American Journal of Mathematics and Statistics 2013, 3(5): 268-280
DOI: 10.5923/j.ajms.20130305.04

Fuzzy Time Series Modeling for Paddy (*Oryza sativa* L.) Crop Production

A. Rajarathinam*, **M. Thirunavukkarasu**

Department of Statistics, Manonmaniam Sundaranar University, Tirunelveli, Tamil Nadu, India

Abstract This paper investigates predictive performance of fuzzy time series analysis method for paddy production trends. Parametric models such as linear, non-linear and time series analysis have been conventionally used for modeling univariate and multivariate time series data sets. However, these analyses have several limitations such as assumption of the model, stationarity, normality, randomness etc. Particularly for non linear dataset, difficulties exist in time series practice. Fuzzy time series analysis is first suggested by Song and Chissom[22, 23] and it is a time-invariant method for modeling. Rather than classical methods, there are no prerequisites like stationarity and normality, and there is no necessity for treatment of missing data. Fuzzy time series methods are applied to paddy production data set and results are reported.

Keywords Fuzzy Time Series, Fuzzy Sets, Fuzzy Relational Equations, Linguistic Values, Linguistic Variables, Non-linear Models, Randomness, Normality

1. Introduction

Many different approaches such as Linear, Non-Linear, stochastic and non-stochastic models have been proposed in literature for the purpose of analyzing time series data. In recent years, the use of non-stochastic models has become widespread. Fuzzy time series forecasting models do not require assumptions that stochastic models do. On the other hand, most of the time series encountered in real life should be considered as fuzzy time series due to the uncertainty that they contain, also they should be analyzed with models appropriate to fuzzy set theory.

Chen[4] presented a simplified method of fuzzy time series forecasting of enrollments using arithmetic operations. Song and Chissom[22-24] used the fuzzy set theory given by Zadeh[28,29] to develop the time variant and time invariant models for fuzzy time series forecasting, and considered a problem of forecasting student enrollments on time series data of the University of Alabama.

A fuzzy time series method consists of three steps, those being fuzzification, identification of fuzzified relations and defuzzification, respectively. Many studies on these three steps have been done in literature because of these steps have either positive or negative impact on the forecasting perfor mance of the method. Some of the approaches proposed in the literature involve first-order forecasting models whereas some of them involve higher order forecasting models. Chen [4], Song and Chissom[22-24], Yolcu et al.,[27] can be given as examples of first-order fuzzy time series forecasting models. Also, Chen[5] and Aladag et.al.,[2] studies involve high-order fuzzy time series forecasting models.

Further, many researchers, Hwang[6], Huarng[9-10], Kim and Lee[11], Sullivan and Woodall[24], Tasi and Wu[25], worked on the development of various models of fuzzy time series forecasting and its implementations.

In the present work the proposed model was implemented on the historical Paddy crop yield forecast, which is a highly non-linear process, where data in general contains imprecision. The study is aimed to get some reliable forecast method for Paddy production during a lead year. This production forecast shall help the farmers as well as the local agro based industries in their process of business planning.

2. Materials and Methods

To achieve the stipulated objectives, the present study had been carried out on the basis of time-series production data of paddy crop pertaining to the period 1950-51 to 2009-10 had been collected through the Tamil Nadu government Agriculture Statistical office in Chennai, India. The main tools used were the nonlinear models were first checked for its appropriateness and then Fuzzy time series model approach and was concluded giving the best model that suited based on the lower values of Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Square Error (MSE) and Average Forecasting Error Rate (AFER)of the model.

2.1. Non-Linear Models

* Corresponding author:
arrathinam@yahoo.com (A. Rajarathinam)
Published online at http://journal.sapub.org/ajms

In parametric model different non-linear models given in table 1 (Bard[3], Draper and Smith[7], Montgomery[14], Ratkowsky[18], Seber and Wild[19]) were employed. Among the non-linear models, the model having highest adjusted R^2 with significant F value was selected, so that it satisfies test for goodness of fit (Montgomery[14]). Normality of residuals was examined by using Shaprio-Wilks test (Agostid'no and Stephens[1]). Further more, while dealing with time-series data it may be possible that successive observations may be auto correlated among themselves (Venugopalan and Shamasundaran[26]). To overcome all these problems, performing residual analysis is strongly advised. Randomness assumption of the residuals needs to be tested before taking any final decision about the adequacy of the model developed. To carry out the above analysis "Run test" procedure developed in the literature (Ratkowsky[18]). Further, to test the presence or absence of autocorrelation in the data set Durbin-Watson test procedure (Lewis-Beck[13]) is utilized In case of more than one model being the good fit for the data, the best model was selected having lower values of RMSE, MAE, MSE and AFER.

Table 1. List of non-linear models

Model No.	Model	Name of the Model
I.	y=a*exp(b*x) + e	Exponential
II.	y= a*exp(-exp(b-c*x)) + e	Gompertz
III.	y=a*b**(1/x)*x**c + e	Modified Hoerl
IV.	y= a/(1+b*exp(-c*x)) + e	Logistic
V.	y=(a*b+c*x**d)/(b+x**d) + e	MMF
VI.	y=a+b*cos(c*x+d) + e	Sinusoidal
VII.	y=(a+b*x)/(1+c*x+d*(x**2)) + e	Rational Function

In table 1, y is the production; x is the time points; a, b, c and d is the parameters; and e is the error term. In addition, a represents the carrying capacity, c is the intrinsic growth rate, b represents different functions of the initial value y(0) and d is the added parameters in the sinusoidal and rational function models.

The widely used Levenberg-Marquardt algorithm (Ratkowsky[18]) is utilized to fit the Modified Hoerl, Sinusoidal and rational function models. Different initial parameters values were used to ensure global convergence. The iterative procedure was stopped whenever the successive iterations parameter estimate values were negligible. The standard SPSS Ver. 10.1 package was used to fit all of the models given in table-1.

