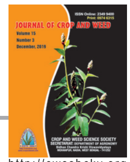
Journal of Crop and Weed, 18(3): 245-250 (2022)

ISSN- O : 2349 9400 ; P : 0974 6315



http://cwssbckv.org
www.cropandweed.com

Development of zonal wheat yield forecast models through Principal Component Analysis

CHETNA, [1]M. DEVI, [2]J. MOHAMMED, [3]Y. S. RAGHAV AND [*4]P. MISHRA

Deptt. of Mathematics & Statistics, [1] Deptt. of Agricultural Economics, CCSHAU, Hisar
[2]Department of General/Liberal Studies, Faculty of Business and Management Studies, Koforidua Technical University, Koforidua, Ghana;
[3]Department of Mathematics, Faculty of Science, Jazan University, KSA,
[4]College of Agriculture, J.N.K.V.V., Rewa, M.P., India

Received : 15.11.2022 ; Revised : 25.11.2022 ; Accepted : 31.11.2022

DOI : https://doi.org/10.22271/09746315.2022.v18.i3.1642

ABSTRACT

For Hisar, Bhiwani, and Sirsa districts, as well as Fatehabad district, the western zone of Haryana has developed zonal wheat yield models based on weather data dating from 1980-81 to 2013-14. Multiple Linear Regression and Principal Component Analysis were employed to achieve this goal. The models' validity was confirmed for the post-sample years 2014-15, 2015-16, 2016-17, 2017-18, and 2018-19. The use of Regression Analysis and Principal Component Analysis in the development of zonal yield models has been crucial. There appears to be an overall preference for using prediction equations with higher loading of weather variables during the period of model creation and model testing Zonal yield models improved district-level yield forecast significantly by showing good agreement with real-time wheat yields. The overall results show a preference for principal component analysis-based prediction equations when it comes to capturing deviations from real-time yields in terms of percent. The wheat harvest can be predicted by zonal weather models 4-5 weeks prior to harvest.

Keywords: Weather variables, principal components, higher loading displaying weather variables, Eigen value.

Various organisations in India and overseas are exploring methodologies for pre-harvest agricultural yield forecasts using various approaches. Empirical statistical models are the most often utilised models. Crop growth is determined by a complicated set of processes, and the final yield is influenced by a variety of factors. Agricultural inputs and weather conditions are the two most important factors that influence crop output. These parameters are used in a wide range of agricultural yield forecasting models. Some of the regression models are by Chmielewski (1992), Agarwal *et al.* (2001), Dadhwal *et al.* (2005), Krishna and Suresh (2009), Pandey *et al.* (2013); principal component analysis by Gervini and Rousson (2004), Xingjie *et al.* (2010), Wang (2012) and Verma *et al.* (2014), Annu *et al.*(2017) and Muema *et al.*(2018) etc. The goal of this research was to create zonal yield models based on time series data on meteorological parameters and crop yield in order to anticipate pre-harvest wheat yields in Haryana. The focus was on examining the forecasting ability of the zonal yield models during the period(s) of model development and model testing to compare yield estimates generated under the two approaches. Finally, a comparison is done to examine if using trend yield and weather factors enhances the efficiency and reliability of pre-harvest wheat yield forecast at the district level in the state.

MATERIALS AND METHODS

The Selection of a crop region and data description

At an elevation between 74 ^{0}25' and 77 ^{0}38' east longitude and 27^{0}40' to 30^{0}55' north latitude, the state of Haryana has a total land area of 44212 square kilometers. Wheat is grown in varying densities throughout the state. The districts of Hisar, Bhiwani, Sirsa, and Fatehabad, which make up the state's western zone, were examined for the model construction of the model. According to Haryana's Bureau of Economics and Statistics, the trend-based yield was calculated using wheat yield data from 1980-81 to 2018-19. Daily data on maximum and minimum temperatures, rainfall, sunlight, and relative humidity were collected in the Hisar district for the same time period. To account for the fact that not all zones have access to the same weather data, the year and time variables were added to the model. However, the neighboring zones' zonal models made use of the same weather data as the zone's own. From 1980-81 to 2018-19, weather data from the first two weeks of November to one month before harvest

Email: pradeepjnkvv@gmail.com

How to cite : Chetna Devi. M., Kumari, Mohammed,J., Raghav, Y. S., and Mishra, P. 2022. Modelling and forecasting of electricity consumption used in agriculture purpose in India. *J. Crop and Weed,* **18** (3): 245-250.

was used for model creation and validity testing (crop growth period: 1st November to 15th April).

