International Journal of Chemical Studies

P-ISSN: 2349–8528
E-ISSN: 2321–4902
www.chemijournal.com
IJCS 2023; 11(3): 56-59

Received: 11-02-2023
Accepted: 18-04-2023

Gendre S
Department of Agril. Statistics, Indira Gandhi Agriculture University, Raipur, Chhattisgarh, India

Pandey KK
Department of Agril. Statistics, Indira Gandhi Agriculture University, Raipur, Chhattisgarh, India

Corresponding Author:
Gendre S
Department of Agril. Statistics, Indira Gandhi Agriculture University, Raipur, Chhattisgarh, India

Model development for yield forecast through composite raw data technique and stepwise regression analysis (Forward method) for chickpea crop of Chhattisgarh plain zone

Gendre S and Pandey KK

Abstract
The time series data on chickpea yield and weather variable *viz*. minimum maximum temperature, relative humidiy, sunshine, rainfall and wind velocity Weekly weather data over a span of 20 years data period (1998- 2017) for Dhamtari district of Chhattisgarh have been used in study. The model have been developed by Stepwise Regression analysis (forward method) and Multiple linear regression on Composite weather data. The Stepwise Regression model fitted on 14 generated weather variables along with T, i.e. 15 parameters have been used. The value of R^2 found i.e. 92% for Dhamtari district for Chickpea crop and multiple regression analysis has been applied on five new generated composite weather variables along with T as independent variables and de- trend yield used as dependent variable. The value of R^2 found 79% for Dhamtari district. The developed model has been validated by the the error parameters *viz.* MAE, MSE, RMSE, PE and PD along with maximum R^2. The above techniques showing that both the models are very reliable for district level forecast for chickpea crop for the district under the Chhattisgarh plain zone of Chhattisgarh. The present study covers under the study of individual effect of weather variables.

Keywords: Stepwise regression analysis and multiple linear regression. chickpea, weather indices

Introduction
Chickpea (*Cicer arietinum*) commonly known as Gram or Bengal Gram is the most important legume crop of India. In Chickpea protein contains 21%, fat 2.2%, carbohydrates 62% it also contain calcium of about 190 mg/100 g, Iron 90.5 mg/100 g, phosphorous 280 mg/100 g. It is used for human consumption as well as for feeding to animals. Fresh green leaves are used as vegetable while straw of Chickpea is an excellent fodder for cattle. The grain are also used as vegetable.

Different weather parameters affect different crops on different stages of crop growth period. Among meteorological variables, rainfall, temperature and soil moisture also are some of the important elements, which significantly affect crop yields, the meteorological parameters and yield data using phenological periods, in which weather elements illustrate significant correlation with crop yield.

Forecasting of the crop yield at appropriate (flowering) stages (which is two and half months before the harvesting for most of the crops) of the crop period well before the harvest are playing an important role in rural economy; and forecast of crop yield is very important for advance planning, formulation and its implementation. Forecasting well before (2-3 months) harvesting is playing an important role for pricing, marketing, packaging, post harvest management and transporting. Chickpea is a very important pulse crop for the state as well as country, so this crop has been covered under this research programmed. The yield forecasting of Chickpea before the crop growing year has been very beneficial to the researchers, state wise as well as country wise policy makers, farmers and district wise local and state wise mandis.

Material and Method
Study area
Dhamtari is located in Chhattisgarh (20.70° N, Latitude and 81.33° E Longitude) the average temperature in Bilaspur is 26.8 °C. Average rainfalls around 1221 mm of the state.

The minimum temperature goes to down to 5.5 °C during winter season. The Mahanadi river is principal river of this district.

Yield data
The yield data on Chickpea crop for 20 years (1998-99 to 2017-18) for Dhamtari District have been procured from the published booklets and official website (http://agridept.cg.gov.in/index.htm) of the Directorate of Agricultural, Govt. of Chhattisgarh.

Weather data
The District wise weekly meteorological data (1990-91 to 2017-18) procured from the department of Agro meteorology, Indira Gandhi Agriculture University, Raipur (CG.) and used for the different districts along with Chhattisgarh plain Zone for Chickpea crop. The data of 7 weather variables have been collected up to the 17 weeks of the chickpea crop cultivation which include 43rd SMW from first year to 7th SMW from 2nd year. The variables are Maximum Temperature (Max temp.), Minimum Temperature (Min. temp), Morning Relative Humidity (RH-I)%, Afternoon Relative Humidity (RH-II)%,Bright Sunshine hours (SS) hours/day, Rainfall (mm), Wind Velocity (WV) respectively.

