Comparative Study on Fuzzy Models for Crop Production Forecasting

Amit Kumar Rana

Department of Mathematics, Swami Vivekanad Subharti University, Meerut, India

Received April 12, 2020; Revised June 2, 2020; Accepted June 16, 2020

Abstract Fuzzy sets theory is a very useful technique to increase effectiveness and efficiency of forecasting. The conventional time series is not applicable when the variable of time series are words variables i.e. variables with linguistic terms. As India and most of the Asian countries are agriculture-based economy with very smaller farmer land holding area in comparison to American, Australian and European counterparts, it becomes more important for these countries to have an approximate idea regarding future crop production. It will not only help in planning policies for future but also will be a great help for farmers and agro based companies for their future management. For small area production, soft computing techniques are important and effective tools for predicting production, as agriculture production involves a high degree of uncertainties in many parameters. In the present study, twenty-one years agricultural crop yield data is used and a comparative analysis of forecast is done with three fuzzy models. The robustness of the model is tested on real time agricultural farm production data of wheat crop of G.B. Pant University of Agriculture and Technology Pantnagar, India. As soft computing techniques involve uncertainty of the system under study, therefore it becomes more and more important for forecasting models to be accurate with the prediction. The efficiency of the three models is examined on the basis of statistical errors. The models under study are judged on the basis of Mean Square Error and average percentage error. The results of the study are in case of small area production prediction and will encourage for predicting large scale production.

Keywords Fuzzy Logical Relations (FLR), Fuzzy Sets (FS), Fuzzy Time Series (FTS), Time Invariant Model (TIM)

1. Introduction

Not many theories in mathematics become such an important application tool in a short span of time as did FS theory. After the introduction by Zadeh [21], FS theory is among the most important tool to deal with the situations involving uncertainties and vagueness in information. Since then FTS is an important soft computing technique for forecasting. Juzhand [6] explained the fuzzy statistics of classification of the data. Song and Chissom [16-18] forecasted with FTS using the data of enrollments of university. This work was highly recognized and provided a platform for forecasting methods. The work was presented in two parts explaining exclusively the enrollments forecasting. Again, the forecasting with the same data of enrollments with improved method is done by Chen [2] and produced better results. Mamdani [8] introduced the concept for approximate reasoning using fuzzy logic by taking linguistic values as variables. Papadais et al [9] proposed a fuzzy model based on load curve tested it for forecasting short term load. Forecasting using a heuristic model of FTS and a study on length of interval [4, 5] which plays a crucial role in FTS forecasting is presented and got good results. Forecasting with raw data of university enrollments using a high order FTS model [3] is given and produced more accurate results. Other soft computing techniques such artificial neural networks techniques are also very useful in FTS, [7] used this technique for prediction of milk production in a dairy. Srivastava [19] studied on soft computing techniques for forecasting. This model is used for the present study along with model of [2]. Based on FTS [14, 15] developed a simple arithmetic method and a computational method and got comparatively better results for production forecasting than other existing methods. FTS is also highly effective in financial forecast, Rubio et al [13] used a new technique in which weighted fuzzy trend time series is used and implemented it on forecasting the stock index. A new rule selection technique based on data partitioning for high order FTS models was presented by [1]. Rana [10] presented a study for FTS models for forecasting rice production. The data variation is used by Vovan [20] and
developed an improved FTSM for forecasting based on it. The nested interval based FTS model [11] was tested for fish production forecasting in India and is also used for the case under study. A study on Indian monsoon rainfall prediction using simplified arithmetic FTS model by Rana [12] provide some interesting results. Thus, it can be undoubtedly said that FTS forecasting have edge over other forecasting methods as other forecasting techniques have limited options wherever information is not clear and certain. In the present case three models are used on the raw crop yield data of the university farm. The models [2, 11, 19] used for the present study are tested on the basis of errors.

2. Fuzzy Time Series

**Definition 2.1** FS A on U is \[ \sum f_a(u_i) \] where U is universe of discourse and \( f_a \) is a function gives membership grade of \( u_i \) in A.