2.2. Fuzzy Time Series

Fuzzy time series is assumed to be a fuzzy variable along with associated membership function. Song and Chissom[23] have proposed a procedure for solving fuzzy time series model described in the following steps. Let U be the universe of discourse, where U={u_1,u_2,........u_n}. A fuzzy set A_i of U is defined by,

$$A_i = \mu_{Ai}(\mu_1)/u_1 + \mu_{Ai}(\mu_2)/u_2 + \ldots + \mu_{Ai}(\mu_n)/u_n \quad (1)$$

where μ_{Ai} is the membership function of fuzzy set A_i, $\mu_{Ai}: U \rightarrow [0, 1]$ $\mu_{Ai}(U_i)$ denotes the membership value of U_i in A_i, $\mu_{Ai}(\mu_i) \in [0, 1]$ and $1 \le i \le n$.

Song and Chissom[22-24] presented the following definitions of the fuzzy time series.

Definition 1: Fuzzy Time Series

Let *Y(t),* (t = 0,1,2,…), is a subset of real number R. Let *Y(t)* be the universe of discourse defined by the fuzzy set $\mu_i(t)$. If *F(t)* consists of $\mu_i(t)$ (i = 1, 2,…). *F(t)* is called a fuzzy time series on Y(t). In definition 1, *F(t)* can be viewed as a linguistic variables. This represents for the main difference between fuzzy time series and classical time series, whose values must be real numbers.

Definition 2: Time – Invariant Fuzzy Time Series

Suppose $F(t)$ is caused only by $F(t-1)$ and is denoted by $F(t-1)\rightarrow F(t)$; if there exists a fuzzy relationship between $F(t)$ and $F(t-1)$ can be expressed as the fuzzy relational equation

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

Here ''∘'' is max–min composition operator. The relation R is called first-order model of $F(t)$. Further, if fuzzy relation $R(t, t-1)$ of $F(t)$ is independent of time t, that is to say for different times t_1 and t_2, $R(t_1, t_1-1) = R(t_2, t_2-1)$, then $F(t)$ is called a time - invariant fuzzy time series. Otherwise is called a time – variant fuzzy time series.

Chen[4] revised the time-invariant models in Song and Chissom[22,23] to simplify the calculations. In addition, Chen's method can generate more precise forecasting results than those of Song and Chissom[22,23]. Chen's method is described as below :

Step 1: Collect the historical data D_{vt}.

Step 2: Define the universe of discourse U.

Find the maximum $D_{\max}$, and the minimum $D_{\min}$ among all D_{vt}. For easy partitioning of U, the small numbers D_1 and D_2 are assigned. The universe of discourse U is then defined by,

$$U = [D_{\min} - D_1, D_{\max} + D_2] \quad (2)$$

Step 3: Determine the appropriate length of interval:

Here, the average-based length method (Huarng, 2001b) can be applied to determine the appropriate l. Length of intervals significantly affects forecasting results in fuzzy time series. Hence, an effective length of intervals can significantly improve the forecasting results. The distribution based is one of the method of fuzzy time series model which can be used to adjust the lengths of intervals determined during the early stages of forecasting when the fuzzy relationship are formulated.

The distribution length of interval l is computed by the following steps:

1. Calculate all the absolute differences between the values D_{vt-1} and D_{vt} as the first differences, and then compute the average of the first differences.
2. Take one-half of the average as the length.

3. Find the located range of the length and determine the base.
4. According to the assigned base, round the length as the appropriate l.

Table 2. Base mapping table

Range	Base
0.1-1.0	0.1
1.1-10	1
11-100	10
101-1000	100
1001-10000	1000

Step 4: Define fuzzy numbers:

The number of intervals (fuzzy numbers), m, is computed by

$$m = \left(D_{\max} + D_2 - D_{\min} + D_1\right)/l. \qquad (3)$$

Thus, there are m intervals and m fuzzy numbers, which are $u_1, u_2, \ldots, u_m$, and $A_1, A_2, \ldots, A_m$, respectively. Assume that the m intervals are $u_1 = [d_1, d_2]$, $u_2 = [d_2, d_3]$, $u_3 = [d_3, d_4], \ldots,$ $u_{m-2} = [d_{m-2}, d_{m-1}]$, $u_{m-1} = [d_{m-1}, d_m]$, and $u_m = [d_m, d_{m+1}]$. The fuzzy numbers $A_1, A_2, \ldots, A_m$, can be defined as follows.

$$A_1 = (d_0, d_1, d_2, d_3),$$
$$A_2 = (d_1, d_2, d_3, d_4),$$
$$\vdots$$
$$A_{m-1} = (d_{m-2},\ d_{m-1},\ d_m,\ d_{m+1}),$$
$$A_m = (d_{m-1},\ d_m,\ d_{m+1},\ d_{m+2})$$

Step 5: Fuzzify the historical data:

If the value of D_{vt} is located in the range of u_j, then it belongs to fuzzy number A_j. All D_{vt} must be classified into the corresponding fuzzy numbers.

Step 6: Generate the fuzzy logical relationships:

For all fuzzified data, derive the fuzzy logical relationships based on definition 3:

The fuzzy logical relationship is like $A_j \to A_k$, which denotes that "if the D_{vt-1} value of time $t-1$ is A_j, then that of time t is A_k".

Step 7: Establish the fuzzy logical relationship groups:

The derived fuzzy logical relationships can be arranged into fuzzy logical relationships groups based on the same fuzzy numbers on the left hand sides of the fuzzy logical relationships. The fuzzy logical relationship groups are like the following

$$A_j \to A_{k1},$$
$$A_j \to A_{k2},$$
$$\vdots$$
$$A_j \to A_{kp}.$$

Step 8: Calculate the forecasted outputs:

The forecasted value at time t, F_{vt}, is determined by the following three heuristic rules. Assume the fuzzy number of D_{vt-1} at time $t-1$ is A_j.

Rule 1:

If the fuzzy logical relationship group of A_j is empty; $A_j \to \phi$, then the value of F_{vt} is A_j, which is $(d_{j-1},\ d_j,\ d_{j+1},\ d_{j+2})$.

