

Weather based fuzzy regression models for prediction of rice yield

RAKHEE¹, ARCHANA SINGH¹ and AMRENDER KUMAR²

¹Amity School of Engineering & Technology, Amity, Noida

²Agricultural Knowledge Management Unit, ICAR- Indian Agricultural Research Institute, New Delhi

Email: rakheesharma234@gmail.com

ABSTRACT

Fuzzy regression models for forecasting rice yield in Kanpur district were developed and compared with the weather indices-based regression model. For this, weekly (23-35 SMW) weather data (1971, 1973-2011) were utilized. Significant variables in fuzzy approach were selected based on index of confidence (IC) and adequacy of models was compared with the weather indices-based regression models. It was found that variables such as total accumulation of minimum temperature, weighted interaction of bright sunshine hours and rainfall, weighted interaction of minimum and maximum temperature, unweighted interaction of maximum temperature and relative humidity in morning and weighted interaction of relative humidity in morning and evening respectively, are significant based on their IC and SSE (sum of square error) values. The validations of models were also attempted for three years (2008-09, 2010-11 and 2011-2012). This study also reveals that the parameters for adequacy of models for linear regression models vis-a-vis their fuzzy counterparts are much higher for all values of fitness criterion (h). Thus, fuzzy regression methodology is more efficient than linear regression technique.

Key Words: Fuzzy linear regression; multiple linear regression; IC values, SSE, SST, SSR

Reliable and timely forecasts of crop production are required for various policy decisions relating to storage, distribution, pricing, marketing, import-export, etc. Fisher *et al.* (1924) has done pioneering work in developing models for crop weather relationship and yield forecasting. Hendricks *et al.* (1943) modified Fisher's technique and divided the crop season into 'n' weekly intervals and assumed that a second-degree polynomial in week number would be sufficiently flexible to express the relationship. This model was extended for two weather variables to study joint effects. Further, since the data for such studies extended over a long period of years, an additional variant 'T' representing the year was included to make allowance for time trend. Multiple linear regression methodology has been widely employed for forecasting the growth and yield of crops (Agarwal and Mehta 2007), and insect pest population etc. (Sagar *et al.*, 2017; Kumar *et al.* 2018). The limitation of statistical regression model is that the model can be applied only if the given data are distributed according to a statistical model.

Zadeh *et al.* (1975) described the fuzzy uncertainty with ambiguity and vagueness and introduced the theory of fuzzy to build such a system as needed to deal with ambiguous and vague information. Tanaka *et al.* (1988) explained fuzzy uncertainty of dependent variables with the fuzziness of

response functions or regression coefficients in regression model and introduces initially the fuzzy regression model. Fuzzy linear regression (FLR) is a fuzzy type of classical regression analysis in which some elements of the model are represented by fuzzy numbers. It is used in evaluating the functional relationship between the dependent and independent variables in a fuzzy environment. Balve *et al.* (2016) predicted the evapotranspiration using fuzzy inference system and found that fuzzy inference system performed better. Therefore, fuzzy regression approaches have been attempted for development of model for forecasting rice yield in Kanpur district as a case study.

MATERIALS AND METHODS

In this study, yield of rice crop of Kanpur district during the year 1971, 1973 to 2011, were procured from Directorate of Economics and Statistics, Department of Agriculture, Cooperation and Farmers Welfare, India and weekly weather data were procured from India Meteorological Department, Pune. District-wise data on rice productivity and weekly weather data for Kanpur location on maximum temperature (X1), minimum temperature (X2), morning relative humidity (X3), evening relative humidity (X4) and rainfall (X5) were considered for model development. As the objective was to forecast yield well in advance of harvest,

Table 1: Coefficient of Determination (R²) for different period utilize for model development

S.No.	Data used	R ² Values
1.	23-32 SMW	0.82
2.	23-33 SMW	0.81
3.	23-34 SMW	0.82
4.	23-35 SMW	0.86

therefore, weather data from 23rd standard meteorological week (SMW) to 35th SMW were used for development of models. In this study, models were developed through weather indices (WI) based regression and fuzzy regression approach. The R-square statistics for multiple linear regression and index of confidence (IC) values for Fuzzy regression approach were utilized for checking the adequacy of developed models.