Average wheat yield data from the district were regressed against yield with time (year) as an independent variable, yielding the trend equation Tr = a+bt, where Tr = Trend projected yield, a = Intercept, b = Slope, and t = Year.

Computation of weather parameters

In November wheat is planted and in April it is harvested. The first fortnight weather parameter(s) are calculated from the average weather value of the 1^{st} to 15^{th} November the second fortnight weather parameter(s) were calculated from the average weather value of the 16^{th} to 30^{th} November and the 10t^{h} and 11^{th} fortnight weather parameter(s) were calculated from the 16^{th} to 31^{st} March and 1^{st} to 15^{th} April in the following year. The fortnightly meteorological parameters were calculated as follows:

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

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

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

Average Relative humidity $= \dfrac{\sum_{l=1}^{15} RH_l}{15}$

Average Sunshine hours $= \dfrac{\sum_{m=1}^{15} SSH_m}{15}$

Where, TMX_{i} = i^{th} day maximum temperature

TMN_j = j^{th} day minimum temperature

ARF_K = $_k{}^{th}$ day rainfall

RH_l = l^{th} day relative humidity

SSH_m = m^{th} day sunshine hours

(i,j,k,l,m refer to daily weather data)

The total wheat growth period was divided into 11 fortnights, with fortnightly weather parameters such as average maximum temperature (TMX), average minimum temperature (TMN), accumulated rainfall (ARF), average relative humidity and average sunshine hours (i.e. 11 fortnights x 5 weather variables = 55 weather parameters, calculated over years) being used in model building. In all of Haryana's districts statistical methodologies such as multiple linear regression and principal component analysis (PCA) were applied to develop zonal forecast models for pre-harvest wheat yield prediction. Dependent variable was regressed with a number of independent variables in a multiple linear regression model which may or may not be connected.

The standard linear regression model was considered. E($\boldsymbol{\varepsilon}$) = 0, an alternative way of writing the model is E(Y) = Xb. The normal equations (X′X) b = X′Y are fitted by least squares technique (here Y, X & b are same as above and (X′X) is the dispersion matrix) providing the solution $\hat{b}$ = bX′X^{-1} X′Y.

Extraction of principal component/higherloading displaying weather variables

The use and interpretation of a Multiple Regression model is frequently predicated on the assumption that the explanatory variables are loosely coupled. Regression findings can be difficult to interpret in circumstances where the explanatory variables are so closely connected. In order to address the problem of multicollinearity in factor extraction, the Principal Component method was utilized to identify the eigen values and eigen vectors. Principal components P_i (i = 1, 2, ….) were obtained as **P = kX**, where **P** and **X** are the column vectors of transformed and 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 are obtained by solving the equation | R – λI | = 0. Generally, we have non- zero roots of this equation. For each λ, the corresponding characteristic vector k is obtained by solving | R–λI | k = 0.

The forecast models were created using data covering 33 years, from 1980-81 to 2013-14 and the models were validated using data from 2014-15 to 2018-19. The model was built using weather data from the first fortnight of November to one month before harvest (as the crop growth period considered: 1st November to 15th April). In Haryana, district-level wheat yield projections were obtained using forecast models created for agro-climatic zones. The year (time) variable was introduced to account for variation between districts within a zone, as weather data was not available for all districts within a zone nonetheless, the zonal model used the same weather data in the zone's neighbouring districts.

Stepwise Regression Methods

It is possible to test a limited number of subset regression models by adding or removing one or more variables at a time, using various methodologies. These operations are referred to as "stepwise-type" approaches. Forward selection, backward elimination, and stepwise regression are all examples of the three types of processes.

