



Forecasting of crop production through fuzzy time series models

Amit Kr. Rana

Assistant Professor, Department of Mathematics, Keral Verma Subharti College of Science,
Swami Vivekanand Subharti University Meerut- 250005, Uttar Pradesh, INDIA.

Abstract: This paper presents the development of fuzzy time series; time invariant model development and its implementation in forecasting the agricultural production. It contains a comparative study of three and its testing for the forecasting of wheat production of University farm. The time series forecasting is based on the historical data of 21 years of University farm. The robustness of the models is examined on the basis of error estimates. The study reveals some interesting feature of time series forecasting wheat production and provides a better platform for short term production forecast for the small area.

1. Introduction

Fuzzy Set theory is an intellectual adventure in which the philosophy of Mathematics: abstraction and idealization are combined. In fuzzy sets and system, the process of abstraction and idealization makes possible the rigorous logical deductions to linguistic variables rather strictly to the crisp numbers. A time series is a sequence of observations taken sequentially in time with an intrinsic feature that the typically adjacent observations are dependent. The time series analysis is concerned with techniques for analysis of such dependence. Future values of time series data of agricultural production process are neither exactly governed by a mathematical function nor by a probability distribution. In such a process non-conventional techniques like fuzzy time series analysis is preferred as it uses the relation of dependency, a generalization of function.

Based upon the fuzzy set theory introduced by Zadeh[1975], Song and Chissom[1993] presented the definition of fuzzy time series and outlined its modeling by means of fuzzy relation equations and approximate reasoning. Song and Chissom[1993] applied the fuzzy time series models to forecast the enrolments of the University of Alabama using historical data. **Chen (1996)** proposed an alternative simplified method of defuzzification using arithmetic operations. **Huarnng (2001; a)** proposed heuristic models of fuzzy time series for forecasting by the problem specific heuristic knowledge with Chen's model to improve the forecasting. **Yu (2005)** proposed a refined fuzzy time series model to adjust the lengths of the intervals during the formulation of fuzzy relationships. **DeOliveira (2014)** explained forecasting through distributed PSO-ARIMA-SVR hybrid system for time series forecasting. **Javedani (2014)** analyses multilayer stock forecasting model using fuzzy time series. **Rubio(2017)** presented improving stock index forecasting by using anew weighted fuzzy-trend time series method. **Ravi (2017)** fuzzy formal concept analysis based opinion mining for CRM in financial services. All these models are tested for forecasting of university enrolment of Alabama. In the present work the time invariant model have been developed refined and applied on the forecasting of wheat crop production, a non linear process, where data in general contains imprecision.

The motivation of applying the fuzzy time series forecasting models is to find ways of modeling the prediction of crop yield, a real non-deterministic process. Further, the area specific crop yield forecasting for a lead year may be applied to help the crop planning and agro based business planning of the area and can be used in economics and business analysis.

II. Fuzzy time series

Let U universe of discourse; $U = \{u_1, u_2, \dots, u_n\}$. A fuzzy set A on U is defined as

$$A = f_A(u_1)/u_1 + f_A(u_2)/u_2 + \dots + f_A(u_n)/u_n \quad (1)$$

Where f_A is membership function of A , $f: U \rightarrow [0,1]$ and $f_A(u_i)$ indicates the grade of membership of u_i in A where $f_A(u_i) \in [0,1]$ and $i = 1, 2, \dots, n$

Definition 2.1 Let $U(t)$ ($t=0,1,2,\dots$) a subset of R be universe of discourse on which fuzzy sets $f_i(t)$; ($i=1,2,\dots$) are defined and $F(t)$ be the collection of $f_i(t)$, then $F(t)$ is called fuzzy time series on $U(t)$.

Definition 2.2 Suppose $F(t)$ is caused by either $F(t-1)$ only or by $F(t-2)$ or $F(t-3)$ or $\dots$ or $f(t-m)$ $m > 0$. This relation can be expressed as the following fuzzy relational equation

$$F(t) = F(t-1) \circ R(t, t-1) \quad (2)$$

Or

$$F(t) = (F(t-1) \cup F(t-2) \cup \dots \cup F(t-m)) \circ R_0(t, t-m) \quad (3)$$

And is called first order model of $F(t)$.