Weather indices (WI) based regression model

The regression model of the developed by Fisher *et al.* (1924) and modified by Hendricks *et al.* (1943) and applied by Agrawal *et al.* (2007) as described below has been used in the present study.

$$Y = a_0 + \sum_{i=1}^p \sum_{j=0}^1 a_{ij} Z_{ij} + \sum_{i=1}^p \sum_{j=0}^1 b_{ii'j} Z_{ii'j} + \varepsilon \quad (1)$$

where,

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

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

- Y variable to forecast
- X_{iw} value of ith weather variable in wth week
- r_{iw} correlation coefficient between Y and ith weather variable in wth week
- r_{ii'w} correlation coefficient between and product of X_i and X_{i'} in wth week
- p number of weather variables
- n₁ initial week for which weather data was included in the model
- n₂ final week for which weather data was included in the model
- E error term

Fuzzy regression approach

Fuzzy linear regression (FLR) of the following form has been used. The basic model assumes a fuzzy linear

function as

$$\tilde{Y} = \tilde{A}_0 X_0 + \tilde{A}_1 X_1 + \dots + \tilde{A}_N X_N = \tilde{A} X$$

Where X=[X₀,X₁,...X_N]^T is a vector of independent variables; $\tilde{A} = [\tilde{A}_0, \tilde{A}_1, \dots, \tilde{A}_N]^T$ is a vector of fuzzy coefficients presented in the form of symmetric triangular fuzzy numbers denoted by $\tilde{A}_j = (\alpha_j, c_j)$ with its membership function described as

$$u_{\tilde{A}_j}(a_j) = \begin{cases} 1 - \frac{|\alpha_j - a_j|}{c_j} & , \alpha_j - a_j \ll a_j \ll \alpha_j + a_j \forall j = 1, 2, 3, \dots, N \\ 0, & otherwise \end{cases} \quad (3)$$

where α_j is its central value and c_j is the spread value. Thus, the equation can be rewritten as

$$\tilde{Y}_i = (\alpha_0, c_0)X_0 + (\alpha_1, c_1)X_1 + \dots + (\alpha_N, c_N)X_N \quad (4)$$

The algorithm developed by Wang *et al.* (2000) for variable selection was utilized in this study. Those set of variables were considered for model development which has lowest error sum of square (SSE) and highest index of confidence (IC) which is a ratio between regression sum of square (SSR) and total sum of square (SST). The IC in fuzzy regression technique is as similar to the coefficient of determination (R²) in multiple linear regression techniques. Greater the value of IC and R² better are the prediction results. The values of IC measure the degree of variation Y is having, between the lower limits & upper limits represented by Y^l and Y^u respectively. In fuzzy linear regression, the regression line y^h=1 has the best ability to interpret the given data Y_i, where the membership of the fuzzy parameter A_j is symmetric as proved by Wang *et al.* (2000). The value of IC may also be calculated by IC=1- (SSE/SST), where, SSE, SST are error sum of square and total sum of square respectively. When SST is very low, the value of IC tends to close to 1, while in case when SSE equals to SST, the value of IC tends to close to 0. Thus, Higher the value of IC, lower the value of SST and better is the Y_i^h=1 used to represent Y_i.

The analysis has been done by using SAS (Statistical Analysis System), version 9.3 software, available at ICAR-IARI, New Delhi. The procedure for development of fuzzy, regression and generation of weather indices were created as per requirement for development of models and adequacy of model for a comparison between weather indices and fuzzy regression approach.