RESULTS AND DISCUSSION

For quantitative forecasting, zonal yield models with increased loadings of meteorological factors and trend yield as regressors, as well as DOA crop yield as regressors, were fitted. In Table 1, the higher loading displaying weather variables is shown. Eigen vectors were utilised as weights to achieve higher loading displaying weather variables. The rotation of factors in an orthogonal direction typically shows a basic structure and aids comprehension. The higher loading displaying weather variables under the retained components was obtained using the Varimax rotation approach. Table 1 shows the selected models, together with their coefficients of determination and standard errors (Table 2). According to PC analysis, the first eight eigen values of the correlation matrix of explanatory variables (fortnightly weather data base) suggested an eight-factor solution, with the remaining components accounting for a smaller portion of overall variation. As a result, those parts were discounted as being of little use. Draper and Smith (1981) developed the stepwise regression method, in which all variables were first included in the required model and then excluded one at a time, with decisions at each stage conditioned by the previous step's result.The zonal yield models were also utilised to predict wheat yields at the district level for the post-sample periods of 2014-15, 2015-16, 2016-17, 2017-18, and 2018-19 (Table 3).

***Percent deviation*(*RD*%) ={(DOA yield- Forecast yield)/ DOA yield} × 100**

The preference for employing prediction equations based on principal component analysis may be seen in the post-sample period forecasted wheat yields and their percent deviations from observed yields. Tr, which is an indicator of technical improvement, improved management methods, and increased usage of high yielding varieties has been noted as an essential parameter appearing in all of the models showing that Tr explains the majority of the variability in yield. On the basis of average absolute percent deviations of forecasts from observed yields and root mean square errors (RMSEs), the performance(s) of the zonal prediction equations were compared in Tables 4 and 5.

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

RMSE is used as a measure of comparing two models, where O_i and E_i are the crop yield's observed and estimated values, and n is the number of years for which forecasting was done. Using zonal yield correlations, the researchers were able to enhance their ability to estimate wheat yields greatly, and the percentage of variance from DOA yields they found was within acceptable boundaries. To top it all off, the created algorithms are able to accurately predict wheat yields up to one month ahead of harvest, whereas D OA yields aren't available until well after harvest has taken place.

CONCLUSION

A examination of the findings shows that PC-based prediction equations based on larger loadings of wealthier variables are preferable. An essential parameter has been found in all models, which indicates that trend yield (Tr) accounts for most yield variability. This is a sign of enhanced fertilizer, pesticide and weedicide use, as well as an increase in the adoption of high-yielding varieties due to technical advancements. The typical absolute percent deviations of post-sample period forecasts range between 3.5 and 6 percent, which favors the use of created models. Zonal yield models can be used to forecast the state's district yield. Predicted district yields are in good according to DOA yield