Statistical Methodology
Forecast Model based on weather variables through Step wise Regression Analysis (forward method)
The model development through Stepwise Regression Analysis (forward method) method However, Earlier this concept for individual effect of weather variable and the procedures laid down Agrawal *et al.* 1986 [1] have been applied. They expressed the effect of change in weather variables on yield in w^{th} week as a linear function of respective correlation coefficients between yield and weather variables. As trend effect on yield is generally expected to be considerable one, its effect need to be removed from yield while calculating correlation coefficients of yield with weather variables to be used as weights. But in this investigation new technique has been used for the study of effect of individual weather parameter and here we has been used for the development of forecasting models.
The two new variables from each weather parameter (consisting of 17 meteorological week data for Chickpea) generated for the model fitting. Now; total 14 variables has been generated (7 weighted and 7 un-weighted) for all the weather parameter. These 14 generated weather variable along with T i.e. 15 parameters has been used for fitting of the model. The variables has been generated on only five month average data and the following techniques has been used:
The un weighted generated variables has been generated as follows:

$$Z_{i(unwt)} = \sum_{w=1}^{n} X_{iw} \Big/ n \qquad i=1, 2, \ldots 7$$

The weighted generated variables has been computed as follows:

$$Z_{i(wt)} = \sum_{w=1}^{n} r_{iw} X_{iw} \Big/ \sum_{w=1}^{n} r_{iw}$$

Let r_{iw} be the simple correlation coefficient between weather variable X_{iw} at w^{th} week and crop yield over the period for respective years (n indicates years). Two new variable for each weather parameter has been generated. Finally only significant variable (entered) after the Step Wise Regression (forward method) has been used for fitting of model. The yield and generated variable has been used for model development. The developed model has been given below: -

$$y = a + b_i Z_{ij} + b_i T$$

Where;
y = crop yield
a = intercept
b_i = (i = 1,2…..n) Coefficients for entered variable.
Z_{ij} = Generated variable entered in the model.
T = Trend
This method has been applicable for each district and zone.

Measures for validation and comparison of the models
Following procedures have been used for the comparison and the validation of the developed models, Out of 28 years, 23 years for the development of the model and 5 years for validation for 4 districts and out of 20 years, 16 years for the development of models and 4 years for validation for 5 districts.

Different procedures for model validation are given bellow
Percent Deviation (PD)
The formula for computation of Percent Deviation of forecast yield from actual yield is given by Md. Azfar *et al.* (2015) [4]. This measures the deviation (in percentage) of forecast from the actual yield data. The formula for calculating the percent deviation of forecast is given below

$$\text{Percent deviation} = \frac{(\text{Actual Yield - Forecasted Yield})}{(\text{Actual yield})} x100$$

Root Mean Square Error (RMSE)
Root Mean Square Error (RMSE): RMSE indicate the magnitude of the average error, but provide no information on the relative size of the average difference between (F) and (O). This method is most popular and important for the validation of the model. The formula for calculating the RMSE is given by Varshneya *et al.* (2010) [11] of forecast is given below:

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

Where;
i = 1, 2, 3.
n = Sample size
F_i = Actual yield
O_i = Predicted yield

Mean Absolute Error (MAE)
Mean absolute error represents the difference between the original and predicted values extracted by averaged the absolu difference over the data set. Kotz *et al.* (2006) [8].

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - x_i|$$

Where:
n =the number of error
$|y_i - x_i|$ = the absolute error

Mean Squared Error (MSE)
Mean squared error (MSE) measures the amount of error in statistical models. It assesses the average squared difference between the observed and predicted values. When a model has no error, the MSE equals zero. Error increase, its value increases. The mean squared error also known as the mean squared deviation Kotz *et al.* (2006) [8].

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y})^2$$

where;
n = Number of data point
Y_i = Observed value
$\hat{Y}$ = Predicted value

Percent error (PE)
PE is defined as ratio of RMSE to mean observed value expressed as percentage This is another important tool of the model validation. The formula for calculating the Percent Error is given by Varshneya *et al.* (2010) [11] of forecast is given below:

$$PE = \frac{RMSE}{\bar{o}} \times 100$$

Where;
PE = Percent error
RMSE = Root mean square error
$\bar{O}$ = Mean of observed value

Coefficient of Determination (R^2)
The models were validated on the basis of **(**R^2**)** which can be computed from the formula given by Draper and Smith (1988) [6]. The R^2 statistic, a measure of the amount of variation about the mean explained by fitted equation, as

$$R^2 = 1 - \frac{RSS_{n-p}}{CTSS}$$

Where, CTSS denotes the corrected total sum of squares. P is the total number of parameters in a fitted model (Including β_0) and RSS_{n-p} is the corresponding residual sum of squares.
R^2 never decreases when a regressed is added to the model, regardless of the value of the contribution of the variable in the model. Therefore, it is difficult to judge whether an increase in R^2 is really important.