Rule 2:

If the fuzzy logical relationship group of A_j is one to one; $A_j \to A_k$, then the value of F_{vt} is A_k, which is $(d_{k-1},\ d_k,\ d_{k+1},\ d_{k+2})$.

Rule 3:

If the fuzzy logical relationship group of A_j is one to many $A_j \to A_k$, $A_j \to A_{k2}, \ldots,$ $A_j \to A_{kp}$, and then the value of F_{vt} is calculated as follows.

$$F_{vt} = \frac{A_{k1} + A_{k2} + \ldots + A_{kp}}{p}$$
$$= \left(\frac{d_{k1-1} + \ldots + d_{kp-1}}{p},\ \frac{d_{k1} + \ldots + d_{kp}}{p},\ \frac{d_{k1+1} + \ldots + d_{kp+1}}{p},\ \frac{d_{k1+2} + \ldots + d_{kp+2}}{p}\right)$$

where,

$$A_{k1} = (d_{k1-1}, d_{k1}, d_{k1+1}, d_{k1+2}),$$
$$A_{k2} = (d_{k2-1}, d_{k2}, d_{k2+1}, d_{k2+2}),$$
$$\vdots$$
$$A_{kp} = (d_{kp-1}, d_{kp}, d_{kp+1}, d_{kp+2}).$$

2.3. Performance of Models

The following measures of goodness of fit have been used to judge the adequacy of the model developed.

Root Mean Square Error (RMSE) =

$$\left[\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 / n\right]^{1/2}$$

Mean Absolute Error (MAE) = $\sum_{i=1}^{n}\left|(Y_i - \hat{Y}_i\right| / n$, and

Mean Square Error (MSE) = $\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 / (n-p)$

Average Forecasting Error Rate (AFER) =

$$\frac{\left|Y_i - \hat{Y}_i\right| / Y_i}{n} \times 100\%$$

where n and p are number of observations and number of parameters, respectively in the model. The lower the values of these statistics, the better were the fitted model.

As pointed out by Kvalseth[12], before taking any final decision about the appropriateness of the fitted model, it is paramount importance to investigate the basic assumptions regarding the error term, viz., randomness and normality.

Randomness assumption of the residuals needs to be tested before taking any final decision about the adequacy of the model developed. To carry out the above analysis "Run test" procedure is developed in the literature

2.4. Test for the Randomness of the Residuals

To test the randomness of the residuals the following hypothesis was tested

H_0 : the set of residual was random Vs H_1 : the set of residual was not random

Let N1 be the number of residuals of one type and N2 be the number of residuals of other type. r is the number of runs (sequence of symbols of one kind separated by symbols of another kind). If both N1 and N2 are less than or equal to 20, then table in the Appendix gives the critical values of r under H_0 for $\alpha = 0.05$. If the observed value of r falls between critical values, H_0 cannot be reject. If the observed value of r is equal to or more extreme than one of the critical values, H_0 should be rejected (Sidney Siegel and John Castellan N, Jr.[21]).

2.5. Test for Normality of the Residuals

The Shapiro – Wilk[20] statistic was used to test whether the residuals are normally distributed or not. The test is based on *n* residuals. These were arranged in non – decreasing sequence and is designated by $e_{(1)}$, $e_{(2)}$, $e_{(3)},\ldots,e_{(n)}$. The following hypothesis was to be tested.

H_0: The residuals are normally distributed Vs H_1: These are not normally distributed.

The required test statistic W is defined as $W = S^2 / b$

where $S^2 = \sum a(k)\left[e(n+1-k) - e(k)\right]$. The parameter K takes the values

$$K = \begin{cases} 1,2,3,4, \ldots, n/2 & \text{when n is even} \\ 1,2,3,4,\ldots, (n-1)/2 & \text{when n is odd and} \end{cases}$$

$$b = \sum_{i=1}^{n}\left(e_i - \bar{e}\right)^2$$

The values of coefficients "a(k)" for different values of n and k are given in table 5 (Shapiro – Wilk[20]). H_0 was accepted if the value of W is very close to one.

3. Results and Discussion

Different non-linear models have been employed to study trends in the paddy crop production data presented in Appendix. The characteristic fitted non-linear models are presented in table 3. Among the non-linear models, appropriate model has been selected based on highest R^2 values, significance of the model parameters, lowest values of RMSE and MAE values. Residuals of the selected model would be tested for randomness and normality. Further Fuzzy time series model was employed and finally appropriate model was selected based on lowest values of MSE, RMSE, MAE and AFER which is desirable. The findings are discussed in sequence, as follows.

3.1. Non-Linear Models Fitting for Paddy Crop Production Data

Among the non-linear model fitted to the production of paddy crop, the result presented in table 3 reveals that, the maximum adjusted R^2 of 63 % was observed in case of Rational function model with minimum values of MSE (692484.40), RMSE (832.16) and MAE (635.93). However the residuals due to this model were not found to be normally distributed since the S-W test statistic value was significant. Hence the Rational functional model was not found to be suitable model to fit the trends in production of the paddy crop.

But the Sinusoidal model had next highest adjusted R^2 of 61 % with comparatively lower values of MSE (728757.80), RMSE (853.68) and MAE (651.04). Also all the estimated parameter values have been found within 95% confidence interval indicating that parameter values were significant (table 3). The model able to explain 62% of variation presents in the paddy production data. Hence among the Non-linear model fitted the following Sinusoidal model was found suitable to fit the trends in Paddy crop production. The observed and predicted values are given in table 4. and depicted in Figure 1.

$Y = 4560.98^{*}+1528.15^{*}\cos(0.0695^{*}x-3.1527^{*})$ (R^2=61%)

Similar type of trend was also reported by Rajarathinam and Vinoth[17] for the wheat crop grown during the period from 1950-51 to 2009-10 in India.

Rajarathinam and Parmar[15] reported that none of the linear and non-linear models have been found suitable to fit the trends on castor crop grown during the period from 1949-50 to 2007-08 in middle Gujarat region, India.