Table 1: Rotated component matrix

Weather parametes	Principal Component(s)							
	1	2	3	4	5	6	7	8
TMN_1	0.98	0.01	0.01	-0.00	0.01	-0.06	0.02	0.03
TMN_8	0.96	0.03	0.02	-0.03	-0.00	-0.02	0.02	-0.00
TMN_6	0.98	0.00	0.01	0.02	0.00	-0.08	-0.00	0.00
TMN_5	0.98	0.01	0.01	-0.01	0.01	-0.02	-0.02	0.05
TMN_2	0.98	-0.00	0.06	0.00	0.00	0.00	0.01	0.00
TMN_3	0.98	0.01	0.01	-0.06	-0.02	-0.05	-0.02	0.02
TMN_7	0.98	0.01	0.10	-0.01	0.00	-0.00	0.04	0.01
TMN_4	0.99	-0.00	0.06	-0.01	-0.01	0.00	0.03	0.00
TMN_9	0.95	0.25	0.03	-0.02	-0.03	-0.01	-0.02	-0.01
TMX_8	0.94	0.04	-0.06	-0.00	-0.07	-0.04	*-0.08*	-0.00
TMX_1	-0.94	0.06	-0.02	0.08	-0.13	0.08	-0.04	-0.03
TMX_7	0.92	0.05	0.00	0.01	-0.09	0.11	0.09	0.01
TMX_6	0.91	0.00	0.02	0.08	-0.01	0.13	0.11	-0.12
TMX_2	-0.89	-0.01	-0.02	0.26	-0.07	0.06	-0.05	0.01
TMX_9	0.88	0.01	-0.23	0.07	-0.16	-0.00	-0.13	-0.00
TMX_5	0.79	-0.03	0.08	-0.00	0.05	0.20	-0.19	-0.22
RH_5	0.14	0.91	-0.02	0.01	0.00	-0.14	0.13	0.22
RH_6	0.14	0.91	0.04	0.00	0.00	-0.30	-0.03	0.10
RH_8	0.11	0.91	0.23	-0.04	0.00	-0.05	0.22	0.00
SSH_9	0.07	0.89	-0.26	0.00	0.06	-0.00	-0.02	0.13
RH_3	0.12	0.89	0.17	-0.31	-0.02	-0.05	0.08	0.03
RH_7	0.12	0.88	0.34	0.02	0.06	-0.11	0.06	0.04
SSH_2	-0.01	0.88	-0.10	0.23	0.06	0 .07	-0.15	-0.12
RH_4	0.13	0.88	0.11	-0.24	0.01	-0.23	0.15	-0.00
RH_9	0.09	0.87	0.30	-0.10	0.09	0 .01	0.15	-0.03
SSH_8	0.08	0.86	-0.04	0.00	-0.06	0.03	-0.16	0.08
SSH_1	-0.06	0.83	-0.05	0.05	-0.04	0.16	-0.11	-0.24
SSH_3	-0.00	0.82	-0.09	0.38	0.12	0.02	0.03	-0.07
SSH_7	0.01	0.80	-0.34	-0.00	-0.08	0.13	0.08	-0.10
RH_1	-0.43	0.80	0.09	-0.10	0.12	-0.05	0.10	0.14
RH_2	-0.51	0.77	0.15	-0.10	0.01	0.03	0.10	0.09
SSH_6	-0.07	0.74	-0.07	0.03	-0.00	0.54	0.14	-0.07
SSH_4	-0.03	0.63	-0.04	0.38	0.07	0.45	-0.14	-0.00
SSH_5	-0.07	0.62	0.20	-0.11	0.15	0.32	-0.10	-0.42
ARF_7	0.04	0.10	0.92	0.10	-0.03	0.08	0.08	0.03
ARF_9	0.06	0.00	0.88	-0.20	0.14	0.04	-0.03	-0.06
ARF_3	0.06	0.05	0.05	-0.91	-0.00	0.02	0.01	-0.01
TMX_3	-0.42	-0.01	-0.09	0.63	-0.10	-0.20	-0.07	0.26
TMX_4	0.49	-0.02	0.11	0.58	-0.03	0.43	-0.06	0.08
ARF_1	-0.04	0.04	0.05	0.08	0.97	-0.02	-0.04	0.00
ARF_2	-0.01	0.05	0.05	-0.12	0.94	0.06	-0.02	-0.02
ARF_6	0.02	0.21	-0.14	0.08	-0.03	-0.74	-0.18	-0.26
ARF_4	0.03	0.05	0.07	0.00	0.01	-0.05	0.87	-0.00
ARF_8	-0.04	.019	-0.04	-0.11	-0.11	0.27	0.71	-0.16
ARF_5	-0.08	0.12	-0.00	0.13	0.00	0.23	-0.17	0.76

Rotation Method : Varimax with Kaiser Normalization

Table 2: Zonal weather-yield models for Hisar, Bhiwani, Sirsa and Fatehabad districts of Haryana

