Crop yield forecasting of paddy, sugarcane and wheat through linear regression technique for south Gujarat

NEERAJ KUMAR, R. R. PISAL, S. P. SHUKLA and K. K. PANDEY*

N. M. College of Agriculture, Navsari Agricultural University, Navsari- 396 450 (Gujarat), India

*V. C. S. G. College of Horticulture, Uttarakhand University of Horticulture and Forestry, Uttarakhand, India

(Received 17 October 2012, Modified 27 May 2013

**Corresponding author:** Neeraj Kumar, N.M. College of Agriculture, Navsari Agricultural University, Navsari - 396 450 (Gujarat), India. Email: neeraj34012@gmail.com

Abstract: Multiple regression models were developed for yield forecasting of paddy, sugarcane and wheat for two districts of Gujarat (Navsari and Bharuch). The historical weather and crop yield data of 31 years of Navsari (1980-2010) and 27 years of Bharuch (1984-2010) were used. The data of de-trend yield and generated weather variables for 27 years of Navsari (1980-2006) and 23 years of Bharuch (1984-2006) were used for generation of the model for both districts. Significant weather variables are obtained on the basis of highest R² and significant P-value. The multiple regression analysis was executed by trial and error method. The models were validated with 4 years independent data set (2007 to 2010) of these two districts. During the validation period, Navsari district model deviations for paddy, sugarcane and wheat were between -7.30 to 3.41%, 1.68 to 2.05% and -8.27 to 11.51% respectively. Similarly Bharuch model deviations for paddy, sugarcane and wheat were between 5.35 to 11.76%, -12.65 to 7.18% and -12.07 to 6.86% respectively.

Key words – Regression models, R², Validation, Yield forecasting.

1. Introduction

Agriculture is an economic sector which depends highly on climatic conditions. Crop models are frequently used to evaluate the ability of climate forecasts in guiding crop management practices. In statistical model approach, one or several variables (representing weather or climate) are related to crop responses such as yield and yield contributing characters. Crop yield is the integrated effect of a number of physical and physiological processes that occur during the crop-growing period. These processes are influenced by the characteristics of the crop, weather, soil and time management factors (Rafi and Rehan, 2005). Statistical models provide simple alternative to process-based models. The main advantages of statistical models are their limited reliance on field calibration data and their transparent assessment of model uncertainties (Lobell and Burke, 2010). Shankar and Gupta (1988) related the meteorological parameters and yield data using the phenological periods, in which weather elements illustrate significant correlation with crop yield, by using multiple linear regressions. The work of Wickham (1973) also clearly brought out that the yield variation in paddy crop production, due to weather, management and biotic factors can be addressed through a statistical modelling approach. Regression models are often used because of their simple and straightforward nature, relating yield with one or more meteorological factors. As such these are used extensively in the studies on soils, agronomic practices and insect-disease; a combination of such factors is still difficult with some of the dynamic simulation models. Many earlier researchers (Norman 1979), Fisher (1924) and Hendrick and Scholl (1943) have suggested models which require small number of parameters to be used
while taking care of distribution pattern of weather over the crop season. Models based on weather parameters can provide reliable forecast of crop yield in advance of harvest (Agrawal and Mehta, 2007). Parthasarathy et al. (1988) and Ramakrishna et al. (2003) have developed forecast equations based on regression model, for total Indian food grain production using monsoon rainfall and soil index. Regression equations have also been developed for forecasting paddy yield (Shankar and Gupta, 1987), for estimation of sugarcane yield (Singh and Bapat, 1988) and for wheat yield (Agarwal et al., 2012). Keeping above facts in view, the present study was undertaken to predict yields of rice, wheat and sugarcane.

Rice is cultivated during Kharif season (June to October), wheat is sown during Rabi season (November to March) and Sugarcane which is one year crop, grown during Nov to Nov in both the districts of Gujarat.

2. Materials and methodology

Weekly weather data of crop growing season of 31 years (1980-2010) of Navsari district (20°57’ latitude, 72°54’ longitude and 11.89 m above mean sea level) and 27 years (1984-2010) of Bharuch district (21°42’ latitude, 72°58’ longitude and 15.0 m above mean sea level) were collected and used in the present study. The yield data was procured from the official site (Directorate of Agriculture, Gujarat) for Navsari and Bharuch for the same time period. Weekly data of maximum temperature (Tmax) °C, minimum temperature (Tmin) °C, morning relative humidity (RH-I)%, afternoon relative humidity (RH-II)% bright sunshine hours (BSS) hours/day and rainfall (mm) were employed according to growing period of each crop. Correlations were worked out between forecasters (weather parameters) with respective year yield of crop. The yield data was detrended.