Rajarathinam et al.,[16] used the Rational function model to fit the trends in production of tobacco crop grown during the period 1949-50 to 2007-08 in Anand region, Gujarat State, India.

From the above discussion it can be concluded that appropriateness of the model is influenced by the type of crop as well as location.

Table 3. Characteristics of fitted Non-Linear models for the paddy crop production

		Model		
		Modified Hoerl	**Rational**	**Sinusoidal**
Parameters	**A**	1763.03* (394.35)	2902.96* (243.04)	4560.98* (274.14)
	B	1.4191* (0.6895)	-33.22* (13.30)	1528.15* (223.03)
	C	0.3076* (0.0592)	-0.0284* (0.0015)	0.0695* (0.0136)
	D	-	0.0002* (0.000023)	-3.1527* (0.5437)
Goodness of fit	**R^2 %**	55**	64**	62**
	Adj. R^2 %	54	63	61
	S-W test	0.414	0.027	0.108
	Run Test	0.037	0.068	0.068
	D-W Test	1.0285	1.2643	1.1999
	MSE	849528.9	692484.4	728757.8
	RMSE	921.6989	832.16	853.68
	MAE	698.9775	635.93	651.04

*Significant at 5% level **Significant at 1% level Values in ( ) indicates standard error

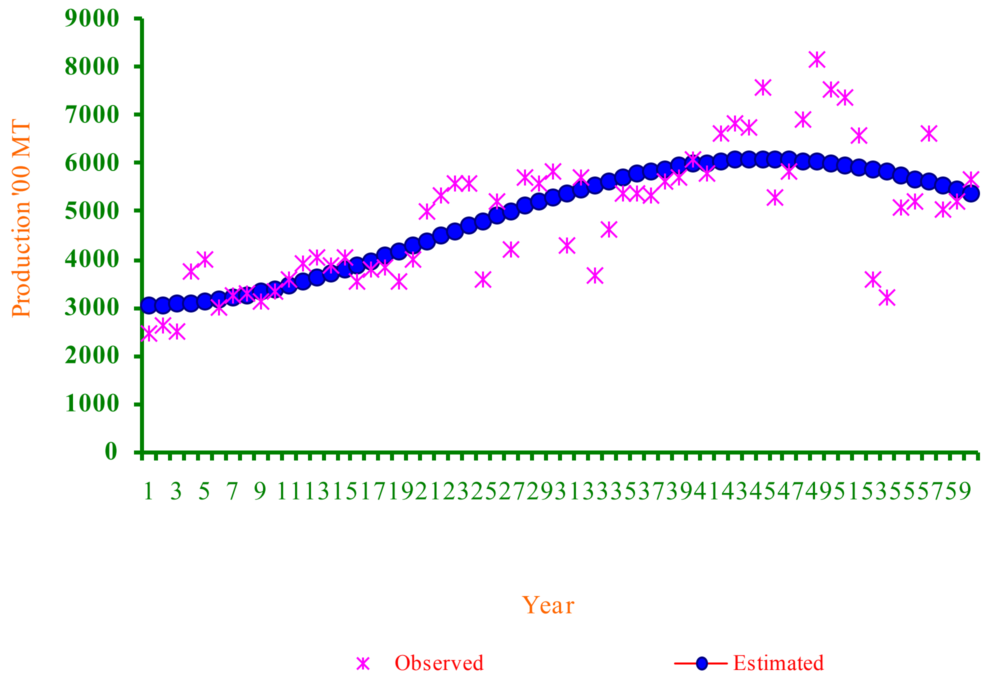

Figure 1. Trends in paddy crop production based on Sinusoidal *Non-linear* Model

Table 4. Observed and Estimated Values for paddy productions based on Sinusoidal *Non-linear* Model

Year	Observed	Estimated	Forecasting error rate	Year	Observed	Estimated	Forecasting error rate	Year	Observed	Estimated	Forecasting error rate
1951	2459	3035	0.3168	1971	5001	4376	0.3048	1991	5782	6021	0.1509
1952	2635	3045	0.2246	1972	5302	4482	0.3561	1992	6596	6049	0.2022
1953	2489	3063	0.3122	1973	5569	4589	0.3065	1993	6806	6069	0.1828
1954	3749	3087	0.3573	1974	5558	4695	0.4255	1994	6750	6083	0.4024
1955	3995	3119	0.4685	1975	3575	4800	0.1013	1995	7559	6089	0.2183
1956	3002	3157	0.0819	1976	5203	4905	0.2638	1996	5290	6087	0.0750
1957	3247	3203	0.0234	1977	4215	5007	0.1949	1997	5805	6079	0.2285
1958	3288	3255	0.0169	1978	5705	5108	0.1130	1998	6894	6063	0.5801
1959	3134	3313	0.0900	1979	5559	5206	0.1571	1999	8141	6039	0.4225
1960	3333	3377	0.0217	1980	5800	5300	0.3440	2000	7532	6009	0.3894
1961	3559	3447	0.0542	1981	4279	5392	0.0615	2001	7366	5971	0.1846
1962	3907	3523	0.1820	1982	5681	5479	0.5751	2002	6584	5927	0.6521
1963	4024	3603	0.1948	1983	3642	5561	0.2973	2003	3577	5876	0.7436
1964	3876	3688	0.0851	1984	4633	5639	0.1011	2004	3223	5819	0.2009
1965	4036	3777	0.1142	1985	5365	5712	0.1178	2005	5062	5755	0.1398
1966	3524	3870	0.1491	1986	5370	5779	0.1449	2006	5209	5686	0.2967
1967	3791	3967	0.0739	1987	5333	5840	0.0797	2007	6611	5612	0.1483
1968	3846	4066	0.0903	1988	5614	5895	0.0674	2008	5040	5532	0.0809
1969	3550	4168	0.2471	1989	5704	5944	0.0216	2009	5183	5448	0.0952
1970	4012	4271	0.1014	1990	6063	5986	0.0660	2010	5665	5359	0.1509

3.2. Fuzzy Time Series Modeling

Fuzzy time series modeling have been employed to the paddy production data and the results are being discussed as follows.

