International Journal of Applied Mathematics
Statistical Sciences (IJAMSS)
ISSN(P): 2319-3972; ISSN(E): 2319-3980
Vol. 7, Issue 3, Apr -May 2018; 29-40

International Academy of Science, Engineering and Technology
Connecting Researchers; Nurturing Innovations

PRE-HARVEST WHEAT YIELD FORECAST THROUGH AGRO-METEOROLOGICAL INDICES FOR NORTHERN REGION OF HARYANA

Megha Goyal

Department of Mathematics, Statistics and Physics, CCS Haryana Agricultural University, Hisar, Haryana, India

ABSTRACT

Parameter estimation in statistical modeling plays a crucial role in the real world phenomena. Several alternative analyses may be required for the purpose. An attempt has been made in this paper to assess the impact of weather variables for district-level wheat yield estimation in the Northern region (Haryana). Phase wise weather data and trend based yield was used for developing the zonal trend-agro meteorological (agromet) models within the framework of multiple linear regression and principal components analysis. The results indicate the possibility of district-level wheat yield prediction, 4-5 weeks ahead of the harvest time. Zonal weather models had the desired predictive accuracy and provided considerable improvement in the district-level wheat yield estimates. The principal component analysis offers a considerable improvement over least squares method in the presence of multicollinearity. The estimated yield(s) from the selected models indicated good agreement with State Department of Agriculture (DOA) wheat yields by showing 2-10 percent average absolute deviations in most of the districts except for the Panchkula district.

KEYWORDS: *Linear Time Trend, Eigen Value, Eigen Vector, Weather Variables, Multicollinearity, Principal Component Score*

Article History

***Received:** 24 Apr 2018* | ***Revised:** 03 May 2018* | ***Accepted:** 12 May 2018*

INTRODUCTION

Predictive modeling is a collection of techniques having in common the goal of finding a relationship between a response and various predictors with the idea of measuring future values of those predictors and inserting them into the derived relationships to predict future values of the target variable. The classical linear regression model in this direction is an important statistical tool, but its use is limited to those settings where the normal distribution is valid and the assumption of a linear function relating the response to the predictors is given.

Reliable, accurate and timely information on types of crop grown and their acreages, crop yield and crop growth conditions are vital components for planning efficient management of natural resources. Crop productivity is affected by technological change and weather variability. It can be assumed that the technological factors will increase yield smoothly through time and therefore, a year or some other parameter of time can be used to study the overall effect of technology on yield. Weather variability both within and between seasons is the uncontrollable source of variability in yield. Weather variables affect the crop differently during different stages of development. This increases the number of variables in the model and in turn, a largenumber of constants are to be evaluated from the long time-series data for precise estimation of the parameters. Thus, a technique based on the relatively smaller number of manageable parameters and at the same time, taking care of entire weather distribution may solve the problem. Keeping in view the importance of the subject matter, multiple regression analysis, and Principal Component Analysis was carried out for wheat yield estimation in Haryana.

Wheat is one of the most important cereal crops in India as it forms a major constituent of the staple diet of a large part of the population. India is the second largest producer among wheat growing countries of the World (Source: www.mapsofindia.com/indiaagriculture). Haryana occupies the third place for wheat production among the various states in India (Source:www.agricoop.nic.in/statistics). Haryana is self-sufficient in food grains production and also one of the top contributors of food grains to the central pool. India is the second largest producer among wheat growing countries of the World (Source: www.mapsofindia.com/ India agriculture). Wheat occupies the foremost position followed by rice, not only in terms of acreage and production but also in the versatility in adopting different soils and climatic conditions. Some similar studies concern these works are, Azfar *et al.* (2015) used principal component analysis for rapeseed and mustard yield forecast models for Faizabad district of U.P. (India). Chandran and Prajneshu (2004), Bazgeer *et al.* (2007) and Esfandiary *et al.(*2009), Mehta *et al.* (2010) etc. have used weather data in the context of crop yield prediction. Verma *et al.* (2011, 16) and Goyal and Verma (2015) have used agromet/spectral indices in context of pre-harvest yield forecasting of different crops in Haryana.

Data Description

Haryana state comprised of 21 districts is situated between 74^o 25’ to 77^o 38’ E longitude and 27^o 40’ to 30^o 55’ N latitude. The total geographical area of the state is 44212 sq. km. Wheat crop is grown in all the districts of the state with varying density. In this research article, the yield estimate for northern zone comprises Ambala, Panchkula, Yamuna Nagar and Kurukshetra districts have been carried out. Time-series yield data for the past 30 years (i.e. 1978-79 to 2007-08) of wheat crop of districts of northern zone of Haryana published by Bureau of Economics and Statistics were used for computing linear yield trend i.e. T_r = a+br, where T_r = Trend yield(q/ha), a = Intercept, b = Slope and r = Year. The meteorological data for the same 30 years were collected from India Meteorological Department (IMD), Delhi and different meteorological observatories in Haryana, India. The weather data of maximum temperature, minimum temperature and rainfall were used for the purpose. Since climatic data from an adequate number of stations, were not available, districts having equable climatic conditions have been grouped into a zone.

