Indian Journal of Agricultural Sciences 71 (7) : 487–90, July 2001

Yield forecast based on weather variables and agricultural inputs on agro-climatic zone basis

RANJANA AGRAWAL[1], R C JAIN[2] and S C MEHTA[3]

Indian Agricultural Statistics Research Institute, New Delhi 110012

Received : 31 March 1999

Key words: Yield forecast, Weather variables, Agricultural inputs, Agroclimatic zone

Crop yield is affected by weather, prevailing during crop season and inputs applied. As such modeling the cause and effect relationship between yield and these factors could provide an approach for reliable pre-harvest yield forecast. Earlier, studies were carried out at the Institute to develop forecast models based on time-series data on weather parameters and agricultural inputs for wheat (*Triticum aestivum* L. emend. Fiori & Paol.) and rice (*Oryza sativa* L.). However, these models were developed at district level and required a long series data of 25–30 years which may not be available for most of the locations. Therefore, it was thought to pool the data of various districts within agro-climatic zone so that a long series may be obtained in relatively shorter period. Hence a study was conducted to develop yield forecast model for agro-climatic zone by introducing some parameters which may take care of the variation between districts within the zone.

The study has been taken up for wheat in Vindhyachal Plateau zone of Madhya Pradesh and for rice, Chattisgarh Plain and Bastar Plateau zone were grouped together. The districts considered for wheat were Guna, Bhopal, Sagar and Raisen, whereas for rice, Durg, Raipur, Bilaspur, Raigarh and Bastar. Bastar partially falls under Chattisgarh zone and partially in Bastar zone, therefore, the 2 zones were considered together for rice as the data were available on district basis.

Time series data on weather parameters were collected for various districts from India Meteorological Department, Pune, from 1971 to 1990. Data on per cent area under irrigation, per cent area under high-yielding varieties and quantities of nitrogen, phosphorus and potassium used were collected from the Fertilizer Association of India. The corresponding district yield figures were taken from the Directorate of Economics and Statistics, Ministry of Agriculture, New Delhi. The wheat crop season is from October-end/November to late-March/April and that of rice is from June/mid-July to mid-October/November.

* Short note

[1]Principal Scientist, [2]Former Head, [3]Scientist (Senior Scale)

Forecast models

Models were developed using data up to 1989 and data for 1990 were used for validation of the models. Weekly weather data starting from a fortnight before sowing up to 1 month before harvest were utilized. Data of the pre-sowing period were considered, as this period is expected to have effect on establishment of the crop. Data for last 1 month of the crop season were excluded, as the objective of the study was to forecast yield at least 1 month before harvest. The models were basically similar to ones developed at district level (Agrawal *et al.* 1980, 1983, 1986, Jain *et al.* 1980), ie for each weather variable, 2 variables were generated—one as simple accumulation of weather variable and the other one as weighted accumulation of weekly data on weather variable, weights being the correlation coefficients of the weather variables, in respective weeks with yield, as it is or yield adjusted for trend and place effect. Similarly, for joint effect of weather variables, weekly interaction variables were generated using weekly products of weather variables taking 2 at a time. Combined models for agro-climatic zone were developed using these generated variables as regressors along with other variables. Previous year's yield, moving averages of yields and agricultural inputs were taken as the variables taking care of variation between districts within the zone. Year variable was included to take care of variation between districts within the zone. Year variable was included to take care of technological changes. On the basis of types of generated weather variables, models can be classified into following 3 groups.

Group I : This group of models was based on generated weather variables, obtained from data pooled over various districts, using data as such or deviations from district averages. For calculation of correlation coefficients to be used as weights in generated variables, yield was used as such or after adjusting for trend and/or place effect taking various combinations of year, previous year yield and 2–5 years moving averages of yield.

Group II : In this group, weather variables were generated separately for each district within the zone and models were developed using these generated variables

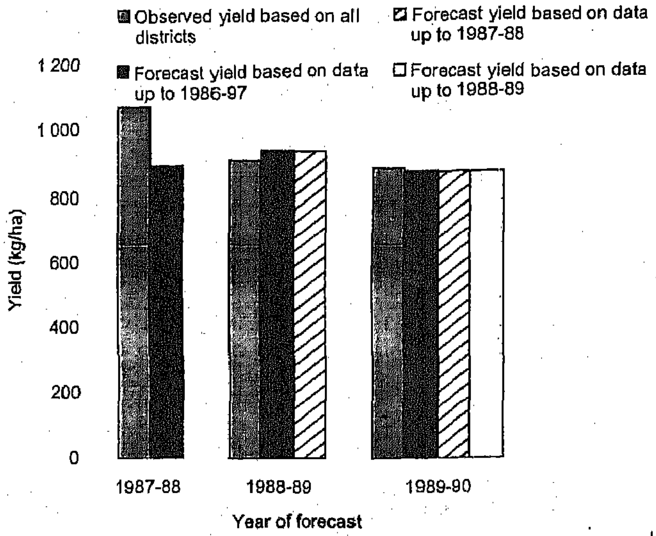

Fig 1 Wheat yield forecast at about 2 months before harvest for Vindhyachal Plateau zone

pooled over districts as such or after taking deviations of generated variables from district averages. In this group also, yield was used as such or after adjusting for trend and / or place effect.