**Definition 2.2** Let \( F(t) = f_j(t), i = 1, 2, \ldots \) is FTS defined on U.

**Definition 2.3** \( F(t) \) is resulted due to \( F(t - 1) \) only or by \( F(t - 2) \), \( F(t - 3) \) ……. \( F(t - m) \). Then fuzzy relational equation is \( F(t) = F(t - 1) R(t, t - 1) \) (2)

And is called first order model of \( F(t) \) is
\[ F(t) = F(t - 1) \cup F(t - 2) \cup F(t - 3) \cup \ldots \cup F(t - m) R_0(t, t - m) \] (3)

**Definition 2.4** \( F(t) \) is resulted by \( F(t - 1) \), \( F(t - 2) \), \( F(t - 3) \) ……. \( F(t - m) \) simultaneously. This fuzzy relational equation (4) is called \( m \) order model.
\[ F(t) = F(t - 1) \ast F(t - 2) \ast F(t - 3) \ast \ldots \ast F(t - m) R_0(t, t - m) \] (4)

**Definition 2.5** In equations (2), (3) or (4) the fuzzy relations \( R(t, t - 1) \), \( R_0(t, t - m) \) or \( R_a(t, t - m) \) is time invariant if for \( t_1 \neq t_1 \)
\[ R(t_1, t_1 - 1) = R(t_2, t_2 - 1), or R_0(t_1, t_1 - m) = R_0(t_2, t_2 - m) \]

3. First Order Fuzzy Relations in Time

FR \( R_i(j(t, t - 1)) \) is such that \( f_j(t) = f_j(t - 1) \ast R_i(t, t - m) \)

Which is equivalent to IF \( f_j(t - 1) \) THEN rule as "IF \( f_j(t - 1) \) THEN \( f_j(t) \)"
\[ R_i(t, t - 1) \cup R_i(t, t - 1) \] (6)
\[ R_i(t, t - 1) = f_i(t) \ast f_i(t - 1) \] (7)

Here the operator \( \ast \) is called Mamdami type max min operator.

4. Algorithm for Production Forecasting

Steps involved in model development are

| STEPS | ACTION |
|-------|---------|
| 1     | Define U i.e. universal set for FTS |
| 2     | Putting U into several equal subintervals |
| 3     | On U define FS |
| 4     | To fuzzify historical data |
| 5     | Getting the FLRs |
| 6     | Getting the forecasted fuzzy output |
| 7     | Defuzzify to get forecasting crisp values |

5. Procedure for Computational Work

Applying the above rules on the 20 years (1981-82) to (2000-01) data.

**Step 1.** Take \( U = [D' - D_1, D' - D_2] \), here \( D', D'' \) are minimum and maximum value of time series data with \( D_1 \) and \( D_2 \) are fixed numbers with positive value. Let \( U = (1400, 2450, 3500) \)

**Step 2.** Divide U into subintervals \( u_a, u_b, \ldots, u_g \) as
\[ u_a = [1400, 3500], u_b = [1700, 2000], u_c = [2000, 2300] \]
\[ u_d = [2300, 2600], u_e = [2600, 2900], u_f = [2900, 3200] \]
\[ u_g = [3200, 3500] \]

**Step 3.** FS \( A_a, A_b, A_c, \ldots, A_g \) are defined labeled with word variables on U as
\( A_a \): poor crop \( A_b \): below average crop
\( A_c \): average crop \( A_g \): good crop
\( A_e \): very good crop \( A_f \): excellent crop
\( A_d \): bumper crop

The above seven FS are given by
\[ A_a : \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \]
\[ A_b : \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \]
\[ A_c : \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix} \]
\[ A_d : \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \]
\[ A_e : \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix} \]
\[ A_f : \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \]
\[ A_g : \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \]

**Step 4.** Table 1 shows fuzzified values of historical data.
Comparative Study on Fuzzy Models for Crop Production Forecasting

Step 5. FLR s obtained from the table 1 are summarized in table 2

Table 2. FLRs of the historical wheat production

| FLR | Year   | Real Yield | A_a | A_b | A_c | A_d | A_e | A_f | A_g | Fuzzify Yield |
|------|--------|------------|-----|-----|-----|-----|-----|-----|-----|---------------|
| 0.0  | 1996   | 2318       | 0.3 | 0.5 | 0.1 | 0.8 | 0.3 | 0.5 | 0.1 | 0.8           |
| 0.0  | 1997   | 2617       | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0           |
| 0.0  | 1998   | 2254       | 0.5 | 1.0 | 0.8 | 0.3 | 0.5 | 0.2 | 0.5 | 0.1           |
| 0.0  | 1999   | 2910       | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0           |
| 0.0  | 2000   | 3434       | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0           |
| 0.0  | 2001   | 2795       | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0           |

Using above table 2, we obtained a total of 15 FLR and are given in table 3.