Definition 2.3 Suppose $F(t)$ is caused by either $F(t-1)$, $F(t-2)$, $F(t-3)$... and $f(t-m)$ $m > 0$ simultaneously. This relation can be expressed as the following fuzzy relational equation

$$F(t) = (F(t-1) \times F(t-2) \times \dots \times F(t-m)) \circ R_a(t, t-m) \quad (4)$$

And is called m^{th} order model of $F(t)$.

Definition 2.4 If in (2) or (3) or (4) the fuzzy relation $R(t, t-1)$, $R_0(t, t-m)$ or $R_a(t, t-m)$ of $F(t)$ is independent of time t that is for different times t_1 and t_2 ,

$R(t_1, t_1-1) = R(t_2, t_2-1)$, or $R_0(t_1, t_1-1) = R_0(t_2, t_2-1)$ or $R_a(t_1, t_1-1) = R_a(t_2, t_2-1)$, then $F(t)$ is called a time invariant fuzzy time series otherwise it is called a time variant fuzzy time series.

III. First order Fuzzy relations in time invariant model

Suppose $F(t)$ as a fuzzy time series ($t = 0, 1, 2, \dots$). For the first order model $R(t, t-1)$ of $F(t)$ for any $f_i(t) \in F(t)$ where $j \in J$ there exist an $f_i(t-1) \in F(t-1)$ where $i \in I$ then the fuzzy relation $R_{ij}(t, t-1)$ is such that

$$f_j(t) = f_i(t-1) \circ R_{ij}(t, t-1) \quad (5)$$

which is equivalent to IF ... THEN rule as "IF $f_i(t-1)$ THEN $f_j(t)$ ",

$$R_i(t, t-1) = \cup_{ij} R_{ij}(t, t-1) \quad (6)$$

$$R_{ij}(t, t-1) = f_i(t-1) \times f_j(t) \quad (7)$$

Here the operator "o" is called Mamdani type max min operator.

IV. Fuzzy time series production forecasting: Algorithm

The development of fuzzy time series forecasting model and its implementation can be made with the following steps:

- | | |
|--------|---|
| Step 1 | Define the universe of discourse with given time series data on which fuzzy sets are to be defined. |
| Step 2 | Partitioning the universe of discourse into several even length intervals. |
| Step 3 | Define the fuzzy sets (linguistic variables) on universe of discourse. |
| Step 4 | Fuzzification of time series data for fuzzy input. |
| Step 5 | Computing the fuzzy relationships. |
| Step 6 | Computing the forecasted production (fuzzy output). |
| Step 7 | Defuzzification of fuzzy output for crisp forecasting. |

V. Computational Procedures

The implementation of the above algorithm for the production forecasting of the wheat crop is based on the 21 years (1995-96 to 2014-2015) time series production data of G.B.Pant Univ. of Ag.& Tech., Pantnagar (Uttarakhand) India as follows

Step 1. Define the universe of discourse to accommodate the time series data. It needs the minimum and maximum production and set as $D_{\min}$ and $D_{\max}$.

Thus universe of discourse U is defined as $[D_{\min} - D_1, D_{\max} - D_2]$, here D_1 and D_2 are two proper positive numbers. In the present case of production forecasting universe of discourse computed is $U = [1400-3500]$

Step 2. Partition the universes of discourse into 7 equal length intervals $u_1, u_2, \dots, u_7$ such that

$$\begin{aligned} u_1 &= [1400-1700], \\ u_2 &= [1700-2000], \\ u_3 &= [2000-2300], \\ u_4 &= [2300-2600], \\ u_5 &= [2600-2900], \\ u_6 &= [2900-3200], \\ u_7 &= [3200-3500] \end{aligned}$$

Step 3. Define 7 fuzzy sets $A_1, A_2, \dots, A_7$ having some linguistic values on the universe of discourse U . The linguistic values to these fuzzy variables are as follows:

A_1 : poor production,

A_2 : below average production

A_3 : average production

A_4 : good production

A_5 : very good production

A_6 : excellent production

A_7 : bumper production

These fuzzy sets in terms of its membership to different intervals are expressed as follows:

A_1	:	$[1/u_1, .5/u_2, 0/u_3, 0/u_4, 0/u_5, 0/u_6, 0/u_7]$
A_2	:	$[0.5/u_1, 1/u_2, 5/u_3, 0/u_4, 0/u_5, 0/u_6, 0/u_7]$
A_3	:	$[0/u_1, .5/u_2, 1/u_3, .5/u_4, 0/u_5, 0/u_6, 0/u_7]$
A_4	:	$[0/u_1, 0/u_2, .5/u_3, 1/u_4, 0.5/u_5, 0/u_6, 0/u_7]$
A_5	:	$[0/u_1, 0/u_2, 0/u_3, 5/u_4, 1/u_5, .5/u_6, 0/u_7]$
A_6	:	$[0/u_1, 0/u_2, 0/u_3, 0/u_4, .5/u_5, 1/u_6, .5/u_7]$
A_7	:	$[0/u_1, 0/u_2, 0/u_3, 0/u_4, 0/u_5, .5/u_6, 1/u_7]$

Step 4. Fuzzification of the time series data for the fuzzy input to the models are obtained as.

Table 1 Fuzzification of time series data of wheat production

Year	Production Kg/hect	A_1	A_2	A_3	A_4	A_5	A_6	A_7	Fuzzified Production
95-96	2730	0	0	.2	.7	1	.5	0	A_5
96-97	2957	0	0	0	.3	.8	1	.5	A_6
97-98	2382	0	.3	.8	1	.5	0	0	A_4
98-99	2572	0	0	.5	1	.8	.3	0	A_4
99-00	2642	0	0	.3	.8	1	.5	0	A_5
00-01	2700	0	0	.2	.7	1	.5	0	A_5
01-02	2872	0	0	0	.5	1	.8	.3	A_5
02-03	3407	0	0	0	0	.3	.5	1	A_7
03-04	2238	0	.5	1	.8	.3	0	0	A_3
04-05	2895	0	0	0	.5	1	.8	.3	A_5
05-06	3276	0	0	0	0	.3	.8	1	A_7
06-07	1431	1	.5	.2	0	0	0	0	A_1
07-08	2248	0	.5	1	.8	.3	0	0	A_3
08-09	2857	0	0	0	.5	1	.8	.3	A_5
09-10	2318	0	.3	.8	1	.5	0	0	A_4
10-11	2617	0	0	.3	.8	1	.5	0	A_5
11-12	2254	0	.5	1	.8	.3	0	0	A_3
12-13	2910	0	0	0	.3	.5	1	.8	A_6
13-14	3434	0	0	0	0	.3	.5	1	A_7
14-15	2795	0	0	0	.5	1	.8	.2	A_5

Step 5. The fuzzy logical relations have obtained from the table 1 are as:

Table 2 Fuzzy logical relationships of the historical wheat production

$A_1 \rightarrow A_3$	$A_3 \rightarrow A_5$	$A_3 \rightarrow A_5$	$A_3 \rightarrow A_6$	$A_4 \rightarrow A_4$
$A_4 \rightarrow A_5$	$A_4 \rightarrow A_5$	$A_5 \rightarrow A_3$	$A_5 \rightarrow A_4$	$A_5 \rightarrow A_5$
$A_5 \rightarrow A_5$	$A_5 \rightarrow A_6$	$A_5 \rightarrow A_7$	$A_5 \rightarrow A_7$	$A_6 \rightarrow A_4$
$A_6 \rightarrow A_7$	$A_7 \rightarrow A_1$	$A_7 \rightarrow A_3$	$A_7 \rightarrow A_5$	

Further the logical relations obtained in table 2 are arranged in 7 groups. Considering the relations only once, ie. leaving the repeated relations, we thus obtain a total of 15 logical relations.

Table 3 Fuzzy logical relationship groups

1	$A_1 \rightarrow A_3$					
2	$A_3 \rightarrow A_5$	$A_3 \rightarrow A_6$				
3	$A_4 \rightarrow A_4$	$A_4 \rightarrow A_5$				
4	$A_5 \rightarrow A_3$	$A_5 \rightarrow A_4$	$A_5 \rightarrow A_5$	$A_5 \rightarrow A_6$	$A_5 \rightarrow A_7$	
5	$A_6 \rightarrow A_4$	$A_6 \rightarrow A_7$				
6	$A_7 \rightarrow A_1$	$A_7 \rightarrow A_3$	$A_7 \rightarrow A_5$			

The fuzzy time invariant relation R from above 15 logical relations can be obtained as:

$R = \bigcup_{i=1}^{15} R_i$, here $\bigcup$ is the Union operator .

