Indian Journal of Agricultural Sciences **88** (10): 1597–9, October 2018/Article

Forecasting of crop yield using weather parameters–two step nonlinear regression model approach

SANJEEV PANWAR[1], ANIL KUMAR[2], K N SINGH[3], RANJIT KUMAR PAUL[4], BISHAL GURUNG[5], RAJEEV RANJAN[6], N M ALAM[7] and ABHISHEK RATHORE[8]

Indian Council of Agricultural Research, Krishi Bhavan, New Delhi 110 001

Received: 15 May 2018; Accepted: 09 July 2018

ABSTRACT

Concept of the paper is firstly to remove the trend of crop yield and then to develop the forecasting models using detrended yield. Not much work is available or development of forecast models or modelling due to their non-linear behaviour. For that, in this paper, methodology developed for forecasting using nonlinear growth models, which will help in forecasting yield, pest and disease incidences etc with high accuracy. Crop yield forecast models for wheat crop have been developed (using non-linear growth models, linear models and weather indices approach with weekly weather data) for different districts of Uttar Pradesh (UP). Weather Indices (WI) were obtained using above two approaches. Weather indices based regression models were developed using weather indices as independent variables while character under study such as crop yield was used as dependent variable for wheat crop, i.e. two step non-linear forecast model. Technique of forecasting using non-linear approach and using weather indices will enrich the knowledge in developing customized models on forecasting for different types of crops and for different locations. The approach provided reliable yield forecast about two months before harvest.

Key words: Detrended yield, Forecasting, Nonlinear regression model, Weather Indices Approach

The agriculture industry represents an important component of the Indian economy both in terms of its contribution to the GDP as well as a source of employment to the majority of the country's population. This sector is currently showing immense opportunities, with India presently being the world's third largest agricultural producer by value (after China and the US). A number of transformations have taken place in this sector over the past few decades. These include - rising penetration of the organised sector, growth in contract farming, agriculture becoming more mechanized, easy loan facilities, rise of exports, use of agrochemicals and high yielding seeds and an increasing role of the private sector in processing, branding and marketing, etc. Crop yield is influenced by technological change and weather variability. Technological factors increase yield smoothly through time and, therefore, years or some other parameters of time can be used to study the overall effect of technology on yield. Thus preharvest forecasting of production is required when crop is still standing in the field (Agarwal *et al*. 1980). An efficient forecasting is thus a pre-requisite for food supply information system at district and state level. Paul *et al*. (2013, 2014, and 2015) have developed some pre harvest forecast models for wheat yield based on weather variables. They have applied autoregressive integrated moving average with exogenous variable (ARIMAX) model and ARIMAX-Generalized autoregressive conditional heteroscedastic (GARCH) model for forecasting wheat yield in Kanpur district of Uttar Pradesh. It is of high importance that with the advancements in non-linear estimation techniques and high computational ability to try for non-linear modeling of yield forecast using different weather parameters and using their indices. In this paper application of nonlinear regression technique has been made for modelling and forecasting yield of wheat crop for Aligarh district of Uttar Pradesh.

MATERIALS AND METHODS

Time series yield data of 40 years (1970-2010) and weather data for the year 1970-71 to 2009-10 of Aligarh district have been utilized. The models have been used to forecast yield in the subsequent three years 2008-09 to 2009-10 (which were not included in model development). District level wheat crop yield data for 40 years (1970-2010) have been collected from the Directorate of Economics and Statistics, Ministry of Agriculture, Government of India, New Delhi and from the Agriculture Directorate, Lucknow (UP) and India Metrological Department, Pune and CRIDA. Wheat is generally sown in the month of October when average daily temperature falls around 23-25^{o}C. The pre-sowing phase of the crop is important because in this phase of two to three weeks, the land is prepared for the cultivation. If

[1]Principal Scientist (e mail: scientist1775@gmail.com), ICAR HQ. [2]Principal Scientist, [3]Scientist, [4]Principal Scientist, [5]Scientist, [6]Scientist, ICAR-IASRI, New Delhi, [7]Scientist, IISWC, Dehradun. [8]Principal Scientist, ICRISAT, Hyderabad.

the weather condition is adverse during the pre-sowing phase the sowing of the crop is generally delayed. After sowing of the crop, germination takes 6-7 days or near about one week after the pre-sowing phase. After germination phase, crown root initiation occurs after 20-25 days of sowing or in about 3 weeks from germination. Tillering phase starts just after the crown root initiation phase and lasts up to 40-45 days after sowing or nearly about 2-3 weeks after crown root initiation phase. Jointing and reproductive phase is the peak plant growth stage and starts after the tillering phase or 45-60 days after sowing. The reproductive phase lasts 60-85 days after sowing. As weather during pre–sowing period is important for establishment of the crop, data starting from two weeks before sowing have been included in model development. Further, as the forecast is required well in advance of harvest, weather data about 2 months before harvesting have been considered. Thus data on four variables, viz. max. temp., min. temp., RH, rainfall during 16 weeks data from 40^{th} SMW to 3^{rd} SMW (next year).