The following steps have been carried out to estimate the paddy production.

Step 1: Algorithm of this method is being implemented on the Paddy crop production grown in the state of the Tamil Nadu, India during the period 1951 to 2010.

Step 2: Appendix shows that the maximum and the minimum number of the paddy productions are 8141 $(D_{\max})$ and 2459 $(D_{\min})$, respectively. For easy computation, let $D_1 = 459$ and $D_2 = 359$. The universe of discourse *U* is defined as follows:

$$U = [2459 - 459,\ 8141 + 359]$$
$$= [2000,\ 8500]$$

Step 3: The appropriate length of interval l can be computed as follows:

1. Based on table 2, we can calculate the average of the first differences, which is 667.
2. Take one-half of 667 as the length, which is 333.5.
3. Since the length 333.5 is located at the range[101, 1000] in table 2, the base is assigned to be 100.
4. According to the base 100, the length 333.5 is rounded off to 300, which is the appropriate length of interval l.

Step 4: Use Eq. 3 to calculate the number of intervals (fuzzy numbers) as follows:

$$m = \frac{8500 - 2000}{300} = 21.67 = 22.$$

Thus, there are 22 intervals, which are

$$u_1 = [2000, 2300],\ u_2 = [2300, 2600],$$
$$u_3 = [2600, 2900],\ u_4 = [2900, 3200],$$
$$u_5 = [3200, 3500],\ u_6 = [3500, 3800],$$
$$u_7 = [3800, 4100],\ u_8 = [4100, 4400],$$
$$u_9 = [4400, 4700],\ u_{10} = [4700, 5000],$$
$$u_{11} = [5000, 5300],\ u_{12} = [5300, 5600],$$
$$u_{13} = [5600, 5900],\ u_{14} = [5900, 6200],$$
$$u_{15} = [6200, 6500],\ u_{16} = [6500, 6800],$$
$$u_{17} = [6800, 7100],\ u_{18} = [7100, 7400],$$
$$u_{19} = [7400, 7700],\ u_{20} = [7700, 8000],$$
$$u_{21} = [8000, 8300],\ u_{22} = [8300, 8600].$$

The fuzzy numbers can be defined by

$$A_1 = (1700, 2000, 2300, 2600),$$
$$A_2 = (2000, 2300, 2600, 2900),$$
$$A_3 = (2300, 2600, 2900, 3200),$$
$$A_4 = (2600, 2900, 3200, 3500),$$
$$\vdots$$
$$A_{19} = (7100, 7400, 7700, 8000),$$
$$A_{20} = (7400, 7700, 8000, 8300),$$
$$A_{21} = (7700, 8000, 8300, 8600),$$
$$A_{22} = (8000, 8300, 8600, 8900).$$

Step 5: Fuzzify the productions. For example, the paddy production in year 1951 is 2459, which is located at the range of $u_2 = [2300, 2600]$. Thus, the corresponding fuzzy number of year 1951 is assigned as A_1.

table 5 lists the corresponding fuzzy number for the paddy production of each year.

Step 6: According to table 5, we can derive the fuzzy logical relationships as shown in table 6. Notice that the repeated relationships are counted only once.

Step 7: Based on the same fuzzy numbers on the left hand side of the fuzzy logical relationships in table 6, 17 fuzzy logical relationship groups are generated as shown in table 7.

Step 8: According to tables 5 and 7, we can calculate the forecasted paddy productions. For instance, the forecasted paddy productions of years 1952 and 1954 can be illustrated below:

Forecasting 1952: The fuzzified paddy production of year 1951 in table 5 is A_1, and from table 7, we can find that there is one fuzzy logical relationships in group 1. $A_3 \to A_2$.

According to *Rule 2,* the forecasted paddy productions of year 1952 is A_2. Thus,

$$F_{v1952} = (2000, 2300, 2600, 2900) = 2450.$$

Forecasting 1953: According to table 7, we can find that there is one fuzzy logical relationships in group 1. $A_2 \to A_6$. The forecasted paddy production of year 1953 is A_6. Thus,

$$F_{v1953} = (2300, 2600, 2900, 3200) = 3650.$$

Forecasting 1954: Because the fuzzified paddy production of 1954 in table 5 is A_6, and from table 7, we can find that there are five logical relationships in group 3.

$A_6 \to A_7$, $A_6 \to A_6$, $A_6 \to A_{11}$, $A_6 \to A_9$, $A_6 \to A_5$.

According to *Rule 3*, the forecasted paddy production of year 1954 is computed as follows:

$$F_{v1954} = \frac{A_7 + A_6 + A_{11} + A_9 + A_5}{5}$$
$$= \frac{3950 + 3650 + 5150 + 4550 + 3350}{5} = 4130.$$

Table 5. Corresponding fuzzy numbers of the paddy productions

Year	Production	Fuzzy Number	Estimated Values	Year	Production	Fuzzy number	Estimated Values	Year	Production	Estimated Values	Fuzzy number
1951	2459	-	-	1971	5001	A_{11}	5450	1991	5782	5536	A_{13}
1952	2635	A_3	2450	1972	5302	A_{12}	4950	1992	6596	5825	A_{16}
1953	2489	A_2	3650	1973	5569	A_{12}	4950	1993	6806	7400	A_{17}
1954	3749	A_6	4130	1974	5558	A_{12}	4950	1994	6750	5825	A_{16}
1955	3995	A_7	3950	1975	3575	A_6	4130	1995	7559	6200	A_{19}
1956	3002	A_4	3350	1976	5203	A_{11}	5450	1996	5290	5450	A_{11}
1957	3247	A_5	3800	1977	4215	A_8	5750	1997	5805	5536	A_{13}
1958	3288	A_5	3800	1978	5705	A_{13}	5536	1998	6894	7400	A_{17}
1959	3134	A_4	3350	1979	5559	A_{12}	4950	1999	8141	7550	A_{21}
1960	3333	A_5	3800	1980	5800	A_{13}	5536	2000	7532	6200	A_{19}
1961	3559	A_6	4130	1981	4279	A_8	5750	2001	7366	6650	A_{18}
1962	3907	A_7	3950	1982	5681	A_{13}	5536	2002	6584	5825	A_{16}
1963	4024	A_7	3950	1983	3642	A_6	4130	2003	3577	4130	A_6
1964	3876	A_7	3950	1984	4633	A_9	4550	2004	3223	3800	A_5
1965	4036	A_7	3950	1985	5365	A_{12}	4950	2005	5062	5450	A_{11}
1966	3524	A_6	4130	1986	5370	A_{12}	4950	2006	5209	5450	A_{11}
1967	3791	A_6	4130	1987	5333	A_{12}	4950	2007	6611	5825	A_{16}
1968	3846	A_7	3950	1988	5614	A_{13}	5536	2008	5040	5450	A_{11}
1969	3550	A_6	4130	1989	5704	A_{13}	5536	2009	5183	5450	A_{11}
1970	4012	A_7	3950	1990	6063	A_{14}	5750	2010	5665	5536	A_{13}