Wheat crop is sown in the month of November and harvested in the month of April. The early three weeks of growing season corresponds to the early growth phase. This includes the period from sowing to the emergence and initial growth of the crop. Grand growth phase includes tillering stage, late jointing stage, and flowering stage. Maturity phase includes the ripening stage of the crop. The derived meteorological indices i.e., Growing Degree Days (GDD) and Temperature Difference (TD) are computed as follows; $TD=\Sigma[T_{max}-T_{min}]$ and $GDD=\Sigma[\{(T_{max}+T_{min})/2\}-T_b]$; where T_{max}= maximum temperature, T_{min} = minimum temperature, T_b= Base temperature (5^o). To integrate GDD, TD and weekly accumulated rainfall (ARF) over different growth phases; total wheat growth period has been divided into seven phonological stages, viz. i) Crown Root Initiation Stage (MNW 44-46), ii) Tillering Stage (MNW 47-49), iii) Jointing Stage (MNW 50-52), iv)

Flowering Stage (MNW 1-3), v) Milking Stage (MNW 4-6), vi) Dough Stage (MNW 7-9) and vii) Maturity stage (MNW 10-14); MNW stands for meteorological week numbers. This helps in identifying critical growth phases influencing the final wheat yield.Yield is a complicated trait which is governed by a number of factors. The main factors affecting the crop yield are agricultural inputs and weather variables. The work was carried out to develop forecasting models on agro-climatic zone basis by combining the data of various districts within a zone. Thus, a longer data series could be obtained in a relatively shorter period (i.e. inclusion of 30 years weather and yield data for each district(s) within the zone) and that provided the basis to use multivariate statistical analyses. The focus was on the comparison of district- level yield estimates obtained under two different procedures by evaluating the forecasting performance of the zonal trend-agromet-yield models during the period of model development (1978-79 to 2004-05 i.e. 27 years) and the model testing period (2005-06 to 2007-08 i.e. 3 post-sample years). Multiple linear regression and principal component analysis were used to achieve the targeted objective.

Statistical Procedure

The Standard Linear Regression model considered may be written in the form $Y=Xb+\varepsilon$; where Y is an ($n\times l$) vector of observations, X is an ($n\times p$) matrix of known form, b is a ($p\times l$) vector of parameters, ε is an ($n\times l$) vector of errors with the assumptions $E(\varepsilon)=0$ and $V(\varepsilon)= I\sigma^2$**,** so the elements of $\boldsymbol{\varepsilon}$ are uncorrelated. Regression models via stepwise regression analysis (Draper and Smith, 2003) were fitted using statistical software SPSS. The selected zonal yield models are given in Table 2.

The principal component analysis offers a considerable improvement over least squares method in the presence of multicollinearity. The presence of multicollinearity among explanatory variables can lead to unstable regression estimates and erroneous results. Bartlett's test of sphericity is used to test the null hypothesis that the variables in the population correlation matrix are uncorrelated. The observed significance level is 0.0000. It is small enough to reject the hypothesis i.e. this test has to be significant for multicollinearity. The result shown in Table 3 confirms the presence of multicollinearity among the weather indices used in regression analysis. Thus, Bartlett's test of sphericity gives the confidence to proceed with principal component analysis.

The principal component method was used for the extraction of factors which consists of finding the eigenvalues and eigenvectors. Principal components P_i (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 P_i is the i^{th} characteristic root λ_i of the correlation matrix R; λ_s were obtained by solving the equation $|R- \lambda I| =0$. For each λ, the corresponding characteristic vector k was obtained by solving $|R-\lambda I|\ k=0$

RESULTS AND DISCUSSIONS

Under this study, first 7 eigenvalues (Table1) of the correlation matrix of explanatory variables (weather parameters) suggested 7-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. Eigenvectors being the weights were used to compute PC scores.