Group III : This group of models was based on weather and agricultural inputs. In these models, agricultural inputs were also taken as regressors along with generated weather variables, year, previous year yield and moving averages of yield. In addition, one more type of model was developed where yield was adjusted for significant variables among year, previous year yield, moving averages of yield and agricultural inputs before calculating correlation coefficient between yield and weather variable to be used as weights in obtaining generated weather variables.

In all the above groups, following category of models comprising various combinations of variables were tried :

(a) $$Y = A_0 + \sum_{i=1}^{p} \sum_{j=0}^{1} a_{ij} z_{ij} + \sum_{i \neq i'}^{p} \sum_{j=0}^{1} a_{ii'j}^{i} z_{ii'j}^{i} + cT + e$$

here $Z_{ij} = \sum_{w=1}^{m} r^{j}_{iw} X_{iw}$; $Z_{ii'j} = \sum_{w=1}^{m} r^{j}_{ii'w} X_{w} X_{iw}$

Y is crop yield (kg/ha.); Ao, a_{ij}, $a_{ii'j}$ (i, i' =1,......., p, j =0,1) and c are constants; X_{iw} is value of i-th weather variable in w-th week (i=1,......6 correspond to weekly average maximum temperature, minimum temperature, total weekly rainfall, number of rainy days, average relative humidity at 830 and 1730 hrs. respectively); r_{iw} and $r_{ii'w}$ are correlation coefficients of yield with i-th weather variable and product of i-th and i'-th weather variable respectively in w-th week; m is week of forecast, e is error term distributed as N $(0,\delta^2)$.

(b) $$Y = A_0 + \sum_{i=1}^{p} \sum_{l=0}^{1} a_{ij} z_{ij} + \sum_{i \neq l'}^{p} \sum_{l=0}^{1} a_{ii'j} z_{ii'j} + by_p cT + c$$

here y_p is previous years' yield, b is constant and other symbols are same as in (a).

(c) $$Y = A_0 + \sum_{i=1}^{p} \sum_{j=0}^{1} a_{ij} z_{ij} + \sum_{i \neq i}^{p} \sum_{j=0}^{1} a_{ii'j} z_{ii'j} + cT + b_1 y_p$$

$$+ b_2 Ma_2 + b_3 Ma_3 + b_4 Ma_4 + b_5 Ma_5 + e$$

here Ma_2, Ma_3, Ma_4 + Ma_5 denote 2 to 5 years' moving averages of yield.

(d) In this category, 2 types of the models were tried, one using N, P, K separately and the other one using total fertilizer (F) besides per cent area under irrigation (I) and area under high yielding varieties (H) as variables in model (c).

Time of forecast and yield forecast

The appropriate models (identified on the basis of coefficients of determination, R^2) were developed at different points of time – week 13 onwards at weekly intervals during the crop season taking partial crop season data. Using the selected models, the yield forecast for subsequent year(s), not included in developing the models were obtained.

Table 1 Selected models

Crop	Model	Regression equation	R^2
Wheat	1	$Y = -1.4 + 13.4 Z_{11} + 0.97 Z_{151} + 2.0 Z_{451} + 2.0 Z_{451} + 0.34 Ma_5$	0.74**
	2	$Y = -907.4 + 4.0 Z_{10} + 22.8 Z_{11} + 1.34 Z_{141} + 11.4 T$	0.75**
	3	$Y = -930.02 + 2.89 Z_{10} + 18.27 Z_{11} + 3.61 Z_{121} + 0.146 Z_{561} + 11.57 T$	0.80**
	4	$Y = 0.017 + 14.1 Z_{11} + 1.6 Z_{20} - 7.6 Z_{40} + 0.09 Z_{251} + 65.2 N - 179.7 K - 44.3 P + 7.2I - 0.15 Y_p$	0.89**
Rice	1	$Y = 95.8 - 4.8 Z_{20} + 1.5 Z_{31} - 8.8 Z_{40} + 2.7 Z_{61} + 13.4 Z_{141} + 0.5 Z_{231} - 0.01 Z_{361}$	0.77**
	2	$Y = 0.00014 + 7.9 Z_{31} - 0.31 Z_{231} + 0.26 Z_{261} + 22.9 N$	0.71**

