

## Soft Computing Techniques: An Application to Short Term Forecast of Inland Fish Production of India

\*<sup>1</sup>Vinod K. Yadav, <sup>3</sup>M.Krishnan, <sup>2</sup>Rama Sharma, <sup>3</sup>Ramasubramanian V, <sup>1</sup>Vidya S.Bharti , <sup>3</sup>N.R.Kumar

<sup>1</sup>Scientist, Central Institute of Fisheries Education, Mumbai 400061, India

<sup>2</sup>Senior Scientist, Central Institute of Fisheries Education, Mumbai 400061, India

<sup>3</sup>Principal Scientist, Central Institute of Fisheries Education, Mumbai 400061, India

\*corresponding author e-mail: [vinod\\_iasri@yahoo.co.in](mailto:vinod_iasri@yahoo.co.in)



**Abstract:** Time series models have been utilized to make accurate predictions in production. Various forecasting methods have been developed on the basis of fuzzy time series data, but accuracy has been matter of concern in these forecasts. The historical data of inland fish production of India have been taken to implement the model. The study uses the fuzzy sets theory of Zadeh [1] and fuzzy time series models based on order and time-invariant method introduced by Song and Chissom [2], Chen [3], Sah and Degtiarev[4].The study also uses an improved and versatile method for fuzzy time series forecasting using a difference parameter as fuzzy relation for forecasting introduced by Singh [5] and first order and time-variant methods given by Chen and Hsu [6]. The forecast to inland fish production have also been obtained by developing an Artificial Neural Network (ANN) model using Back propagation algorithm. The study is aimed to find the inland fish production forecast for a lead year by using different fuzzy time series models and back propagation algorithm for the forecast. The forecasted inland fish productions, obtained through these techniques, have been compared and their performance has been examined. Present infers that performance of ANN model is similar or less performer as comparison to fuzzy time series methods when the dependence is linear. In such cases, uses of ANN only lead to increase in complexity without significant improvement in the performance.

**Keywords:** Fuzzy Time Series, fuzzy Set, Production, Forecasting, Linguistic Value, fuzzified production. fuzzy logical relationships, back propagation algorithm.

### INTRODUCTION

Forecasting plays an important role in our daily life. Future value of a time series data of many processes are neither exactly governed by mathematical function nor by probability distribution. The fish production and productivity is one of such processes, which is not governed by any deterministic process due to highly non – linearity caused by various affective production parameters like weather, sunshine , diseases, disaster etc. In such processes, the soft computing techniques: Fuzzy time series and artificial neural network be preferred as these use the relation of dependency, a knowledge by learning process of set of training data. A classical time series methods can not deal with forecasting problems in which the values of time series are linguistic terms represented by fuzzy sets Zadeh

[1]. Fuzzy time series forecasting emerged as a novel approach for predicting the future values in a situation where neither a trend is viewed nor a pattern in variations of time series are visualized and moreover when information (data) are imprecise and vague. Both the models: stochastic and dynamic in the time series analysis use the probability structure of sequence of observations to construct the forecast function. The unique features of fuzzy time series models are that it uses the relation, a generalization of functions. Therefore, Song and Chissom [2] presented the theory of fuzzy time series to overcome this drawback of the classical time series methods. Based on the theory of fuzzy time series, Song [2][7][8] presented some forecasting methods to forecast the enrollments of the University of Alabama. Chen [3] presented a method to forecast the enrollments of the University of Alabama based on fuzzy time series. It had the advantage of reducing the steps in calculation, time and simplifying the calculation process. Hwang, Chen and Lee [9][10] used the differences of the enrollments to present a method to forecast the enrollments of the University of Alabama based on fuzzy time series. Huarng [11] used simplified calculations with the addition of heuristic rules to forecast the enrollments using Chen [3]. Chen [12] presented a forecasting method based on high-order fuzzy time series for forecasting the enrollments of the University of Alabama. Chen and Hsu [6] presented a first order time variant method for fuzzy time series for forecasting the enrollments of the University of Alabama. Sah and Degtiarev [4] utilized the variations of the available historical data as fuzzy time series instead of direct usage of raw numerical values.

Singh [5] presented an improved and versatile method for fuzzy time series forecasting using a difference parameter as fuzzy relation for forecasting. All these models have been implemented to forecast the enrollments of the University of Alabama.

In the present paper the different fuzzy time series models introduced by Chen [3], Chen and Hsu [6], Sah and Degtiarev [4], Singh[5] and Artificial Neural Network model using Back Propagation Algorithm(BPA) , have been implemented on the historical inland fish production forecast. The study is aimed to get some reliable forecast for inland fish production for a lead year. This production forecast will help the fish farmers as well as the local fish based industries in their business planning.