The step-wise forward model (1) can be used for obtaining the most significant weather variable, by which we can fit the multiple regression model (2) and these most significant variables can be used for correlation coefficient.

\[ Y = a + b_1X_1 + b_2X_2 + b_3X_3 + \ldots + b_nX_n \]  
\[ n = 1, 2, \ldots, 5 \]  
(1)

\[ Y = a + b_1X_1 + b_2X_2 + b_3X_3 + \ldots + b_nX_n \]  
\[ n = 1, 2, 3 \]  
(2)

The summary measures describe the quality of simulation while the difference measures try to locate and quantify errors. The latter include the Mean Absolute Error (MAE), the Mean Bias Error (MBE) and the Root Mean Square Error (RMSE) (Varshneya et al., 2010).

\[ \text{MAE} = \frac{1}{n} \sum_{i=1}^{n} (F_i - O_i) \]

\[ \text{MBE} = \frac{1}{n} \sum_{i=1}^{n} (F_i - O_i) \]

\[ \text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (F_i - O_i)^2} \]

MAE and RMSE indicate the magnitude of the average error, but provide no information on the relative size of the average difference between (F) and (O). The statistic MBE describes the direction of the error bias. The value of MBE is related to the magnitude of the values under investigation. A negative MBE indicates that the forecasting is smaller in values than those of the corresponding observations. PE is defined as ratio of RMSE to mean observed value expressed as percentage (Varshneya et al., 2010).

\[ \text{PE} = \frac{\text{RMSE}}{O} \times 100 \]

3. Results and discussion

3.1. Validation of models

The observed and forecasted yields for period (2007-2010) have been presented in Table 2 and various error analysis of independent and all data set in Table 3. The regressions models were validated with the four years (2007-2010) of independent data set. The data exposed that the Navsari district paddy model deviated between -7.30 to 3.41% from the observed yield, the error analysis revealed that the MAE, MBE, RMSE, PE and average error of model were 94.25 kg ha\(^{-1}\), -94.25 kg ha\(^{-1}\), RMSE ± 96.42 kg ha\(^{-1}\), 2.87% and 4.9% respectively. Bharuch regression model for paddy deviated between 5.35 to 11.76% and error analysis values of MAE, MBE, RMSE, PE and average error of model were 89.21 kg ha\(^{-1}\), -89.21 kg ha\(^{-1}\), ± 102.87 kg ha\(^{-1}\), 2.68% and 3.6% respectively.

For sugarcane crop the Navsari district model diverged between 1.68 to 2.05% and error analysis values of MAE, MBE, RMSE, PE and average error of model were 526.22 kg ha\(^{-1}\), -526.22 kg ha\(^{-1}\), ± 466.46 kg ha\(^{-1}\), 5.66 % and 6.66% respectively. Similarly for Bharuch district the model deviated between -12.65 to 7.18% and error analysis values of MAE, MBE, RMSE, PE and average error of model were 479.89 kg ha\(^{-1}\), -479.89 kg ha\(^{-1}\), ± 378.59 kg ha\(^{-1}\), 14.68% and 7.9%, respectively.
TABLE 1

Statistical models for paddy, sugarcane and wheat forecast at Navsari and Bharuch districts

<table>
<thead>
<tr>
<th>Model</th>
<th>Crops</th>
<th>Equation</th>
<th>$R^2$</th>
<th>Model Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Navsari</td>
<td>Paddy</td>
<td>$Y = 20.791 + 0.019\times Rain + 0.721\times BSS + 0.362\times EP$</td>
<td>0.69</td>
<td>4.9%</td>
</tr>
<tr>
<td></td>
<td>Sugarcane</td>
<td>$Y = 875.667 - 83.72\times T\text{min} + 0.846\times BSS$</td>
<td>0.63</td>
<td>6.6%</td>
</tr>
<tr>
<td></td>
<td>Wheat</td>
<td>$Y = 64.942 - 1.687\times T\text{min}$</td>
<td>0.57</td>
<td>7.2%</td>
</tr>
<tr>
<td>Bharuch</td>
<td>Paddy</td>
<td>$Y = 17.458 + 0.328\times Rain - 0.871\times BSS + 0.687\times RH-II$</td>
<td>0.78</td>
<td>3.6%</td>
</tr>
<tr>
<td></td>
<td>Sugarcane</td>
<td>$Y = 677.254 - 7.017\times T\text{max} + 0.595\times T\text{min}$</td>
<td>0.58</td>
<td>7.9%</td>
</tr>
<tr>
<td></td>
<td>Wheat</td>
<td>$Y = 7.992 - 3.367\times T\text{max} + 3.768\times T\text{min}$</td>
<td>0.53</td>
<td>8.7%</td>
</tr>
</tbody>
</table>