Similarly, we can calculate the forecasted paddy production for the remaining years. Table 8 shows that the paddy productions during 1951 to 2010.

The observed and estimated paddy production data based on fuzzy time series modeling is given in the table 8 and depicted in the Fig.2.

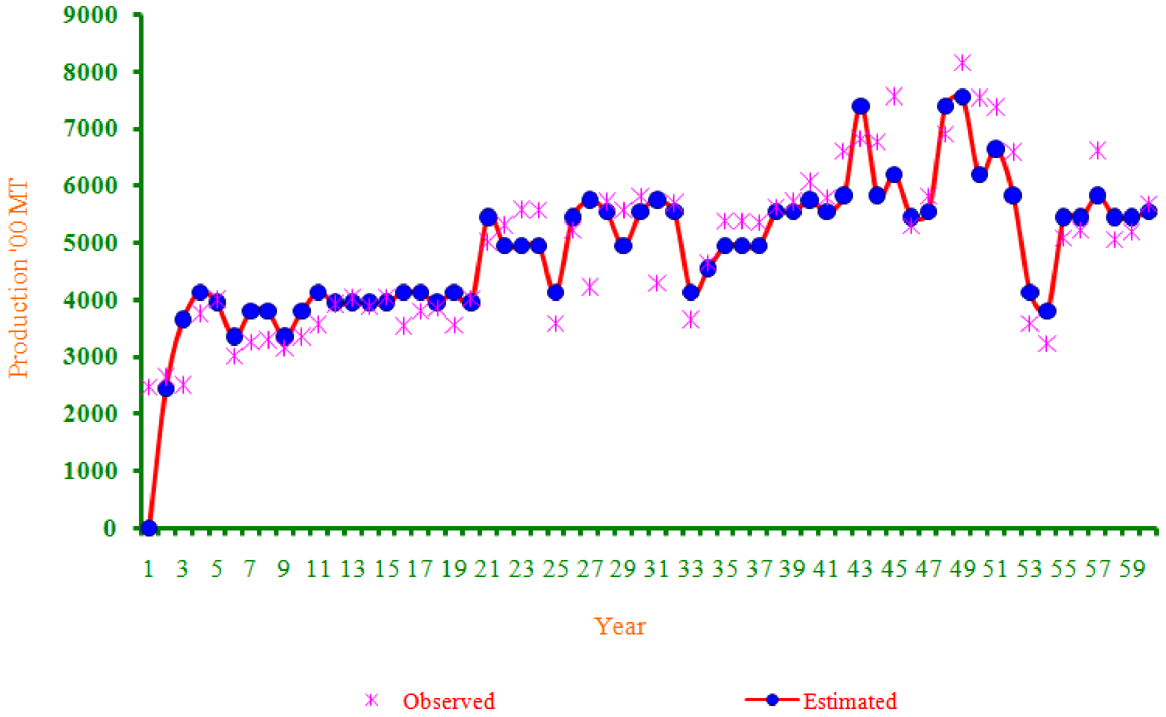

Figure 2. Trends in paddy production based on Fuzzy time series modeling

Table 6. Fuzzy Logical relationships of the paddy productions

$A_3 \rightarrow A_2$	$A_2 \rightarrow A_6$	$A_6 \rightarrow A_7$	$A_7 \rightarrow A_4$	$A_4 \rightarrow A_5$	$A_5 \rightarrow A_5$
$A_5 \rightarrow A_4$	$A_4 \rightarrow A_5$	$A_5 \rightarrow A_6$	$A_6 \rightarrow A_7$	$A_7 \rightarrow A_7$	$A_7 \rightarrow A_7$
$A_7 \rightarrow A_7$	$A_7 \rightarrow A_6$	$A_6 \rightarrow A_6$	$A_6 \rightarrow A_7$	$A_7 \rightarrow A_6$	$A_6 \rightarrow A_7$
$A_7 \rightarrow A_{11}$	$A_{11} \rightarrow A_{12}$	$A_{12} \rightarrow A_{12}$	$A_{12} \rightarrow A_{12}$	$A_{12} \rightarrow A_6$	$A_6 \rightarrow A_{11}$
$A_{11} \rightarrow A_8$	$A_8 \rightarrow A_{13}$	$A_{13} \rightarrow A_{12}$	$A_{12} \rightarrow A_{13}$	$A_{13} \rightarrow A_8$	$A_8 \rightarrow A_{13}$
$A_{13} \rightarrow A_6$	$A_6 \rightarrow A_9$	$A_9 \rightarrow A_{12}$	$A_{12} \rightarrow A_{12}$	$A_{12} \rightarrow A_{12}$	$A_{12} \rightarrow A_{13}$
$A_{13} \rightarrow A_{13}$	$A_{13} \rightarrow A_{14}$	$A_{14} \rightarrow A_{13}$	$A_{13} \rightarrow A_{16}$	$A_{16} \rightarrow A_{17}$	$A_{17} \rightarrow A_{16}$
$A_{16} \rightarrow A_{19}$	$A_{19} \rightarrow A_{11}$	$A_{11} \rightarrow A_{13}$	$A_{13} \rightarrow A_{17}$	$A_{17} \rightarrow A_{21}$	$A_{21} \rightarrow A_{19}$
$A_{19} \rightarrow A_{18}$	$A_{18} \rightarrow A_{16}$	$A_{16} \rightarrow A_6$	$A_6 \rightarrow A_5$	$A_5 \rightarrow A_{11}$	$A_{11} \rightarrow A_{11}$
$A_{11} \rightarrow A_{16}$	$A_{16} \rightarrow A_{11}$	$A_{11} \rightarrow A_{11}$	$A_{11} \rightarrow A_{13}$		