## MATERIALS AND METHODS

### Fuzzy Time Series Models

Let  $Y(t)$  ( $t = \dots, 0, 1, 2, \dots$ ), is a subset of  $R_1$ , be the universe of discourse on which fuzzy sets  $f_i(t)$  ( $i = 1, 2, \dots$ ) are defined and  $F(t)$  is the collection of  $f_i(t)$  ( $i = 1, 2, \dots$ ). Then  $F(t)$  is called fuzzy time series on  $Y(t)$  ( $t = \dots, 0, 1, 2, \dots$ ). Further  $F(t)$  can be understood as a linguistic variable and  $f_i(t)$  ( $i = 1, 2, \dots$ ) as the possible linguistic values of  $F(t)$ .

Definition 1: Suppose  $F(t)$  is caused by a  $F(t-1)$  only or by  $F(t-1)$  or  $F(t-2)$  or...or  $F(t-m)$  ( $m > 0$ ). This relation can be expressed as the following fuzzy relational equation:

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

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

The equation is called the first order model of  $F(t)$ .

Definition 2: Suppose  $F(t)$  is caused by a  $F(t-1)$ ,

$F(t-2), \dots$ , and  $F(t-m)$  ( $m > 0$ ) simultaneously. This relation can be expressed as the following fuzzy relational equation

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

and is called the  $m^{\text{th}}$  order model of  $F(t)$ .

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

$$R(t_1, t_1-1) = R(t_2, t_2-1), \text{ or } R_a(t_1, t_1-m) = R_a(t_2, t_2-m) \text{ or } R_0(t_1, t_1-m) = R_0(t_2, t_2-m), \text{ then}$$

$F(t)$  is called a time invariant fuzzy time series. Otherwise it is called a time variant fuzzy time series,

In the case of time invariant fuzzy time series,

$$R(t, t-1) = R,$$

$$R_a(t, t-m) = R_a(m),$$

$$R_0(t, t-m) = R_0(m)$$

In general at different times  $t_1$  and  $t_2$ ,  $R(t_1, t_1-1) \neq R(t_2, t_2-1)$ ,  $R_a(t_1, t_1-m) \neq R_a(t_2, t_2-m)$  and  $R^\circ(t_1, t_1-m) \neq R^\circ(t_2, t_2-m)$ .

There are two reasons for this: first, the universes of discourse on which the fuzzy sets are defined may be different at different times: second the value of  $F(t)$  at different times may be different.

Depending upon the complexity of the system, fuzzy time series modeling for a forecast process may use type of relations  $R(t, t-1)$ ,  $R_a(t, t-m)$ ,  $R_0(t, t-m)$ . Development of fuzzy time series model essentially depends on the procedure of fuzzy relations generated between the observations at a time  $t$  and among the observations at

different times. Several methods Dubois and Parde (1991), Wu (1986) and Mamdani (1977) are available to determine these relations.

### Inland Fish Production Forecasting using fuzzy time series model (method)

Fuzzy time series model deals with situation where the data are linguistic values, in contrast to the conventional time series approaches that typically manipulate numerical data. If data are available in crisp form, it is to be fuzzified before the fuzzy time series methodology can be applied. Fuzzification process starts with defining the universe of discourse, which contains the historical data and upon which the fuzzy sets are defined. The study deals with the inland fish production of India (in Lakh Kg) in various years starting from 1995-96 to 2009-2010 with assumption that it includes some vagueness incurred due to statistical sampling.

The inland fish production forecasting is obtained and compared using fuzzy time series method given by Chen and Hsu [6], Singh[5], Chen [3] and Sah and Degtiarev[4] method (Table 1).

With the comparison of actual production of inland fish production of India with the forecasted production by Chen and Hsu [6], one can conclude that the forecasted results are very close to that of actual result. The MSE and Average Error (%) of the forecasting results of the Chen and Hsu [6] method is smaller than that of Singh [5], Sah & Degtiarev [4] & Chen [3] methods. Hence, the Chen and Hsu [6] method can get a higher forecasting accuracy rate for forecasting inland fish production of India than the Singh[5], Sah and Degtiarev [4] & Chen[3] other methods studied

### Inland Fish Production Forecasting using Artificial Neural Network Model

Back Propagation through time is a powerful tool of artificial neural network with application to many areas as pattern recognition, dynamic modeling and nonlinear systems. Back propagation algorithm (BPA) provides an efficient way to calculate the gradient of the error function using chain rule of differentiation. The error after initial computation in the forward pass is propagated backward from the output units, layer by layer. BPA, a generalized Delta rule is commonly used algorithm for supervised training of multi layer feed forward artificial neural network. It requires a teacher that knows, or can calculate, the desired output for any given input (fig.1). It is most useful for feed-forward networks (networks that have no feedback, or simply, that have no connections that loop). The term is an abbreviation for "backwards propagation of errors". Back propagation requires that the activation function used by the artificial neurons (or "nodes") be differentiable.