Where as the computations of R_i ($i=1, 2, \dots, 15$) are carried out by "IF. ..THEN" with the procedure:

If sets A and B are row vectors of order p and the logical relation between them is such that $A \rightarrow B$, then fuzzy relation R_1 is obtained in form of a matrix $C = A^T \times B$ of dimension $p \times p$, here $C_{ij} = \min(A_i, B_j)$; ($i, j = 1, \dots, p$).

Thus computing all the fuzzy logical relations $R_1, \dots, R_{15}$ and taking the union of these computed relations, we obtain a fuzzy time invariant relation R as :

$$R = \begin{pmatrix} 0 & .5 & 1 & .5 & 0 & 0 & 0 \\ 0 & .5 & .5 & .5 & .5 & .5 & .5 \\ 0 & 0 & .5 & .5 & 1 & 1 & .5 \\ 0 & .5 & .5 & 1 & 1 & .5 & .5 \\ 0 & .5 & 1 & 1 & 1 & 1 & 1 \\ .5 & .5 & .5 & 1 & .5 & .5 & 1 \\ 1 & .5 & 1 & .5 & 1 & .5 & .5 \end{pmatrix}$$

Step 6. Computation of fuzzy forecast of the crop production have been carried by the three models: Chen's arithmetic model (Model-1), Srivastava and Singh's refined arithmetic model (Model-2) and Song and Chissom model (Model-3)

Model-1

1. If the fuzzified production of the year i is A_j , and there is only one fuzzy logical relationship in the fuzzy logical relationship groups in Table 3 in which the current state of the production is A_j , then the fuzzy forecasted production of the year $i+1$ is A_j .
2. If the fuzzified production of the year i is A_j and there are fuzzy logical relationships in the fuzzy logical relationship group as:
 $A_j \rightarrow A_{k1}, A_j \rightarrow A_{k2}, \dots, A_j \rightarrow A_{kp}$, the A_j forming a relation with $A_{k1}, A_{k2}, \dots, A_{kp}$ is the fuzzy forecasted production for the year $i+1$

Model-2

Similar to model-1 but it also counts the repeated relations and the frequency is recorded and is used for defuzzification (crisp output).

Model-3

It uses the fuzzy time invariant relation R computed in sep 5. If A_{i-1} is the production of the year $i-1$, the fuzzy forecasted production for the year i will be A_i and will be computed by $A_i = A_{i-1} \circ R$. The computed fuzzy output is as

Table 4 Fuzzy out put of wheat forecasting

Year	A_1	A_2	A_3	A_4	A_5	A_6	A_7
95-96	.5	.5	1	1	1	1	1
96-97	.5	.5	.8	1	.8	.8	1
97-98	0	.5	.5	1	1	.8	.5
98-99	.3	.5	.8	1	1	.8	.8
99-00	.5	.5	1	1	1	1	1
00-01	.5	.5	1	1	1	1	1
01-02	.5	.5	1	1	1	1	1
02-03	1	.5	1	.5	1	.5	.5
03-04	0	.5	.5	.8	1	1	.5
04-05	.5	.5	1	1	1	1	1
05-06	1	.5	1	.8	1	.5	.8
06-07	0	.5	1	.5	.5	.5	.5
07-08	0	.5	.5	.8	1	1	.5
08-09	.5	.5	1	1	1	1	1
09-10	0	.5	.5	1	1	.8	.5
10-11	.5	.5	1	1	1	1	.5
11-12	0	.5	.5	.8	1	1	.5
12-13	.8	.5	.8	1	.8	.5	1
13-14	1	.5	1	.5	1	.5	.5
14-15	.2	.5	1	1	1	1	1

Step 7. Defuzzification is the process by which fuzzy output of model is transformed to crisp values for getting the forecasted values. The output of each of the three models are defuzzified in the following ways:

Model-1

- (1) If the production of the year i is A_j and fuzzy logical relation is $A_j \rightarrow A_k$ and A_k has max membership in interval u_k , then the forecasted production for the year $i+1$ will be mid point of A_k .
- (2) If the fuzzified production of the year i is A_j and there are fuzzy logical relationships in the fuzzy logical relationship group as:
 $A_j \rightarrow A_{k1}, A_j \rightarrow A_{k2}, \dots, A_j \rightarrow A_{kp}$
 $A_{k1}, A_{k2}, \dots, A_{kp}$ has max membership in the intervals $u_{k1}, u_{k2}, \dots, u_{kp}$ respectively and $m_1, m_2, \dots, m_p$ are their respective midpoints, then the forecasted production for the year $i+1$ will be $(m_1 + m_2 + \dots + m_p)/p$.
- (3) If the fuzzified production of a year is A_j , and no logical relationship is found in logical relationship groups, whose current state of production is A_j , where the maximum membership value of A_j occurs at interval u_j and the midpoint of u_j is m_j , then the forecasted production of year $i+1$ is m_j .

Model-2

Similar procedures of defuzzification as in model-1 with additional concept of repeated relations and according weighted mean is computed keeping in view of their frequencies.

Model-3

A combined approach is utilized to have the defuzzification of the fuzzy output in table 4 into crisp output. The rules are as follows;

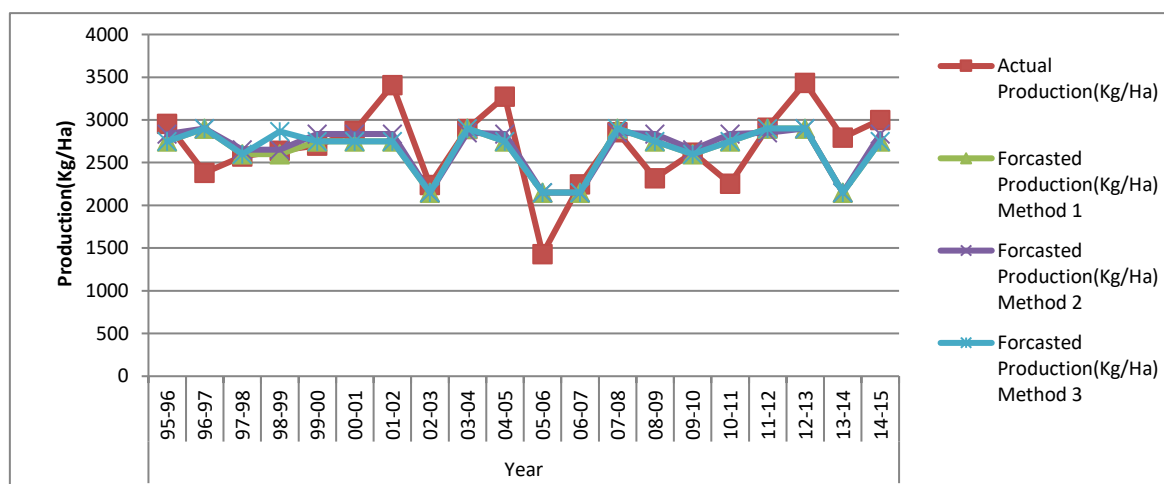
- (1) If fuzzy output set has only one max, then choose the mid point of the interval corresponding to that max as forecasted production.
- (2) If the fuzzy output has only one consecutive max, then choose the mid point of the corresponding conjunct interval as the forecasted value.
- (3) Otherwise compute the forecasted value by center of area (COA) or Centroid (COG) method.

VI. Computational Results

The forecasted production of wheat obtained by the above three models have been place in the following table and is presented in fig.1.

Table 5 Wheat production forecast

Year	Actual production kg/ ha	Forecasted Method- 1	production Forecasted Method-2	production Forecasted production Method-3
95-96	2957	2750	2836	2750
96-97	2382	2900	2900	2900
97-98	2572	2600	2650	2600
98-99	2642	2600	2650	2867
99-00	2700	2750	2836	2750
00-01	2872	2750	2836	2750
01-02	3407	2750	2836	2750
02-03	2238	2150	2150	2150
03-04	2895	2900	2850	2900
04-05	3273	2750	2836	2750
05-06	1431	2150	2150	2150
06-07	2248	2150	2150	2150
07-08	2857	2900	2850	2900
08-09	2318	2750	2836	2750
09-10	2617	2600	2650	2600
10-11	2254	2750	2836	2750
11-12	2910	2900	2850	2900
12-13	3434	2900	2900	2900
13-14	2795	2150	2150	2150
14-15	3000	2750	2836	2750



