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

SANJEEV PANWAR¹, ANIL KUMAR², K N SINGH³, RANJIT KUMAR PAUL⁴, BISHAL GURUNG⁵, RAJEEV RANJAN⁶, N M ALAM⁷ and ABHISHEK RATHORE⁸

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 Meteorological Department, Pune and CRIDA. Wheat is generally sown in the month of October when average daily temperature falls around 23-25°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
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. Tilling 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 tilling 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 40th SMW to 3rd 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 were developed using weather indices as independent variables. 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:

\[
y = a_0 + \sum_{j=1}^{p} a_j Z_{ij} + \sum_{j=1}^{p} b_j Z_{ij} + \varepsilon
\]

\[
Z_i = \sum_{w=1}^{n} r_{i,w} X_w
\]

\[
Z_{ij} = \sum_{w=1}^{n} r_{i,w} X_{iw} X_{wj}
\]

where, \( Y \) is detrended yield forecast, \( X_{iw} \) is value of i-th weather variable in w-th week, \( r_{i,w} \) is correlation coefficient between \( Y \) and i-th weather variable in w-th week, \( r_{i,w} \) is correlation coefficient between Y and the 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 \( b_{ij} \) is final week for which weather data are included in the model and model, \( a_{ij} \) and \( b_{ij} \) are parameters to be estimated, \( \varepsilon \) 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 (\( w=1(1)n \)), \( r_{i,w} \) be the simple correlation coefficient between weather variable \( X_i \) at the \( w \)-th week.
and yield over a period of k years.

The generated variables are given by:

\[ Z_0 = \sum_{i=1}^{n} r_iX_i \]

For j=0 we have unweighted generated variable as:

\[ Z_{0j} = \frac{\sum_{i=1}^{n} X_{ij}}{n} \]

And weighted generated variables as:

\[ Z_{nj} = \frac{\sum_{i=1}^{n} r_iX_{ij}}{n} \]

For each year [1].

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

\[ Y = a_0 + a_1 Z_{0j} + a_2 Z_{10} + cT + \varepsilon \]

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 \( \varepsilon \) 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:

\[ R_{adj}^2 = 1 - \frac{ss_{res} / (n-p)}{ss_y / (n-1)} \]

where \( ss_{res} / (n-p) \) is the residual mean square and \( ss_y / (n-1) \) 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):

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

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 1

<table>
<thead>
<tr>
<th>Model Name</th>
<th>Forecasting model</th>
<th>Goodness of fit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonlinear</td>
<td>( Y_t = 1.61 - 0.54Z_{21} - 0.015Z_{41} + 0.021Z_{121} + 0.0031Z_{241} )</td>
<td><strong>R^2</strong> 0.89 <strong>RMSE</strong> 2.35</td>
</tr>
<tr>
<td>Linear</td>
<td>( Y_t = 0.71 - 0.11Z_{11} - 0.0017Z_{41} + 0.024Z_{121} + 0.0041Z_{241} + 0.01Z_{341} )</td>
<td><strong>R^2</strong> 0.85 <strong>RMSE</strong> 2.40</td>
</tr>
</tbody>
</table>

### Table 2

<table>
<thead>
<tr>
<th>Districts</th>
<th>Two step nonlinear models</th>
<th>Linear models</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aligarh</td>
<td>9.15</td>
<td>10.06</td>
</tr>
</tbody>
</table>

### 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