Table 7. Fuzzy logical relationship groups

Group	Fuzzy Logical Relationships
1	$A_3 \rightarrow A_2$
2	$A_2 \rightarrow A_6$
3	$A_6 \rightarrow A_7$, $A_6 \rightarrow A_6$, $A_6 \rightarrow A_{11}$, $A_6 \rightarrow A_9$, $A_6 \rightarrow A_5$
4	$A_7 \rightarrow A_4$, $A_7 \rightarrow A_7$, $A_7 \rightarrow A_6$, $A_7 \rightarrow A_{11}$
5	$A_4 \rightarrow A_5$
6	$A_5 \rightarrow A_5$, $A_5 \rightarrow A_4$, $A_5 \rightarrow A_6$, $A_5 \rightarrow A_{11}$
7	$A_{11} \rightarrow A_{12}$, $A_{11} \rightarrow A_8$, $A_{11} \rightarrow A_{13}$, $A_{11} \rightarrow A_{11}$, $A_{11} \rightarrow A_{16}$, $A_{11} \rightarrow A_{13}$
8	$A_{12} \rightarrow A_{12}$, $A_{12} \rightarrow A_6$, $A_{12} \rightarrow A_{13}$
9	$A_8 \rightarrow A_{13}$
10	$A_{13} \rightarrow A_{12}$, $A_{13} \rightarrow A_8$, $A_{13} \rightarrow A_6$, $A_{13} \rightarrow A_{13}$, $A_{13} \rightarrow A_{14}$, $A_{13} \rightarrow A_{16}$, $A_{13} \rightarrow A_{17}$
11	$A_9 \rightarrow A_{12}$
12	$A_{14} \rightarrow A_{13}$
13	$A_{16} \rightarrow A_{17}$, $A_{16} \rightarrow A_{19}$, $A_{16} \rightarrow A_6$, $A_{16} \rightarrow A_{11}$
14	$A_{17} \rightarrow A_{16}$, $A_{17} \rightarrow A_{21}$
15	$A_{19} \rightarrow A_{11}$, $A_{19} \rightarrow A_{18}$
16	$A_{21} \rightarrow A_{19}$
17	$A_{18} \rightarrow A_{16}$

Table 8. Observed and Estimated Values for paddy productions based on Fuzzy Time series Model

Year	Observed	Estimated	Forecasting error rate	Year	Observed	Estimated	Forecasting error rate	Year	Observed	Estimated	Forecasting error rate
1951	2459	-	-	1971	5001	5450	0.1497	1991	5782	5536	0.0711
1952	2635	2450	0.1169	1972	5302	4950	0.1107	1992	6596	5825	0.1949
1953	2489	3650	0.7775	1973	5569	4950	0.1853	1993	6806	7400	0.1455
1954	3749	4130	0.1695	1974	5558	4950	0.1824	1994	6750	5825	0.2284
1955	3995	3950	0.0189	1975	3575	4130	0.2589	1995	7559	6200	0.2996
1956	3002	3350	0.1932	1976	5203	5450	0.0792	1996	5290	5450	0.0504
1957	3247	3800	0.2836	1977	4215	5750	0.6071	1997	5805	5536	0.0774
1958	3288	3800	0.2598	1978	5705	5536	0.0494	1998	6894	7400	0.1224
1959	3134	3350	0.1150	1979	5559	4950	0.1825	1999	8141	7550	0.1210
1960	3333	3800	0.2335	1980	5800	5536	0.0759	2000	7532	6200	0.2948
1961	3559	4130	0.2672	1981	4279	5750	0.5730	2001	7366	6650	0.1621
1962	3907	3950	0.0183	1982	5681	5536	0.0426	2002	6584	5825	0.1920
1963	4024	3950	0.0307	1983	3642	4130	0.2231	2003	3577	4130	0.2576
1964	3876	3950	0.0317	1984	4633	4550	0.0300	2004	3223	3800	0.2985
1965	4036	3950	0.0356	1985	5365	4950	0.1290	2005	5062	5450	0.1279
1966	3524	4130	0.2865	1986	5370	4950	0.1305	2006	5209	5450	0.0770
1967	3791	4130	0.1491	1987	5333	4950	0.1196	2007	6611	5825	0.1981
1968	3846	3950	0.0452	1988	5614	5536	0.0231	2008	5040	5450	0.1356
1969	3550	4130	0.2723	1989	5704	5536	0.0491	2009	5183	5450	0.0857
1970	4012	3950	0.0256	1990	6063	5750	0.0861	2010	5665	5536	0.0381

It is clear from the table 9 that the values of RMSE, MAE, MSE and AFER obtained using Fuzzy time series models were found to be much lower than that obtained through the non-linear model, thereby indicating the superiority of Fuzzy time series over the parametric approach. Hence Fuzzy time series model have been justified to fit the trends in production of paddy crop. The graph of the fitted trend for production of paddy crop using Fuzzy time series model is depicted in Fig.2.

Table 9. Characteristics of Model Performance parameters

	RMSE	MAE	MSE	AFER
Non-Linear Model	853.68	651.04	728757.80	0.2156%
Fuzzy Time Series Model	582.12	463.74	338865.45	0.1660%

4. Conclusions

After the through analytical implementation of different non-linear and fuzzy time models over the paddy production data it can be finally concluded that, the Fuzzy time series method was justified to be a suitable method to fit the trends in paddy crop production. Fuzzy time method can be used as a more suitable one since it proved to be dynamic and versatile enough to be considered for the statistical interpretation for the trends in paddy crop production for the years to come.