Table 3. FLR groups

| FLR Groups | FLR | FLR | FLR | FLR | FLR | FLR | FLR | FLR | FLR | FLR | FLR | FLR | FLR | FLR | FLR | FLR |
|------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 0.0        | 1996| 1997| 1998| 1999| 2000| 2001| 1996| 1997| 1998| 1999| 2000| 2001| 1996| 1997| 1998| 1999|

Step 6. Models, we choose - Chen's arithmetic, Srivastava and Singh's refined model and Rana’s model for the current study.

Chen’s arithmetic model
1. If it has yield A_i and is only one FLR A_i → A_j in Table 3 then for the next year is A_j.
2. If it has yield A_i and FLR in table 3 are A_i → A_k1, A_i → A_k2, ……A_i → A_k_p then production for next year is mean of A_k1, A_k2, ……A_k_p

Srivastava and Singh’s refined arithmetic model
It has similar procedures of defuzzification as in Chen’s arithmetic model but it also considers the repeated FLR and keeping in view of their frequencies the weighted mean is computed and is used for defuzzification.

Rana’s Model
1. Get FLR after fuzzifying data using “if A_c, A_n is current and next year fuzzify production then the FLR is A_c → A_n where A_c is called current state and A_n is next state of fuzzified data.
2. Use FLR, obtain the fuzzified output.
3. Defuzzify values obtained in 2.
Table 4. Fuzzified Output

| Year   | $A_a$ | $A_b$ | $A_c$ | $A_d$ | $A_e$ | $A_f$ | $A_g$ |
|--------|-------|-------|-------|-------|-------|-------|-------|
| 81-82  | .5    | .5    | 1     | 1     | 1     | 1     | 1     |
| 82-83  | .5    | .5    | .8    | 1     | .8    | .8    | 1     |
| 83-84  | 0     | .5    | .5    | 1     | 1     | .8    | .5    |
| 84-85  | .3    | .5    | .8    | 1     | 1.    | .8    | .8    |
| 85-86  | .5    | .5    | 1     | 1     | 1     | 1     | 1     |
| 86-87  | .5    | .5    | 1     | 1     | 1     | 1     | 1     |
| 87-88  | .5    | .5    | 1     | 1     | 1     | 1     | 1     |
| 88-89  | 1     | .5    | 1     | .5    | 1     | .5    | .5    |
| 89-90  | 0     | .5    | .5    | .8    | 1     | 1     | .5    |
| 90-91  | .5    | .5    | 1     | 1     | 1     | 1     | 1     |
| 91-92  | 1     | .5    | 1     | .5    | 1     | .5    | .5    |
| 92-93  | 0     | .5    | 1     | .5    | .5    | .5    | .5    |
| 93-94  | 0     | .5    | .5    | .8    | 1     | 1     | .5    |
| 94-95  | .5    | .5    | 1     | 1     | 1     | 1     | 1     |
| 95-96  | .5    | .5    | 1     | 1     | 1     | 1     | 1     |
| 96-97  | .5    | .5    | 1     | 1     | 1     | 1     | 1     |
| 97-98  | 0     | .5    | .5    | .8    | 1     | 1     | .5    |
| 98-99  | .8    | .5    | .8    | 1     | .8    | .5    | 1     |
| 99-00  | 1     | .5    | 1     | .5    | 1     | .5    | .5    |
| 00-01  | 2     | .5    | 1     | 1     | 1     | 1     | 1     |

Step 7. Defuzzifying the above values:

The defuzzification of output of each of the two models is carried out as

(I) Chen’s arithmetic model

(1) If $A_i \rightarrow A_k$, then the forecast for next year will be $m_k$ i.e. mid value of $u_k$, which has maximum membership value for $A_k$.

(2) If $A_i \rightarrow A_{k_1}$, $A_i \rightarrow A_{k_2}$, ....... $A_i \rightarrow A_{k_p}$ then forecast for next year will be mean of $m_r$; $r = 1, 2, .... p$ where $m_r$ i.e. mid value of $u_r$, which has maximum membership value for $A_r$.

(3) If no FLR for $A_i$ then forecast for next year will be $m_i$ i.e. mid value of $u_i$, which has maximum membership value for $A_i$.

(II) Srivastava and Singh’s refined arithmetic model

It has similar procedures of defuzzification as in Chen’s arithmetic model but it also considers the repeated FLR and keeps in view of their frequencies, the weighted mean is computed.