RESULTS AND DISCUSSION

Weather indices were generated for each weather variables along with its interaction and these indices were

Table 2: Analysis of variable selection through SSE and IC values

S.No.	Independent Variables Selected in the model	SSE	IC value	R2 value
1	Z20	20.75	0.925	0.652
2	Z11	20.45	0.923	0.614
3	Z121	13.33	0.929	0.702
4	Z451	20.56	0.917	0.620
5	Z231	23.06	0.893	0.631
6	Z341	18.00	0.928	0.611
7	Z121 Z131	10.46	0.924	0.748
8	Z11 Z231	20.65	0.906	0.631
9	Z41 Z141	12.69	0.927	0.765
10	Z30 Z151	18.61	0.922	0.615
11	Z130 Z251	18.90	0.922	0.619
12	Z41 Z141 Z121	11.90	0.907	0.770
13	Z50 Z341 Z251	19.64	0.901	0.684
14	Z150 Z31Z231	23.14	0.892	0.677
15	Z230 Z41 Z351	19.90	0.919	0.635
16	Z11 Z120 Z451	17.62	0.927	0.690
17	Z20 Z41 Z121 Z141	11.90	0.907	0.770
18	Z50 Z11 Z231 Z451	22.58	0.888	0.677
19	Z40 Z31 Z131 Z451	12.31	0.896	0.756
20	Z50 Z11 Z251 Z451	19.92	0.900	0.713
21	Z10 Z231 Z451 Z231	18.81	0.912	0.611
22	Z41 Z11 Z251 Z451	19.00	0.921	0.647
23	Z20 Z41 Z121 Z141 Z231	11.77	0.902	0.826
24	Z31 Z120 Z351 Z141 Z231	11.41	0.911	0.765
25	Z20 Z451 Z121 Z130 Z341	10.22	0.932	0.757
26	Z11 Z41 Z121 Z141 Z231	11.77	0.902	0.862
27	Z40 Z230 Z121 Z131 Z451	10.26	0.910	0.782
28	Z31 Z451 Z120 Z141 Z231 Z251	11.56	0.906	0.776
29	Z51 Z141 Z121 Z231 Z351 Z451	11.59	0.896	0.806
30	Z40 Z31 Z140 Z351 Z131 Z241	11.81	0.895	0.760

used as an explanatory input for a model development. Table 1 represents values of coefficient of determination (R^2) in different periods of weather data used for model development. This shows that coefficient of determination was maximum for 23-35 SMW (13 weeks after sowing), after that as week increases the coefficient of determination does not improve therefore, the data 23-35 SMW is appropriate time of forecast. For fuzzy regression technique as well, we have considered same 13 weeks (23-35 SMW) data for the

model development. Weather data of forty years were utilized (1971, 1973-2011) of which the validation of the models was attempted considering three years (2008-09, 2010-11 and 2011-2012).

In this study the objective function in fuzzy linear regression is to minimize the range (the width between upper and lower limits) of the predicted values i.e. minimizing the total spread of the fuzzy number Y_i . Weather indices along with interaction of two weather variable at a time for 5

Table 3: Observed and forecasted values along with of weather indices-based regression models and Fuzzy Regression models.

S.No.	Year	Multiple linear regression			Weather indices based fuzzy regression	
		Observed	Forecast	Width	Forecast range	Width
	2008-09	2.17	3.039	3.660	2.928±1.218	1.218
	2010-11	2.25	2.411	2.469	2.638±0.514	0.514
	2011-12	2.71	3.010	3.474	2.746±0.904	0.904

weather variables leads to 30 explanatory variables (5_{C_2} , each for weighted and unweighted) were generated. All the variables couldn't be utilized to develop a model for predicting a crop yield. Thus, for all individual and combination of variables SSE and IC values were calculated and only those set of variables were selected which were having highest IC and smallest SSE values. The results were reported for such 30 combinational sets as shown in Table 2. Based on these combinations, the variable set (Z20, Z451, Z121, Z130, Z341) has the smallest SSE with value of 10.22 and highest IC value of 0.932. Z20 is unweighted minimum temperature; Z451 interaction of relative humidity in evening with rainfall; Z121 interaction of maximum and minimum temperature; Z130 unweighted interaction of maximum temperature and relative humidity in the evening; Z341 interaction of relative humidity morning and evening. The subset (Z20, Z451, Z121, Z130, Z341) has got the highest IC value among all subsets reported, which means that the center regression line represented by $Y_{i,h=1}$ of this subset has the best ability to predict Y_i . Using these set of variables for model development for predicting rice yield will give high accuracy in the result.