Crop yield forecast models have been developed for districts of Aligarh of UP using weekly weather data. Residuals obtained from the selected non-linear models and linear models. Weather Indices (WI) were obtained using nonlinear and linear modelling approaches (Panwar *et al*. 2009 and 2016). Weather indices based regression models were developed using weather indices as independent variables while character under study such as crop yield was used as dependent variable for Wheat crop. Stepwise regression technique has been used for selecting significant variables in all the models. Statistical approaches have been used for development forecast model based on weather variables, which are as follows:

Linear regression approach: Suppose we have a simple linear regression equation:

Mathematically, $y_i = \theta_0 + \theta_1 x_i + \varepsilon_i, i = 1,2,...,n$

where, y_i is the dependent variable and the variable under study, x_i is the explanatory variable and ε_i is the error component. After estimating the parameters θ_0 and θ_1 of the regression equation through different parameter estimating techniques we can obtain the predicted value of the dependent variable as: $y_i = \theta_0 + \theta_1 x_i + \varepsilon_i, i = 1,2,...,n$

$\hat{y}_i$ obtained from above models has been used (as y_i) for the further development of forecast models and for further prediction respectively.

Non-linear models: Some nonlinear growth models (Seber and Wild 1989) will provide a reasonable representation of yields which are expected to provide a reasonable representation of crop yield as compared to linear models will be tried for fitting the yield of selected crop in a selected location using the weather data.

Logistic Model: $Y = \dfrac{\alpha}{1+\exp(b-ct)} + e$;

Gompertz Model: $Y = a\exp(-\exp(b-ct)) + e$

Richards Model: $Y = \dfrac{\alpha}{1+\exp(b-ct)^{1/d}} + e$;

Wiebull Model: $Y = a - b\exp(-ct^d) + e$

where a, b, c and d are the unknown parameters to be estimated.

For nonlinear estimation procedures, Gauss-Newton Method is used which yields efficient estimates with proper convergence was applied in this study.

The residuals from above model will be used in the subsequent linear model for fitting against different forms of the weather variables including their indices. The weather variables within an agricultural year will be aggregated through mean or through indices as mentioned above. Further, indices will be computed based on the influence (positive or negative) of the selected weather variable on the crop yield.

Weather Indices Approach: For each weather variable, two indices will be developed, one as simple total of values of weather variables parameter in different weeks and the other one as weighted total, weights being correlation coefficients between detrended yield and weather variable in respective weeks. The first index represents the total amount of weather parameter received by the crop during the period under consideration. While the other one takes care of distribution of weather parameters with special reference to its importance in different weeks in relation to the detrended yield. On similar line, indices were computed with products of weather variables (taken two at a time) for joint effects.

These indices are computed as follows:

$$y = a_0 + \sum_{i=1}^{p} \sum_{j=0}^{1} a_{ij} Z_{ij} + \sum_{i \neq i'=1}^{p} \sum_{j=0}^{1} b_{ii'} Z_{ii'j} + \varepsilon$$

$$Z_{ij} = \sum_{w=n_1}^{n_2} r_{iw}^{j} X_{iw}$$

$$Z_{ii'j} = \sum_{w=n_1}^{n_2} r_{ii'w}^{j} X_{iw} X_{i'w}$$

where, y is detrended yield forecast, X_{iw} is value of i^{th} weather variable in w^{th} week, $r_{ii'w}$ is correlation coefficient between Y and i^{th} weather variable in w^{th} week, $r_{ii'w}$ is correlation coefficient between Y and product of X_{iw} and n_1 in w^{th} week, p is number of weather variables, n_1 is initial week for which weather data are included in the model, a_{ij} and b_{ii} is final week for which weather data are included in the model and model, a_{ij} and b_{ii} are parameters to be estimated, ε is the random error.

Stepwise regression technique was used for retaining significant variables only in the forecast models in each approach.

If we use the weekly data on m weeks in p variables, now new weather variables and interaction components can be generated with respect to each of the weather variables using the below mentioned procedure. Forecast model has been developed by considering all the generated variables simultaneously including the time trend (T). In order to study the individual effect of each weather variables, two new variables from each variable can be generated as follows:

Let X_{iw} be the value of the i^{th} (i=1(1)p) weather variable at the w^{th} week (w=1(1) n), r_{iw} be the simple correlation coefficient between weather variable X_i at the w^{th} week

Table 1 Aligarh wheat yield forecast models (Two steps non-linear and linear models) based on weather indices approach

Model Name	Forecasting model	Goodness of fit	
		R^2	RMSE
Nonlinear	$Yt = 1.61 - \underset{(0.04)}{0.54}Z21 - \underset{(0.009)}{0.015}Z41 + \underset{(0.0015)}{0.021}Z121 + \underset{(0.001)}{0.0031}Z241$	0.89	2.35
Linear	$Yt = 0.71 - \underset{(0.015)}{0.11}Z11 - \underset{(0.0010)}{0.0017}Z41 + \underset{(0.001)}{0.024}Z121 + \underset{(0.0010)}{0.0041}Z241 + \underset{(0.0085)}{0.01}Z341$	0.85	2.40

and yield over a period of k years.