(III) Rana’s Model

The Algorithm used in this model is $[S_n^*]$ is corresponding interval $u_n$ for which $A_n$ has membership $= 1$

$L[S_n^*]$ l.b. of $u_n$

$U[S_n^*]$ u.b. of $u_n$

$l[S_n^*]$ length of $u_n$

$M[S_n^*]$ mid value of $u_n$

For FLR $A_c \rightarrow A_n$

$E_c$ real data of nth year

$E_{c-1}$ real data of (n-1)th year

$E_{c-2}$ real data of (n-2)th year

$F_n$ forecasted value (n + 1)th year

Applying Algorithm for forecasting production of fish in India is as -

Fish production forecasting for (n + 1)th year, $c = 3$ to...

FLR for kth year to (k + 1)th year is $A_c \rightarrow A_n$

Now calculating the nested differences $\alpha_c$’s, $\beta_c$’s and $\gamma_c$’s using $\frac{D_c}{4}$, $\frac{3D_c}{4}$ and $D_c$ as

$D_c = ||E_c - E_{c-1}| - |E_{c-1} - E_{c-2}||$

$\alpha_c = E_c + \frac{D_c}{4}$, $\alpha'_c = E_c - \frac{D_c}{4}$, $\beta_c = E_c + \frac{3D_c}{4}$, $\beta'_c = E_c - \frac{3D_c}{4}$, $\gamma_c = E_c + D_c$, $\gamma'_c = E_c - D_c$

For using “If Then rule” for nested interval forecasting value $F_i$, $l = 1 to 6$

If $L[S_n^*] \leq \alpha_c \leq U[S_n^*]$ then $F_1 = \alpha_c$, $n_1 = 1$, Else $F_1 = 0, n_1 = 0$
If $L[S_n^*] \leq \alpha c \leq U[S_n^*]$ then $F_2 = \alpha c, n_2 = 1$, Else $F_2 = 0, n_2 = 0$

Next $I$

If $L[S_n^*] \leq \beta c \leq U[S_n^*]$ then $F_3 = \beta c, n_3 = 1$, Else $F_3 = 0, n_3 = 0$

Next $I$

If $L[S_n^*] \leq \gamma c \leq U[S_n^*]$ then $F_5 = \gamma c, n_5 = 1$, Else $F_5 = 0, n_5 = 0$

Next $I$

If $L[S_n^*] \leq \gamma' c \leq U[S_n^*]$ then $F_6 = \gamma' c, n_6 = 1$, Else $F_6 = 0, n_6 = 0$

Now $F = \sum_{i=1}^{6} F_i$

If $F = 0$ then $F_n = M[S_n^*]$

Else $F_n = (F + M[S_n^*])/(n_1 + n_2 + n_3 + n_4 + n_5 + n_6 + 1)$

Next $k$

### 6. Computational Results

The predicted outcomes are shown in the table 5.

| Year | Real production kg/ha | Chen’s Arithmetic Model | Srivastava and Singh’s Model | Rana’s Model |
|------|------------------------|-------------------------|----------------------------|--------------|
| 81-82 | 2957                   | -                       | -                          | -            |
| 82-83 | 2382                   | 2900                    | 2900                       | -            |
| 83-84 | 2572                   | 2600                    | 2650                       | -            |
| 84-85 | 2642                   | 2600                    | 2650                       | 2760         |
| 85-86 | 2700                   | 2750                    | 2836                       | 2706         |
| 86-87 | 2872                   | 2750                    | 2836                       | 2707         |
| 87-88 | 3407                   | 2750                    | 2836                       | 3350         |
| 88-89 | 2238                   | 2150                    | 2150                       | 2150         |
| 89-90 | 2895                   | 2900                    | 2850                       | 2779         |
| 90-91 | 3273                   | 2750                    | 2836                       | 3341         |
| 91-92 | 1431                   | 2150                    | 2150                       | 1550         |
| 92-93 | 2248                   | 2150                    | 2150                       | 2150         |
| 93-94 | 2857                   | 2900                    | 2850                       | 2750         |
| 94-95 | 2318                   | 2750                    | 2836                       | 2450         |
| 95-96 | 2617                   | 2600                    | 2650                       | 2750         |
| 96-97 | 2254                   | 2750                    | 2836                       | 2150         |
| 97-98 | 2910                   | 2900                    | 2850                       | 3050         |
| 98-99 | 3434                   | 2900                    | 2900                       | 3277         |
| 99-00 | 2795                   | 2150                    | 2150                       | 2750         |
| 00-01 | 3000                   | 2750                    | 2836                       | 2980         |
The results are shown graphically in figure 1.