TABLE 2

Validation of model for Navsari and Bharuch districts

<table>
<thead>
<tr>
<th>Crops</th>
<th>Years</th>
<th>Observed</th>
<th>Forecast</th>
<th>Error (%)</th>
<th>Observed</th>
<th>Forecast</th>
<th>Error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paddy (kg ha$^{-1}$)</td>
<td>2007</td>
<td>2580</td>
<td>2492</td>
<td>3.41</td>
<td>1646</td>
<td>1558</td>
<td>5.35</td>
</tr>
<tr>
<td></td>
<td>2008</td>
<td>2469</td>
<td>2567</td>
<td>-3.97</td>
<td>1671</td>
<td>1573</td>
<td>5.86</td>
</tr>
<tr>
<td></td>
<td>2009</td>
<td>2487</td>
<td>2419</td>
<td>2.73</td>
<td>1590</td>
<td>1482</td>
<td>6.79</td>
</tr>
<tr>
<td></td>
<td>2010</td>
<td>2577</td>
<td>2765</td>
<td>-7.30</td>
<td>1684</td>
<td>1486</td>
<td>11.76</td>
</tr>
<tr>
<td>Sugarcane (kg ha$^{-1}$)</td>
<td>2007</td>
<td>72000</td>
<td>70789</td>
<td>1.68</td>
<td>74640</td>
<td>76491</td>
<td>-2.48</td>
</tr>
<tr>
<td></td>
<td>2008</td>
<td>68700</td>
<td>67289</td>
<td>2.05</td>
<td>71010</td>
<td>73121</td>
<td>-2.97</td>
</tr>
<tr>
<td></td>
<td>2009</td>
<td>70300</td>
<td>68989</td>
<td>1.86</td>
<td>72600</td>
<td>67389</td>
<td>7.18</td>
</tr>
<tr>
<td></td>
<td>2010</td>
<td>72200</td>
<td>70989</td>
<td>1.68</td>
<td>72810</td>
<td>80211</td>
<td>-12.65</td>
</tr>
<tr>
<td>Wheat (kg ha$^{-1}$)</td>
<td>2007</td>
<td>3179</td>
<td>3333</td>
<td>-4.84</td>
<td>1930</td>
<td>2163</td>
<td>-12.07</td>
</tr>
<tr>
<td></td>
<td>2008</td>
<td>2503</td>
<td>2215</td>
<td>11.51</td>
<td>1374</td>
<td>1306</td>
<td>4.95</td>
</tr>
<tr>
<td></td>
<td>2009</td>
<td>2571</td>
<td>2749</td>
<td>-6.92</td>
<td>1559</td>
<td>1452</td>
<td>6.86</td>
</tr>
<tr>
<td></td>
<td>2010</td>
<td>3121</td>
<td>3379</td>
<td>-8.27</td>
<td>1600</td>
<td>1788</td>
<td>-11.75</td>
</tr>
</tbody>
</table>