Types	Yield Forecast Model(s)	Adj.R^2	SE
1	$Y=7.15+1.14*Tr+0.20*TMN_5-0.60*TMX_2+0.01*RH_4-0.07*ARF_9+ 0.44*SSH_1$ (8.16) (0.06) (0.19) (0.27) (0.04) (0.02) (0.28)	0.827	3.14
2	$Y=7.52+1.07*Tr-0.49*TMX_2+0.11*TMN_5-0.10*ARF_9-0.50*SSH_8+ 0.104*RH_7$ (7.32) (0.04) (0.22) (0.18) (0.02) (0.26) (0.04)	0.838	3.04
3	$Y=20.31+1.06*Tr-0.12*ARF_9-0.10*RH_1+0.15*RH_7-0.70*SSH_8-0.70*TMX_2-0.31*TMN_6$ (9.12) (0.05) (0.03) (0.06) (0.05) (0.26) (0.23) (0.19) 0.847 2.96		
4	$Y=-22.73+0.93*Tr-0.08*ARF_9-0.85*TMX_8-1.25*SSH_2+1.53*TMX_1+0.53*TMN_6+0.19$ (11.29) (0.05) (0.01) (0.14) (0.30) (0.29) (0.15) (0.06) $*RH-0.53*TMN_2$ (0.21)	0.890	2.50

Western zone comprised of Hisar, Bhiwani, Sirsa and Fatehabad districts. Figures in parentheses indicate the standard error and all the regressors are significant at pd” 0.05 in above zonal yield models.

Model 1, 2, 3, 4 - Higher loading displaying weather variables and trend yield as regressors

where,

- Y - Model predicted yield (q ha^{-1})
- Tr - Trend yield (q ha^{-1})
- TMX - Av. Maximum Temperature
- TMN - Av. Minimum Temperature
- ARF - Accumulated Rainfall
- RH - Av. Relative Humidity
- SSH - Av. Sunshine Hours (1,2, 3, ……,9 refer to different fortnights)
- R^2 - Coefficient of determination
- SE - Standard error of the estimate

Table 3 : Estimated wheat yield(s) based on zonal models and their associated per-centage deviation (RD%) = 100 × (observed yield-estimated yield)/observed yield

District/ Forecast years	DOA Yield (q ha^{-1})	Model-1 Fitted Yield (q ha^{-1})	Model-1 RD (%)	Model-2 Fitted Yield (q ha^{-1})	Model-2 RD (%)	Model-3 Fitted Yield (q ha^{-1})	Model-3 RD (%)	Model-4 Fitted Yield (q ha^{-1})	Model-4 RD (%)
Hisar									
2014-15	41.80	43.62	4.36	42.30	1.20	41.47	0.77	37.77	9.64
2015-16	46.48	48.10	3.49	47.98	3.23	46.10	0.81	44.23	4.81
2016-17	47.58	48.15	1.21	49.00	3.00	46.41	2.45	45.03	5.34
2017-18	49.14	50.12	1.99	51.40	4.59	49.11	0.06	45.45	7.50
2018-19	49.55	51.20	3.32	52.40	5.75	52.25	5.44	45.20	8.77
Bhiwani									
2014-15	38.50	37.09	1.52	36.99	3.90	36.21	5.93	35.80	6.99
2015-16	39.46	42.41	7.48	42.69	8.20	40.86	3.54	42.34	7.32
2016-17	43.21	42.48	1.66	43..74	1.23	41.19	4.66	43.25	0.09
2017-18	43.26	41.26	4.62	42.20	2.45	40.00	7.53	40.56	6.24
2018-19	44.25	45.20	2.14	45.20	2.14	43.45	1.80	42.25	4.51
Sirsa									
2014-15	47.06	46.49	1.20	44.98	4.40	44.13	2.92	42.72	9.21
2015-16	48.30	51.09	5.77	50.78	5.14	48.87	0.57	49.34	2.16
2016-17	50.71	51.26	1.09	51.92	2.38	49.30	1.40	50..33	0.74
2017-18	50.62	52.3	3.31	51.34	1.42	49.26	2.68	50.87	0.49
2018-19	50.62	51.5	1.73	49.87	1.48	49.89	1.44	51.23	1.20
Fatehabad									
2014-15	48.58	45.74	5.83	45.47	6.38	44.57	8.24	43.10	11.27
2015-16	45.41	50.77	11.81	51.21	12.78	49.25	8.47	49.67	9.39
2016-17	52.65	51.73	1.73	52.29	0.66	49.62	5.74	50.61	3.87
2017-18	53.85	52.89	1.78	54.28	0.79	50.14	6.88	55.23	2.56
2018-19	53.85	54.20	0.64	52.90	1.76	50.63	5.97	49.78	7.55