Wheat

Weekly weather data from 15 October to 4 March (20 weeks) were used for selection of models. Results indicated that the models in Group-1 were not appropriate as the coefficients of determination in different models under this group ranged between 0.35 and 0.62. Therefore, these models were dropped from further investigation. The models in Group II were found to be appropriate. The coefficients of determination in different models under this group were of the order of 0.75. Further, inclusion of agricultural inputs increased the coefficient of determination to 0.89.

The selected models (based on data up to 1988-89) along with coefficient of determination are given in Table 1. The first 3 models belong to Group-II. In Model 1, deviations of variables from corresponding district averages were utilized to adjust for variation due to districts. Using these deviations, weather variables were generated separately for

Table 2 Coefficient of determination for various models at different time of forecast

Crop	Model	Coefficient of determination for forecast week							
		13	14	15	16	17	18	19	20
Wheat	1	0.76	0.75	0.74	0.72	0.72	0.74	0.74	0.74
	2	0.66	0.66	0.68	0.68	0.73	0.75.	0.77	0.77
	3	0.74	0.73	0.79	0.74	0.73	0.77	0.79	0.80
	4	0.61	0.82	0.85	0.86	0.89	0.89	0.90	0.89
Rice	1	0.75	0.74	0.75	0.79	0.78	0.79	0.77	
	2	0.66	0.65	0.65	0.65	0.74	0.71	0.71	

each district and models were developed by taking these generated variables through step-wise regression technique. The Model 2 is based on generated weather variables using weather data as such and yield adjusted for trend effect. In Model 3, yield was adjusted for trend effect and also deviations from district averages were considered for developing the model. Model 4 belongs to Group III where deviations from district averages were used and yield was adjusted for significant variables among year, previous year's yield, moving averages of yield and agricultural inputs before calculating correlation coefficients between yield and weather variables to be used as weights for obtaining generated weather variables.

The coefficients of determination of the models at different weeks of forecast are given in Table 2. In general the earlierst appropriate time of forecast may be taken as week 13 or 14 (11 or 12 weeks after sowing, i e about 2 years and 6 months before harvest).

The yield forecast for 1989–90 was worked out using Models 1, 3 and 4 at weeks 13 and 14. Model 2 was dropped, as it did not show reliability before week 17 which is quite late. The forecasts thus obtained were compared with the observed yields. The observed yields for the zone were computed in 2 ways—one based on the districts which were used in developing forecast models and the other one based on all the districts in the zone. The idea of comparing the forecast with the observed zonal yield based on all the districts was to study the reliability of the model in forecasting the yield of the zone if data on all the districts in the zone are not available. Results indicate (Table 3) that Model 4 (using weather as well as inputs) gives forecast with 0.8% and 2.7 % deviation in week 13 and 14 respectively from the observed zonal yield based on the districts actually used in developing forcasts. The coresponding deviations from the observed yield based on all the districts were 2.5 and 1.1% respectively. Looking at these deviations along with the values of R^2, week 14 seems to be the proper time of forecast because the forecast is obtained with merely 1% deviation from observed yield based on all the districts in the zone and R^2 is 0.82.

However, in the absence of data on agricultural inputs, the Model 3, which involves data on weather only can be used. With this model, the deviations from 2 types of observed zonal yields are 4.9 and 3.3% respectively in week 14.

To study the consistency of the recommended model (Model 4) at week 14, the models were developed deleting data of 1988–89 and 1987–88 and the forecasts for subsequent years were computed. The comparison of these forecasts with the observed zonal yield based on all districts are presented in Fig 1. The per cent deviations of yield forecasts from observed yields were in general of low order except the forecast for 1987–88 from the model based on data up to 1986–87. This higher deviation may be due to the fact that the forecast in this case was based on the available data for three districts only (out of 6 in the zone).