TABLE 3

Error analysis of forecasting models for Navsari and Bharuch districts

<table>
<thead>
<tr>
<th>Crops</th>
<th>Error analysis</th>
<th>Navsari</th>
<th>Bharuch</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paddy</td>
<td>MAE (kg ha$^{-1}$)</td>
<td>87.75</td>
<td>94.25</td>
</tr>
<tr>
<td></td>
<td>MBE (kg ha$^{-1}$)</td>
<td>-16.46</td>
<td>-94.25</td>
</tr>
<tr>
<td></td>
<td>RMSE (kg ha$^{-1}$)</td>
<td>22.67</td>
<td>96.42</td>
</tr>
<tr>
<td></td>
<td>PE (%)</td>
<td>3.64</td>
<td>2.87</td>
</tr>
<tr>
<td></td>
<td>Average error (%)</td>
<td>2.2</td>
<td>4.9</td>
</tr>
<tr>
<td>Sugarcane</td>
<td>MAE (kg ha$^{-1}$)</td>
<td>917.79</td>
<td>526.22</td>
</tr>
<tr>
<td></td>
<td>MBE (kg ha$^{-1}$)</td>
<td>-64.33</td>
<td>-526.22</td>
</tr>
<tr>
<td></td>
<td>RMSE (kg ha$^{-1}$)</td>
<td>523.06</td>
<td>466.46</td>
</tr>
<tr>
<td></td>
<td>PE (%)</td>
<td>3.33</td>
<td>5.66</td>
</tr>
<tr>
<td></td>
<td>Average error (%)</td>
<td>0.88</td>
<td>6.6</td>
</tr>
<tr>
<td>Wheat</td>
<td>MAE (kg ha$^{-1}$)</td>
<td>68.53</td>
<td>67.42</td>
</tr>
<tr>
<td></td>
<td>MBE (kg ha$^{-1}$)</td>
<td>-17.64</td>
<td>-67.42</td>
</tr>
<tr>
<td></td>
<td>RMSE (kg ha$^{-1}$)</td>
<td>89.82</td>
<td>77.28</td>
</tr>
<tr>
<td></td>
<td>PE (%)</td>
<td>8.76</td>
<td>2.49</td>
</tr>
<tr>
<td></td>
<td>Average error (%)</td>
<td>0.97</td>
<td>7.2</td>
</tr>
</tbody>
</table>

In Navsari district wheat model oscillated between -8.27 to 11.51% and error analysis values of MAE, MBE, RMSE, PE and average error of model were 67.42 kg ha$^{-1}$, -67.42 kg ha$^{-1}$, ±77.28 kg ha$^{-1}$, 2.49% and 7.2% respectively, while at Bharuch district model variations were between -12.07 to 6.86% and error analysis values of MAE, MBE, RMSE, PE and average error of model were 39.6 kg ha$^{-1}$, -39.6 kg ha$^{-1}$, ± 62.11 kg ha$^{-1}$, 6.42% and 8.7%, respectively. All the models performed very well under diverse conditions.
3.2. Regression model evaluation for paddy (kg ha\(^{-1}\))

The step wise regression analysis was indicating that rainfall, bright sunshine hours and evaporation at Navsari, while at Bharuch rainfall, bright sunshine hours and maximum relative humidity showed maximum and significant effect on paddy yield. The \(R^2\) and model error percent values for paddy crop for Navsari were 0.69 and 4.9% respectively and for Bharuch, value of \(R^2\) was 0.78 and model error percent was 3.6% (Table 1).

3.3. Regression model evaluation for sugarcane (kg ha\(^{-1}\))

The step wise regression analysis was indicating that minimum temperature and bright sunshine hours at Navsari, while at Bharuch maximum temperature and minimum temperature showed maximum and significant effect on sugarcane yield. Navsari district \(R^2\) value and model error percent were 0.63 and 6.6% respectively and for Bharuch values of \(R^2\) along with model error percent were 0.58 and 7.9% respectively (Table 1). These results are comparable with the sugarcane forecasting by Singh and Bapat (1988) and Mehta et al. (2000).

3.4. Regression model evaluation for wheat (kg ha\(^{-1}\))

The step wise regression analysis was indicating that only one parameter like minimum temperature at Navsari while at Bharuch maximum temperature and minimum temperature showed maximum and significant effect on sugarcane yield. Navsari district \(R^2\) value and model error percent were 0.57 and 7.2% respectively and for Bharuch district values of \(R^2\) and model error percent were 0.53 and 8.7% respectively (Table 1). Similar results have also been reported by Singh et al. (2008).

4. Conclusion

In this paper an attempt has been made to obtain yield forecast for paddy, wheat and sugarcane by using weather data. Analysis is based on time series weather and yield data. The performance of weather variables indicate that rainfall is important for rice crop, while minimum temperature for sugarcane and wheat are more important in both the districts. The multiple regression models were found to be efficient in forecasting yields satisfactorily in respect of the three crops.

References