The generated variables are given by:

$$Z_{ij} = \frac{\sum_{w=1}^{n} r_{iw}^{j} x_{iw}}{\sum_{w=1}^{n} r_{iw}^{j}}; j = 0,1$$

For j=0 we have unweighted generated variable as:

$$Z_{i0} = \frac{\sum_{w=1}^{n} X_{iw}}{n}$$

And weighted generated variables as:

$$Z_{i1} = \frac{\sum_{w=1}^{n} r_{iw} X_{iw}}{\sum_{w=1}^{n} r_{iw}}$$

For each year [1].

The following model is then fitted to study the effect of individual weather variables.

$$Y = a_0 + a_1 Z_{i0} + a_2 Z_{i1} + cT + \varepsilon$$

where, Y is yield, T is variable expressing time effect, a_0, a_1, a_2 and c are constant entities known as the parameters that needs to be evaluated from the model and ε ε is the error term which is supposed to be distributed with a null expectation and a constant dispersion R_{adj}^2. Thus for each of the weather parameter two variables will be obtained which along with the time component and the intercept term makes a total of 10 parameters thus 10 parameters have been estimated in order to detrend the yield.

Goodness of fit: Different regression models were compared on the basis of adjusted efficient of determination (R_{adj}^2) which is as follows:

$$R_{adj}^2 = 1 - \frac{ss_{res} / (n-p)}{ss_t / (n-1)}.$$

where ss_{res} /(n-p) is the residual mean square and ss_t/(n-1) is the total mean square.

From the fitted models, wheat yield forecasts for the years 2008-09 to 2009-10 were obtained and forecasts were compared with the actual yield on the basis of Root Mean Square Deviation (RMSE):

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

where O_i and the E_i are the observed and forecast value of crop yield respectively and n is the number of years for which forecasting has been done.

Table 2 MAPE of forecast of wheat yield from models developed

Districts	Two step nonlinear models	Linear models
Aligarh	9.15	10.06

RESULTS AND DISCUSSION

The two step nonlinear forecasting model was found to be superior to the linear model as its RMSE value (2.35) is much lower as compared that of linear model (2.40). For fitting this residual model, the nonlinear model (Logistic) which was found to have better fit was used to output the residuals. The negative values of the coefficients for variables such as Z2 (Min) and Z4 (rh), showed that increase in the variables results in decrease in the yield. On the other hand, variable such as Z12 (max*min) and Z24 (min*rh) yielded positive coefficients which means that increase in the variables will increase the yield of wheat crop in Aligarh district (Table 1) of Uttar Pradesh. Similarly, RMSE values are much lower as compared to that of linear model in almost all selected districts of Uttar Pradesh, thereby showing that two step nonlinear forecasting models were found to be superior for yield forecasting of wheat crop (Table 2).

REFERENCES

Agrawal, Ranjana, Jain R C, Jha M P and Singh D. 1980. Forecasting of wheat yield using climatic variables. *Indian Journal of Agricultural Sciences* **50(9)**: 680–4.

Gurung Bishal, Panwar Sanjeev, Singh K N, Banerjee Rahul, Gurung Sisir Raj and Rathore Abhishek. 2017. Wheat yield forecast using detrended yield over a sub-humid climatic environment in five districts of Uttar Pradesh, India. *Indian Journal of Agricultural Sciences* **87** (1): 87–91.

Panwar S, Singh K N, Kumar A, Paul R K, Sarkar S K, Gurung B and Rathore A. 2016. Performance evaluation of yield crop forecasting models using weather index regression analysis. *Indian Journal of Agricultural Sciences* **87** (2): 270–2.

Panwar S, Sharma S C and Kumar A. 2009. A critical study of Onion (*Allium cepa*) export of India: Nonlinear approach. *Indian Journal of Agriculture Sciences* **78** (2): 178–80.

Paul R K, Prajneshu and Ghosh H. 2013. Statistical modelling for forecasting of wheat yield based on weather variables. *Indian Journal of Agricultural Sciences* **83** (2): 180–3.

Paul R K, Ghosh H and Prajneshu. 2014. Development of out-of-sample forecast formulae for ARIMAX-GARCH model and their application. *Journal of the Indian Society of Agricultural Statistics* **68**(1): 85–92.

Seber G A F and Wild C J. 1989. *Non-linear Regression*. John Wiley and Sons, New York.