Fuzzy time series forecasting of Wheat production

VII. Conclusion

The results obtained by three fuzzy time series forecasting models are almost similar as the average forecasting error of the three models about 11%. One of the unique feature of the fuzzy time series forecasting is that it provides appreciable forecast in any trend of data, where the other stochastic or mathematical models fail. Most of work done in the area of crop production forecasting deals with the stochastic or probabilistic models, which necessarily need some probability distribution or trend in the time series data. But the fact is that agricultural production depends on various factors like weather or disease attack is a non-linear and complex phenomenon for which a probability or trend factor is difficult to locate.

Some interesting features can be observed in the fuzzy output of the model-3, which needs to be tackled by some index based defuzzification procedure for better crisp forecast. The model-3 clearly indicates that defuzzification methods like COA, COG or MOM methods are not suitable for present problem. The interesting features of fuzzy output of the model-3 can be harvested by using some higher order defuzzification method on introducing some indicators. This feature of model-3 gives an advantageous basis in comparison to other two models. The above study can provide an advantageous basis to Farm administration for better post harvest management and the local industries in planning for their raw material requirement management. The fuzzy time series forecasting can be optimally utilized in agri-business management for a lead year.

Acknowledgement

The author is thankful to University farm administration for providing the time series production data.

References

- Chen Shyi-Ming** Forecasting enrollments based on fuzzy time series, *Fuzzy sets and systems* 81[1996],311-319.
- Chen Toli, Mao-Juin J. Wang,** Forecasting methods using fuzzy concepts, *Fuzzy sets and systems* 105[1999],339-352.
- de Oliveira, J.F.L., Ludermir, T.B.:** A distributed PSO-ARIMA-SVR hybrid system for time series forecasting. In: 2014 IEEE International Conference on Systems, Man and Cybernetics (SMC), pp. 3867–3872(2014). IEEE
- Duboi S.D. and H. Parde,** Fuzzy sets in approximate reasoning part-I : interface with possibility distributions, *Fuzzy sets and systems*, 40[1991], 143-200
- E. Olmedo,** Forecasting Spanish unemployment using near neighbour and neural net techniques. *Computational Economics*, **43**(2),(2013), 183–197 (2013).
- Huang Kunhuang,** Heuristic models of fuzzy time series for forecasting, *Fuzzy sets and systems* 123[2001;a], 369-386.
- Hwang Jeng-Ren, Chen Shyi-Ming and Lee China-Hoang,** Handling forecasting problems using fuzzy time series, *Fuzzy sets and Systems* 100[1998], 217-228.
- Javedani Sadaei, H., Lee, M.H.:** Multilayer stock forecasting model using fuzzy time series. *Sci. World J.* (2014)
- Mamdani E.H.,** Application of fuzzy logic to approximate reasoning using linguistic synthesis, *IEEE Trans concept* 26[1977], 1182-1191
- Ravi, K., Vadlamani, R., Prasad, P.,** Fuzzy formal concept analysis based opinion mining for CRM in financial services. *Appl. Soft Comput.* 58, (2017), 35–52
- Rubio, A., Bermúdez, J.D., Vercher, E.:** Improving stock index forecasts by using a new weighted fuzzy-trend time series method. *Expert Syst. Appl.* 76(2017), 12–20
- Song Q. and Brad S Chissom,** Fuzzy time series and its models, *Fuzzy sets and systems*, 54[1993; a], 269-277.
- Song Q. and Brad S. Chissom** Forecasting enrollments with fuzzy time series -Part I, *Fuzzy sets and systems* 54[1993; b], 1-9.
- Song Q. and Brad S. Chissom,** Forecasting enrollments with fuzzy time series -Part II, *Fuzzy sets and systems* 62[1994],1-8.
- Wu W.,** Fuzzy reasoning and fuzzy relational equations, *Fuzzy sets and systems* 20[1986], 67-78.
- Yu, Hui-Kuang,** A refined fuzzy time series model for forecasting. *Physica A.* 346[2005], 657-681