A comparison among multiple linear regression (R^2 value) techniques and fuzzy linear regression (IC value) are shown in Table 2 for the selected set of variables respectively. It is derived from the table that all the IC values calculated from fuzzy regression are much greater than the R^2 values of linear regression which means fuzzy regression approach is much efficient than multiple linear regression technique and have more model stability.

The weather based fuzzy regression model for prediction of rice yield is presented as follows:

$$\hat{Y} = (1.37886, 3.21110) + (0.00001, 0.00002)Z_{20} + (0.00002, 0.00008)Z_{451} + (0.003061, 0.00000)Z_{121} + (0.00011, 0.00000)Z_{130} + (0.00027, 0.00026)Z_{341} \quad (5)$$

The developed model using set of variables (Z20, Z451, Z121, Z130, Z341) at serial number 25 were also validated for prediction of rice yield for three years viz. 2008-09, 2010-11 and 2011-2012. Table 3 represents the observed

and forecasted values of the developed model under multiple linear regression methodology and forecasted range of weather indices based fuzzy regression methodology. This table also reveals that fuzzy regression approach has narrow width as compared to the multiple regression, thus, indicates prediction closed to observed one.

CONCLUSION

Comparison of fuzzy regression and multiple linear regression method revealed that there is a considerable difference in the adequacy of models in terms of the values of IC and R^2 . The validation of models developed using variable set (Z20, Z451, Z121, Z130, Z341) having highest IC and lowest SSE values were also attempted. It was found that the fuzzy regression approach has a narrow width as compared to the multiple linear regression methodology. The forecasted range of the fuzzy approach revealed that this approach is more efficient and has more potential for predicting the crop yield in agriculture and may be considered as a good alternative over multiple linear regression.

REFERENCES

- Agrawal, R. and Mehta, S.C. (2007). Weather based forecasting of crop yields, pests and diseases-IASRI Models. *J. Ind. Soc. Agril. Stat.*, 61(2): 255-263.
- Balve P.N. and Patel J.N. (2016). Prediction of evapotranspiration using Fuzzy logic. *J. Agrometeorol.*, 18(2): 311-314.
- Fisher R. A. (1924). The influence of rainfall on the yield of wheat at Rothamsted. *R. Soc. (Land) Phil. Tans. Ser. B* 213: 89-142.
- Handrick, W.A. and Scholl, J.C. (1943). Technique in measuring joint relationship. The joint effects of temperature and precipitation on crop yield. N. Carolina Agric. Exp. Sta. Tech. Bull. 74.
- Kumar, A., Nemade, P.W., Sharma, R., Tanwar. R.K., Chattopadhyay, C., Wanjari, S.S. and Rathod. T.H. (2018). Statistical forewarning models for sucking pests of cotton in Maharashtra. *J. Agrometeorol.*, 20(1): 62-65.

- Sagar D., Nebapure S.M. and Chander S. (2017). Development and validation of weather based prediction model for *Helicoverpa armigera* in chickpea. *J. Agrometeorol.*, 19(4): 328-333.
- Tanaka H. and Watada J. (1988), Possibilistic linear systems and their application to the linear regression model, *Fuzzy Sets Sys.*, 27: 275 - 289.
- Wang H. and Tsaur R.C. (2000). Insight of a fuzzy regression model, *Fuzzy Sets Sys.*, 112: 355-369.
- Zadeh L.A. (1975). The concept of a linguistic variable and its application to approximate reasoning I. *Inform. Sci.* 8; 199-249.

Received : June 2018 ; Accepted : November 2018