Table 1: Eigen Vectors, Eigen Values, Variance (%) and Cumulative (%) of Total Variance Explained by Different Principal Components of Northern Region

Weather Indices	Components						
	1	2	3	4	5	6	7
ARF_1	-0.09	-0.23	0.04	0.17	0.03	-0.02	-0.47
TD_1	0.19	0.16	0.14	-0.07	-0.01	-0.04	0.09
GDD_1	0.04	-0.19	-0.06	-0.01	0.25	0.39	0.12
ARF_2	-0.19	0.17	-0.13	0.01	-0.09	0.11	-0.01
TD_2	0.22	0.03	0.08	-0.11	-0.02	-0.01	-0.35
GDD_2	-0.06	-0.26	0.00	0.00	0.12	0.30	0.11
ARF_3	-0.15	0.14	-0.11	-0.11	-0.09	0.00	0.20
TD_3	0.20	-0.07	0.11	0.16	-0.03	-0.02	-0.04
GDD_3	0.09	-0.05	0.22	-0.24	-0.04	0.25	0.38
ARF_4	0.11	-0.03	-0.13	0.20	-0.33	0.08	0.32
TD_4	0.02	0.04	0.19	0.33	0.17	-0.27	0.29
GDD_4	-0.02	0.13	0.11	0.28	0.38	0.08	0.09
ARF_5	-0.05	-0.12	0.29	-0.28	-0.09	-0.16	-0.05
TD_5	0.13	0.16	-0.14	0.11	-0.10	0.33	-0.11
GDD_5	-0.02	0.24	0.10	-0.08	0.14	0.23	-0.32
ARF_6	-0.07	0.09	0.30	0.13	-0.18	0.27	-0.16
TD_6	0.17	-0.06	-0.29	-0.01	0.04	-0.09	-0.17
GDD_6	0.05	0.08	-0.12	-0.25	0.39	-0.09	0.07
Eigen value	**3.70**	**2.72**	**2.09**	**1.86**	**1.64**	**1.47**	**1.08**
Percent variance explained	**20.53**	**15.09**	**11.62**	**10.35**	**9.11**	**8.18**	**5.98**
Cumulative Percentage of total variance	**20.53**	**35.62**	**47.25**	**57.59**	**66.70**	**74.88**	**80.86**

The analysis was carried out to see the impact of weather parameters for pre-harvest wheat yield forecasting on the agro-climatic zone basis in Haryana state. The developed zonal models are based on time-series data of weather parameters from 1978-79 to 2004-05 and trend based yield as well, however, the data from 2005-06 to 2007-08 were used for validation of the models. Data for the last one month of wheat crop season were excluded from the analysis, as the idea behind the study was to predict yield(s) about one month in advance of the actual harvest. The multiple linear regression and principal component analysis were used to obtain different zonal trend-agromet-yield equations. The best subsets of weather variables were selected using stepwise regression method in which all variables were first included in the model and eliminated one at a time with decisions at any particular step conditioned by the result of the previous step. The best-supported weather variables were retained in the model if they had the highest adjusted adj. R^2 and lowest standard error (SE) of the estimate at a given step. The selected zonal trend-agromet-yield models are as follows:

Table 2: Selected Zonal Trend-Agromet Wheat Yield Models

<u>Model-I</u> Model Variable	Coefficients		<u>Model-II</u> Model Variable		Coefficients
Constant	c_1	16.44	Constant	c_2	0.30
T_r	a_1	1.056	T_r	b_1	0 .98
TD_1	a_2	-0.30	PC_3	b_2	0.72

Table 2: Contd.,

GDD_3	a_3	0.21	PC_5	b_3	-0.60
GDD_6	a_4	- 0.26	PC_7	b_4	-0.62
Adj.R^2= 0.87	SE= 2.53		Adj. R^2= 0.86		SE= 2.62

$$Yield_{est.}\ (Model\text{-}1) = \{c_1 + (a_1 \times T_r) + (a_2 \times TD_1) + (a_3 \times GDD_3) + (a_4 \times GDD_6)\}$$

$$Yield_{est.}\ (Model\text{-}2) = \{c_2 + (b_1 \times T_r) + (b_2 \times PC_3) + (b_3 \times PC_5) + (b_4 \times PC_7)\}$$

Regression Models Model-1: Weather parameters and trend yield as repressors, **Model-2:** Principal scores and trend yield as repressors,

Where,

$Yield_{est.}$: Model predicted yield (q/ha)**Tr:** Trend yield (q/ha)**TD:** Temperature difference **GDD:** Growing degree days **ARF:** Accumulated rainfall (1,2,3,……,7 refer to different phases) **PC_i :** i^{th} principal component score (i = 1,2,3,4,5,6,7)**SE:** Standard error of the estimate **Adj. R^2:** Adjusuted Coefficient of determination

Table 3: Bartlett's Test of Sphericity for Checking Multicollinearity

Approx. Chi-Square	989.64
df	153
Sig.	0.00

Table 4: District-Specific Wheat Yield Estimates Along with Percent Deviations from DOA Yield(s) Using Fitted Models