Results and Discussion
Model Developed through Stepwise regression analysis (forward method) for Dhamtari
The stepwise regression model has been obtained
y =-211.88 + 89.42 Z21 + 10.00 Z40 + 523.77 Z70 -612.89 Z71

Table 1: Model Developed through Stepwise regression analysis (forward method) for Dhamtari

District	Entered Variable in model	Coefficient	P-value	Standard Error	R^2	significance
Dhamtari	Intercept	-211.88	0.378	230.804	0.92***	0.001
	Z_{21}	89.42***	0.0001	13.930		
	Z_{40}	10***	0.0001	122.897		
	Z_{70}	523.77**	0.002	129.649		
	Z_{71}	-612.89*	0.020	3.664		

***p<0.001, ** p<0.01,* p<0.05,+ p<0.1

The variable Z_{21} (weighted Minimum temperature) along with Z_{40} (un-weighted RH-ll), Z_{70} (un-weighted wind velocity) and Z_{71} (weighted wind velocity) variable entered in model for Dhamtari district. The value of R^2 found i.e. 92% which is highly significant The entered variable Z_{21}, Z_{40}, Z_{70} and Z_{71} is significant at different level of significance. The detail of result has been given table 1

Table 2: Validation and comparison of developed model by SRA for Dhamtari

	Year	Actual yield	Predicted yield	MAE	PE	MSE	RMSE	R^2	PD	Sig.
DHM	2014	1360	1004.467	227.36	24.39	79269.96	281.55	0.92***	26.14	0.001
	2015	900	1325.9685						-47.33	
	2016	1261	1302.152						-3.26	
	2017	1097	1183.7957						-7.91	

*** p<0.001, ** p<0.01,* p<0.05,+ p 0.1

DHM = dhamtari, Sig. = siginficance

The value of MAE (227.39), PE (24.39), MSE (79269.69), RMSE (281.55, the value of R^2 found i.e. 92% which is highly significant at 0.1% level of significance). The validation of model minimum PD show for 2015 years followed by 2017 and maximum 2014 followed by 2016 the predicted yield of different years showing on table 2, the detail result of validation show on table 2.

References

1. Agrawal R, Jain RC, Jha MP. Models for studying rice crop weather relationship. Mausam. 1986;37(1):67-7.
2. Agrawal R, Mehta SC. Weather Based Forecasting of Crop Yields, Pests and Diseases - IASRI Models. J Ind. Soc. Agril. Stati. Ft. 2007;61(2):255-263.
3. Agrawal R, Jain RC, Singh D. Forecasting of rice yield using climatic variables. Ind. J Agric. Sci. 1980;50(9):680-84.
4. Azfar Mohd. Sisodia BVS, Rai VN, Monika. Pre-harvest forecasting models for rapeseed and mustard yield using principle component analysis of weather variable. Mausam. 2015;66(4):761-766.
5. Banakara KB, Panadya HR, Garde YA. Pre-harvest forecast of Kharif Rice yield using PCA and MLR technique in Navsari of Gujarat. Jr. of Agrometeorology. 2019;21(3):336-343.
6. Draper NR, Smith H Applied Regression Analysis. 3rd edition, John Wiley & Sons Inc, 1988.
7. Garde YA, Thorat VS, Pisal RR, Shinde VT. Pre harvest forecasting of Kharif Rice yield using weather parameters for strategic Decision Making in agriculture. Int. J of Env. clim. chag. 2020;10(12):162-170.
8. Kotz S, Johnson NL. Encyclopedia of Statistical Sciences, 2nd.edition, John Wiley & Sons Inc, 2006.
9. Pandey KK, Rai VN, Sisodia BVS, Singh SK. Effect of weather variables on rice crop in eastern Uttar Pradesh, India. Plant archives. 2015;15(1):575-579.
10. Pandey KK, Maurya D, Gupta G, Mishra SV. Yield forecasting models based on weather parameters for eastern UP. Vegetos. 2016;29(1):22-24.
11. Varshneya MC, Chinchorkar SS, Vaidya VB, Pandey Vyas. Forecasting models for seasonal rainfall for different regions of Gujarat. Jr. of Agrometeorology. 2010;12(2):202-207.
12. Smith AM, Blackshaw RE. Weed–crop discrimination using remote sensing: a detached leaf experiment. Weed Technology. 2003;17(4):811-820.