with observed yields and per cent relative deviations for all districts with in western zone are given in table below.

District-specific wheat yield estimates alongwith percent deviations from DOA yield(s) using fitted models.

Districts/ Years	Hisar			Sirsa		
	Obs. Yield (q/ha)	Fitted Yield (q/ha)	RD(%)	Obs.Yield (q/ha)	Fitted Yield (q/ha)	RD(%)
2010-11	46.22	48.02	-3.90	51.3	47.67	7.08
2011-12	50.98	49.92	2.07	53.57	49.56	7.49
2012-13	42.73	48.48	-13.45	48.42	48.10	0.67
2013-14	44.77	49.17	-9.83	53.47	48.77	8.79
Av. abs. Dev	7.31			6.01		

Districts/ Years	Bhiwani			Fatehabad		
	Obs. Yield (q/ha)	Fitted Yield (q/ha)	RD(%)	Obs.Yield (q/ha)	Fitted Yield (q/ha)	RD(%)
2010-11	44.65	41.79	6.41	50.81	48.71	4.13
2011-12	43.06	43.67	-1.41	54.72	50.54	7.64
2012-13	40.55	42.20	-4.06	46.81	49.02	-4.72
2013-14	42.28	42.85	-1.34	53.18	49.63	6.67
Av. abs. Dev	3.31			5.79		

Percent Relative Deviation (RD%) = 100 × [(observed (obs.) yield – fitted yield) / observed yield]

The zonal weather-yield models have been developed to achieve district-level pre-harvest wheat yield estimation in Haryana. The predictive performance(s) of the fitted models were observed in terms of the percent deviations of wheat yield forecasts in relation to real time wheat yield(s). The estimated yield(s) from the selected models indicated good agreement with DOA wheat yield estimates by showing 3-8 percent average absolute deviations for most of the districts

References

- Azfar, M., Sisodia, B. V. S., Rai, V. N. and Devi, M.(2015). Pre-harvest forecast models for rapeseed and mustard yield using principal component analysis of weather variables, *Mausam*, **66**(4): 761-766.
- Bal, S.K., Mukherjee, J., Mallick, K. and Hundal, S.S. (2004). Wheat yield forecasting models for Ludhiana district of Punjab state. J. Agrometeorology. 6: 161-65.
- Bazgeer, S., Gh. Kamali and A. Mortazavi (2007). Wheat yield prediction through agrometeorological indices for Hamedan, Iran.*Biaban*, **12**, 33-38.
- Draper, N.R. and Smith, H. (2003). Applied Regression Analysis. 3rd edition, *John Wiley & Sons. New York.*
- Esfandiary, F., G. Aghaie and A. D. Mehr (2009). Wheat yield prediction through agro meteorological indices for Ardebil district. *World Academy of Science: Engineering and Technology*, **49**, 32-35.
- Gill, K.K. , Sandhu, S.S. , Kaur, P., Babuta, R. and Bhatt, K. (2015). Wheat Yield Prediction using Weather based Statistical Model in Central Punjab. *Journal of Agricultural Physics.* 15(2), 157-162.
- Goyal, M. and Verma, U. (2015). Development of weather-spectral models for pre-harvest wheat yield prediction on agro-climatic zone basis in Haryana, *International J. of Agricultural and Statistical Sciences*, **11**(1): 73-79.
- Jain, R.C., Sridharan, H and Agarwal, R. (1984). Principal component technique for forecasting of sorghum yield. Indian J. Agric. Sci. 54 (6): 467-70.
- Walker, G. K. (1989). Model for operational forecasting of western Canada wheat yield. Agric. For. Meteor. 44: 339-51.