Appendix

Paddy production of the Tamil Nadu data during the period 1951 to 2010

Year	Production	Year	Production	Year	Production
1951	2459	1971	5001	1991	5782
1952	2635	1972	5302	1992	6596
1953	2489	1973	5569	1993	6806
1954	3749	1974	5558	1994	6750
1955	3995	1975	3575	1995	7559
1956	3002	1976	5203	1996	5290
1957	3247	1977	4215	1997	5805
1958	3288	1978	5705	1998	6894
1959	3134	1979	5559	1999	8141
1960	3333	1980	5800	2000	7532
1961	3559	1981	4279	2001	7366
1962	3907	1982	5681	2002	6584
1963	4024	1983	3642	2003	3577
1964	3876	1984	4633	2004	3223
1965	4036	1985	5365	2005	5062
1966	3524	1986	5370	2006	5209
1967	3791	1987	5333	2007	6611
1968	3846	1988	5614	2008	5040
1969	3550	1989	5704	2009	5183
1970	4012	1990	6063	2010	5665

REFERENCES

[1] Agostid'no, R.B., and Stephens, M.A. (1986). Goodness of fit technologies. Marcel Dekker, New York.

[2] Aladag , C.H., Basaran, M.A., Egrioglu, E., Yolcu, U. and Uslu, V.R. (2009), Forecasting in high order fuzzy times series by using neural networks to define fuzzy relations, Expert Systems with Applications, Volume 36, Issue 3, 4228-4231.

[3] Bard, Y. (1974). Nonlinear Parameter Estimation. Academic Press: New York.

[4] Chen, S.M. (1996). Forecasting enrollments based fuzzy time series. Fuzzy Sets and Systems, 81, 311-319.

[5] Chen, S.M. (2002). Forecasting enrollments based on higher-order fuzzy time series. Cybernetics and Systems: An International Journal, 33, 1-16.

[6] Chen, S.M. and Hwang, J.R. (2000). Temperature prediction using fuzzy time series. IEEE Transactions on Systems, Man and Cybernetics – Part B : Cybernetics, 30, 263 – 275.

[7] Draper, N.R., and Smith, H. (1998). Applied Regression Analysis, 3rd Edn., John Wiley & Sons, USA.

[8] Liu, H.T. (2007). An improved fuzzy time series forecasting method using trapezoidal fuzzy numbers, 6, 63-80.

[9] Huarng, K.H. (2001a). Heuristic models of fuzzy time series for forecasting. Fuzzy Sets and Systems, 123, 369-386.

[10] Huarng, K.H. (2001b). Effective lengths of intervals to improve forecasting in fuzzy time series. Fuzzy Sets and Systems, 100, 217-228.

[11] Kim, I., and Lee, S.R. (1999). A fuzzy time series prediction method based on consecutive values. IEEE international fuzzy fuzzy systems conference proceedings, II, August 22-25, Seoul, Korea , 730-707.

[12] Kvalseth, T. O. (1985). Cautionary note about R^2. The Amer. Statistician, 39(4):279-285.

[13] Lewis Beck, S.M. (1993). Regression Analysis. Sage Publications, New York.

[14] Montgomery, D.C., Peck, E.A., and Vining, G.G. (2003). Introduction to Linear Regression Analysis, John Wiley & Sons, Inc.

[15] Rajarathinam, A., and Parmar, R.S. (2011). Application of parametric and nonparametric regression models for area, production and productivity trends of Castor (Ricinus communis L.) crop. Asian Journal of Applied Sciences. 4(1), 45-52.

[16] Rajarathinam, A., Parmar, R.S. and Vaishnav, P.R. (2010). Estimating models for area, production and productivity trends of tobacco (Nicotaina tabacum) crop for Anand region of Gujarat state, India. Journal of Applied Sciences. 10(20) : 2419-2425.

[17] Rajarathinam, A., and Vinoth B. (2012). Statistical modeling for wheat (Triticum Aestivum) crop production. International Journal of Statistics and Applications. 2(4) : 40-46.

[18] Ratkowsky, D.A. (1990). Handbook of Non-linear Regression Models. Marcel Dekker, New York.

[19] Seber, G.A.F. and Wild, C.J. (1989). Non-Linear Regression. John Wiley and Sons.

[20] Shapiro, S. S. and Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika,* 52: 591-611.

[21] Sidney Siegel and John Castellan, N. Jr. (1988). Nonparametric statistics for the behavioral sciences. McGraw-Hill, New York.

[22] Song, Q. and B.S. Chissom (1993a). Fuzzy forecasting enrolments with fuzzy time series-Part 1. Fuzzy Sets and Systems, 54, 1-9.

[23] Song, Q. and B.S. Chissom (1993b). Fuzzy time series and its models. Fuzzy Sets and Systems, 54, 269-277.

[24] Sullivan, J. and Woodall, W.H. (1994). A comparison of fuzzy forecasting and Makov modeling. Fuzzy Sets and Systems, 64, 279-293.

[25] Tsai, C. C., Wu, S.J. (2001). Forecasting local region data with fuzzy time series. In ISIE 2001, Pusan, Korea, Proceedings, IEEE, 122-133.

[26] Venugopalan, R and Shamasundaran, K.S. (2003). Nonlinear Regression. A realistic modeling approach in Horticultural crops research. Jour.Ind.Soc.Statistics 56(1), 1-6.

[27] Yolcu, U., Egrioglu, E. and Uslu, V.R. Basaran, M.A. and Aladag, C.H., (2009). A New Approach for Determining the Length of Intervals for Fuzzy Time Series, Applied Soft Computing, 9, 647-651.

[28] Zadeh, L.A. (1965). Fuzzy set. Information and Control, 8, 338-353.

[29] Zadeh, L.A. (1975). The concept of a linguistic variable and its applications to approximate reasoning – Part 1. Information Sciences, 8, 199-249.