Districts/ Years	**Ambala**				
	DOA Yield(q/ha)	Fitted Yield(q/ha)	RD(%)	Fitted Yield(q/ha)	RD(%)
		Model-1		**Model-2**	
2005-06	37.89	33.91	-10.49	37.51	-0.99
2006-07	38.06	39.12	2.78	38.34	0.74
2007-08	39.82	38.06	-4.42	38.87	-2.38
Districts/ Years	**Kurukshetra**				
	DOA Yield (q/ha)	Fitted Yield(q/ha)	RD(%)	Fitted Yield(q/ha)	RD(%)
		Model-1		Model-2	
2005-06	45.82	42.47	-7.32	45.47	-0.76
2006-07	46.72	47.72	2.14	46.35	-0.80
2007-08	47.72	46.71	-2.11	46.93	-1.66
Districts/ Years	**Yamunanagar**				
	DOA Yield (q/ha)	Fitted Yield (q/ha)	RD(%)	Fitted Yield (q/ha)	RD(%)
		Model-1		Model-2	
2005-06	36.82	34.00	-7.65	37.60	2.11
2006-07	41.55	39.06	-5.98	38.29	-7.85
2007-08	37.71	37.86	0.39	38.69	2.59
Districts/ Years	**Panchkula**				
	DOA Yield (q/ha)	Fitted Yield (q/ha)	RD(%)	Fitted Yield (q/ha)	RD(%)
		Model-1		**Model-2**	
2005-06	18.81	18.87	0.30	23.51	24.96
2006-07	23.75	23.06	-2.89	23.39	-1.50
2007-08	25.75	20.99	-18.47	22.99	-10.72

Regression models via step-wise regression have been fitted, taking yield as the dependent variable and weather indices and trend yield as the regressors. Bartlett's test of sphericity was used to confirm the presence of multicollinearity among the weather indices used in regression analysis. The significance of the test gave the confidence to proceed with principal component analysis. Further, trend yield along with principal component scores was used as regressors to obtain the suitable zonal yield model. Looking at forecast figures, the zonal models based on principal component scores have been retained for district-level wheat yield prediction in the state. The estimated yield(s) from the selected models indicated good agreement with State Department of Agriculture (DOA) wheat yields by showing 2-10 percent average absolute deviations in most of the districts except for the Panchkula district.

REFERENCES

1. *Azfar, M., Sisodia, B. V. S., Rai, V. N. and Devi, M. (2015). Pre-harvest forecast models for rapeseed & mustard yield using principal component analysis of weather variables, Mausam,* ***66****(4), 761-766*
2. *Bazgeer, S., Kamali, G.H. and Mortazavi, A. (2007). Wheat yield prediction through agro-meteorological indices for Hamedan. Iran. Biaban, 12 : 33-38.*
3. *Chandran, K.P. and Prajneshu (2004). Modelling effect of sunshine and temperature on rice (Oryza sativa) tiller production using non parametric regression. Ind. J. Agri. Sci. 74(10) : 563-65.*
4. *Jog, Sr, And Sp Hande. "On Some New Polynomial And Its Eigen Values."*
5. *Draper, N.R. and Smith, H. (2003). Applied Regression Analysis. 3rd edition, John Wiley & Sons. New York.*
6. *Esfandiary, F., Aghaie, G. and Mehr, A. D. (2009). Wheat yield prediction through agro meteorological indices for Ardebil district. World Academy of Science. Engineering and Technology, 49 : 32-35.*
7. *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.*
8. *Mehta, S.C., Pal, S. and Kumar, V. (2010). Weather based models for forecasting potato yield in Uttar Pradesh. Project Report: IASRI/P.R.-01/2010, IASRI, New Delhi.*
9. *Verma, U., Dabas, D. S., Hooda, R. S., Kalubarme, M. H., Yadav, M. and Sharma, M. P. (2011). Remote Sensing Based wheat acerage and spectraltrend-agrometeoro-logical yield forecasting: Factor Analysis Approach. Society of Statistics, Computers and Applications,* ***9****(1&2), 1-13.*
10. *Najah Ali Abidalkarem, Effect of Farm Possession on the Economic Efficiency of Wheat Yield Production in the Irrigated Regions of Iraq (Al-Azezia, Khan Bne-Saad), Case - Study, International Journal of Humanities and Social Sciences (IJHSS), Volume 6, Issue 3, April-May 2017, pp. 47-54*
11. *Verma, U., Piepho, H. P., Goyal, A., Ogutu, J. O. and kalubarme, M. H. (2016). Role of climatic variables and crop condition term for mustard yield prediction in Haryana (India). International J. of Agricultural and Statistical Sciences,* ***12****(1), 45-51.*