![Figure 1. Real and predicted value](image)

7. Conclusion

Because of the edge over conventional time series when series variables are linguistics, FS theory is an important tool for forecasting. A model is judged on the basis of the difference between real and forecasting value and this is decided by MSE. The testing and results show that between two models- APE (average percentage error) are approx. 11% whereas the MSE mean sq. error (mean sq. error) for Chen model is 138458 and MSE in Srivastava and Singh model is 135105. The MSE of Rana’s model is 11671.47 with an APE 3.8585 %. This shows that Rana’s model performs far better than other two models under considerations i.e. Srivastava and Singh’s model and Chen’s model. The results also show that if there is much fluctuations in real production values then two models deviate more from actual values than Rana’s model. The study shows that soft computing techniques are more comfortable tools for forecasting over statistical models as information required in forecasting models generally involve uncertainty.

Conflict of Interest

The authors have no conflict of interest.

Acknowledgement

The data provided by the University farm administration is duly acknowledged by the authors.

REFERENCES

[1] Bose, M. & Mali, K. (2018). A Novel Data Partitioning and Rule Selection Techniques for Modeling High Order Fuzzy Time Series. Applied Soft Computing. 63, 87-96.

[2] Chen, S.M. (1996). Forecasting Enrollments Based on Fuzzy Time Series. Fuzzy Sets and Systems. 81, 311–319.

[3] Chen, S.M. (2002) “Forecasting Enrollments Based on High-Order Fuzzy Time Series”, Cybernetics and Systems: An International Journal, vol. 33, pp. 1–16.

[4] Huarng, K. (2001). Heuristic Models of Fuzzy Time Series for Forecasting. Fuzzy Sets and Systems. 123, 369–394.

[5] Huarng, K. (2001). Effective Length of Intervals to Improve Forecasting in Fuzzy Time Series. Fuzzy Sets and Systems. 123, 387–394.

[6] Juzhand, L. (1982). Fuzzy Statistics of Classification. Fuzzy Mathematics.2(4), 107.

[7] Kominakis, A.P. (2002). A Preliminary Study of the Application of Artificial Neural Networks to Prediction of Milk Yield in Dairy Sheep. Computers and Electronics in Agriculture. 35, 35-48.

[8] Mamdani, E.H. (1997). Application of Fuzzy Logic to Approximate Reasoning Using Linguistic Synthesis. IEEE Trans computers. 26, 1182-1191.

[9] Papadais, S.E., Theocharis, J.B. & Bairtzis, A.G. (2000). A Load Curve Based Fuzzy Modeling Technique for Short Term Load Forecast. Fuzzy sets and system.135, 279-303.
[10] Rana, A.K. (2018). Rice Production Forecasting Through Fuzzy Time Series. American Journal of Research in Science, Technology, Engineering & Mathematics. 23(1), 158-162.

[11] Rana, A.K. (2020). Fish Production Forecasting in India Using Nested Interval Based Fuzzy Time Series Model. International Journal of Recent Technology and Engineering. 8(6), 5534-5537

[12] Rana, A.K. (2020). Prediction of Indian Monsoon Rainfall by Interval Based Simplified High Order Fuzzy Time Series Model. International Journal of Engineering and Advanced Technology. 9(4), 782-785

[13] Rubio, A., Bermudez, J.D. & Vercher, E. (2017). Improving Stock Index Forecasts by Using a New Weighted Fuzzy-Trend Time Series Method. Expert System with Application. 76, 12–20.

[14] Singh, S.R (2007). A Simple Method of Forecasting Based on Fuzzy Time Series. Applied Mathematics and Computation. 186(1), 330-339.

[15] Singh, S.R (2008). A Computational Method Based on Fuzzy Time Series. Mathematics and Computer in Simulation. 79(3), 539-554.

[16] Song, Q. & Chissom, B.S. (1993). Fuzzy Time Series and its Models. Fuzzy Sets and Systems. 54(3), 269–277.

[17] Song, Q. & Chissom, B.S. (1993; b). Forecasting Enrollments with FTS -Part I. Fuzzy sets and systems. 54, 1-9.

[18] Song, Q. & Chissom, B.S. (1994). Forecasting Enrollments with FTS -Part II. Fuzzy sets and systems. 62,1-8.

[19] Srivastava, R.K. (2004). Study on Soft Computing Techniques for Forecasting. Ph.D. Thesis. G.B. Pant University of Agriculture and Technology, Pantnagar, India

[20] Vovan, T. (2019). An Improved Fuzzy Time Series Forecasting Model Using Variations of Data. Fuzzy Optimization and Decision Making. 18, 151-173.

[21] Zadeh, L.A. (1968). Probability Measures of Fuzzy Events. Journal of Mathematical Analysis and Applications. 23, 421-42.