Rice

For studying the relationship of yield with weather parameters and agricultural inputs, data from 21 May to 30 September (19 weeks) were utilized. In case of rice also, the results were almost similar to those of wheat. The models under Group-I were not appropriate as these could explain only 50% variation in yield, whereas models under Group II and III were found to be appropriate, the coefficient of determination being of the order of 0.7. The selected models are given in Table-1. Models 1 and 2 for rice are same as Models 2 and 4 in wheat respectively. The proper time of forecast was 13 or 14 weeks (Table 2). Using these models,

Table 3 Comparison of observed yield and forecasts

Crop (year of forecast)	Model	Forecast at		Observed yield		Deviation (%) of forecast from observed yield			
		Week 13	Week 14	I*	II**	I*		II**	
						Week 13	Week 14	Week 13	Week 14
Wheat (1989–90)	1	832.3	822.3	899.4	884.5	7.5	8.6	5.9	7.0
	3	798.6	855.1	899.4	884.5	11.2	4.9	9.7	3.3
	4	906.5	875.0	899.4	884.5	0.8	2.7	2.5	1.1
Rice (1990–91)	1	908.0	942.0	1282.7	1290.9	29.2	26.6	29.7	27.0
	2	1 458.7	1 206.0	1 282.7	1 290.9	13.7	6.0	13.0	6.6

I*, Observed yield based on districts used in the development of forecast models
II**, Observed yield based on all districts within the agro-climatic zone

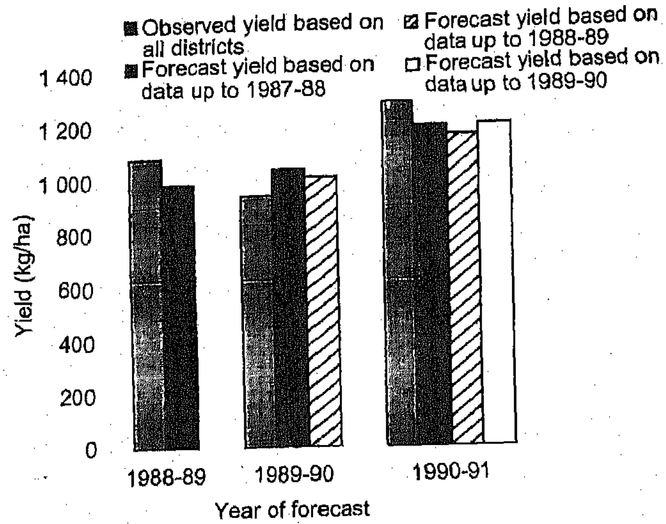

Fig 2 Rice yield forecast at about 2 months before harvest for Chhattisgarh plain and Bastar planteau zone

forecasts for 1990–91 could be obtained on the basis of only 3 districts, viz Raigarh, Bastar and Raipur, as weather data for 1999–91 were not available for other 2 districts. The model using data on weather and inputs (Model 2) was found to be appropriate in week 14 (Table 3). The per cent deviations from the 2 zonal yield averages—one based on districts actually used in yield forecast and the other based on all the districts was 6.0 and 6.6 respectively. The other model based on weather alone did not perform well. However, the deviations in rice are of higher order than those for wheat, probably due to the reason that models are based on data pertaining to 2 agro-climatic zones instead of one. It seems that the model is not able to take care of heterogeneity between 2 zones. This reflects that models should be developed zone-wise.

The consistency of the recommended model (Model 2 at week 14 was studied by developing the models taking data up to 1987–88 and 1988–89. Forecasts for subsequent years are presented in Fig 2. The per cent deviations were in the range 6.6–10.6. It may also be mentioned that due to non-availability of weather data in 1986–87 and 1988–89 for Raigarh, the forecasts in some cases are based on only 2 districts (out of 7) that too from different zones.

Thus the model using weather and agricultural inputs where weather variables could be generated separately for each district within the agro-climatic zone after adjusting the yield for place and trend effect was found to be appropriate for both the crops. Further, it is possible to forecast the yield for the zone even if data on some districts are not available either at the stage of model development or at forecasting stage.

SUMMARY

A study was carried out to develop forecast model on agro-climatic zone basis using time series data on weather variables and agricultural inputs in various districts within the zone. The study was carried out for wheat (*Triticum aestivum* L. emend. Fiori & Paol.) and rice (*Oryza sativa* L.) in Madhya Pradesh. The results indicated that reliable yield forecasts could be obtained using about 15 years data when the crops were 12 weeks old, ie about 2 months before harvest. The methodology worked even if partial data for various districts or complete data for some of the districts within the zone were missing.

REFERENCES

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

Agrawal, Ranjana, Jain R C and Jha M P. 1983. Joint effects of weather variables on rice yield. *Mausam* **34**(2) : 189–94.

Agrawal, Ranjana, Jain R C, Jha M P. 1986. Models for studying rice crop-weather relationship. *Mausam* **37**(1) : 67–70.

Jain R C, Agrawal, Ranjana and Jha M P. 1980. Effects of climatic variables on rice yield and its forecast. *Mausam* **31**(4) : 591–6.