Table 4: Average absolute percent deviations of post-sample period forecasts for all the districts

Districts	Model-1	Model-2	Model-3	Model-4
Hisar	2.87	3.87	2.90	5.06
Bhiwani	3.91	3.48	4.43	4.71
Sirsa	2.62	2.19	2.44	1.95
Fatehabad	4.36	4.39	6.45	5.90

Table 5: RMSEs of the post-sample period forecasts based on zonal weather-yield models

Districts	Model-1	Model-2	Model-3	Model-4
Hisar	1.41	1.97	1.64	2.78
Bhiwani	1.79	1.62	2.13	2.12
Sirsa	1.54	1.37	1.46	1.33
Fatehabad	2.78	2.79	3.51	3.34

projections in all districts. The established zonal models, on the other hand, provide wheat yield projections at least a month in advance of harvest, but DOA yield estimations arrive significantly later. For district-level wheat yield forecasting in Haryana's western agro-climatic zone, zonal yield models may be used because the percent relative variations are acceptable.

REFERENCES

Agarwal, R., Jain, R. C. and Mehta, S. C. 2001. Yield forecast based on weather variable and agricultural inputs on agro-climatic zone basis. *Indian j. Agric.Sci.,* **71**(7):487-490.

Agarwal, R. and Mehta, S. C. 2007. Weather based forecasting of crop yields, pests and diseases - IASRI Models. *J. Indian Soc.Agric. Stat.*, **61**(2): 255-263.

Annu, Sisodia, B.V.S. and Rai, V.N. 2017. An application of principal component analysis for pre harvest forecast model for wheat crop based on biometrical characters. *Int. Res.J. Agric. Econ.Stat.*, **8**(1):83-87.

Bazgeer, S., Kamali, G. H. and Mortazavi, A. 2007. Wheat yield prediction through agro-meteorological indices for Hamedan. *Iran. Biaban,* **12**:33-38.

Chmielewski, Frank-M. 1992. Impact of climate changes on crop yields of winter rye in Halle (Southeastern Germany). *Climate Res.,* **2**:23-33.

Dadhwal, V. K., Sehgal, V. K., Singh, R. P. and Rajak, D. R. 2005. Wheat yield modeling using satellite remote sensing with weather data Recent Indian experience. *Mausam,* **54**:253-262.

Draper, N. R. and Smith, H. 1981. *Applied Regression Analysis*, 2nd ed. New York: John Wiley.

Gervini, D. and Rousson, V. 2004. Criteria for evaluating dimension – Reducing components for multivariate data. *The American Statistician,* **58**(1):72-76.

Kaiser, H.F.1960. The application of electronic computers to factor analysis. *Educational and Psychological Measurement,* **20**:141-151.

Krishna, P. S. R. and Suresh, K. K. 2009. A study on pre-harvest forecast of sugarcane yield using climatic variables. *Statistics and Applications,* **7&8**(1&2):1-8.

Muema, F., Home, P. and Raude, J. 2018. Application of benchmarking and principal component analysis in measuring performance of public irrigation schemes in Kenya. *J. Agric.*, **8**(10):162.

Pandey, K. K., Rai, V. N., Sisodia, B. V. S., Bharati, A. K. and Gairola, K. C. 2013. Pre - harvest forecast models based on weather variable and weather indices for Eastern U.P. *Adv. Biores.,* **4**(2):118-122.

Verma, U., Piepho, H. P., Ogutu, J. O., Kalubarme, M. H. & Goyal, M. 2014. Develoment of agromet models for district-level cotton yield forecasts in Haryana State. *Int. J. Agric. Stat. Sci.*, **10**(1):59-65.

Wang, W. 2012. Bayesian principal component regression with data-driven component selection. *J. Appl. Stat.,* **39**(6):1177-1189.

Xingjie, J., Yongqiang, Y. and Zhang, W. 2010. The Harvest Index Model of Winter Wheat in China based on meteorological data. *Scientia Agric. Sinica,* **43**(20):4158-4168.