In general following steps are followed to implement the algorithm:

1. Select the training pattern; design the neural network (say network) by choosing the number of nodes for the input, hidden and output layers
2. Initialize weights for all node connections to arbitrary values between -1 .0 and 1.0
3. Normalized the production data so that every value must be in between 0 and 1.
4. Choose a value between 0 and 1.0 for the learning parameter
5. Repeat for each training pattern X in the training set

DO

- i.  $E$  = neural-net-output (network, X); forward pass
- ii.  $O$  = teacher output for X
- iii. Calculate error ( $E - O$ ) at the output units
- iv. Compute change in weights for all weights from hidden layer to output layer; backward pass
- v. Compute change in weights for all weights from input layer to hidden layer; backward pass continued.
- vi. Update the weights in the network.

END

6. Until all patterns classified correctly or stopping criterion satisfied, return trained network.
7. Test the accuracy of the network on a test dataset. If the accuracy is less than optimal, change one or more parameters of the network topology and start over.

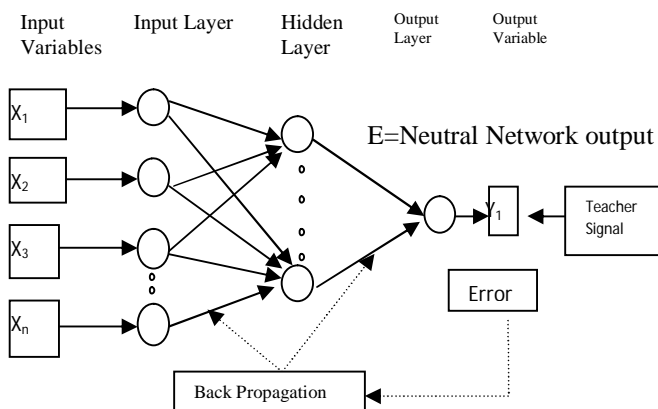


Fig. 1: Typical structure of artificial neural network

Input to Hidden node ( $j$ ) =  $I_j = \sum x_i W_{ji}$  where  $x_i$  =input value for  $i$  th year and  $W_{ji}$  = the wt associated with link between hidden node  $j$  and input node  $i$

Output from Hidden node ( $j$ ) =  $O_j = y_j = f(I_j) = 1 / [1 + e^{-I_j}]$

Error at output node ( $k$ ) = Error ( $k$ ) =  $(T - O_k) * O_k * (1 - O_k)$  where  $T$  = the target output and  $O_k$  = the computed output at node  $k$

Error at hidden node ( $j$ ) =  $(\sum \text{Error}(k) W_{jk}) * O_j * (1 - O_j)$  where  $O_j$  = computed output at hidden node  $j$  and  $W_{jk}$  = the wt associated with link between hidden node  $j$  and output node  $k$

Error at hidden node ( $j$ ) if output has only one node = Error ( $k$ ) \*  $W_{jk} * O_j * (1 - O_j)$

Weight adjustment formula-  $W_{jk}(\text{new}) = W_{jk}(\text{current}) + \Delta W_{jk}$

And  $\Delta W_{jk} = r [\text{Error}(k)] (O_j)$  where  $r$  = the learning rate parameter with  $1 > r > 0$

#### Method Used:

**Method I:** In this method, the actual production obtained from historical data have been used for the training in the development process of the ANN model and forecasted production obtained by **Chen and Hsu [6]** method through fuzzy time series have been used as desired output for validation of developed model outside the training set as higher forecasting accuracy rate for forecasting inland fish production are obtained by this method as compared to Singh [5], Sah and Degtiarev[4] & Chen[3] other methods studied (**Table 1**). The models have been developed for forecast based on different training sets and validation the steps adopted as:

**Step1:** Production of year 1995-96 to 2004-2005 as input set and fuzzy output of 2005-06 as desired output.

**Step2:** Production of year 1995-96 to 2005-2006 as input set and fuzzy output of 2006-07 as desired output.

**Step3:** Production of year 1995-96 to 2006-2007 as input set and fuzzy output of 2007-08 as desired output.

**Step4:** Production of year 1995-96 to 2007-2008 as input set and fuzzy output of 2008-09 as desired output.

**Step5:** production of year 1995-96 to 2008-2009 as input set and fuzzy output of 2009-10 as desired output.

**Method II:** In this method, forecasted inland fish production obtained by **Chen and Hsu [6]** method through fuzzy time series have been used for the training of ANN model development and actual production has been used for the validation of developed model outside the training set. The model has been developed for forecast based on different training sets and validation the steps adopted are as:

**Step 1:** Fuzzy output of year 1988-89 to 2004-2005 as input set and actual production of 2005-06 as desired output.

**Step 2:** Fuzzy output of year 1988-89 to 2005-2006 as input set and actual production of 2006-07 as desired output.

**Step 3:** Fuzzy output of year 1988-89 to 2006-07 as input set and actual production of 2007-08 as desired output.

**Step 4:** Fuzzy output of year 1988-89 to 2007-08 as input set and actual production of 2008-09 as desired output.

**Step 5:** Fuzzy output of year 1988-89 to 2008-09 as input set and actual production of 2009-10 as desired output.

### Computation of inland fish production forecast with Neural Network method

The Neural Network method is being implemented on the data of inland fish production of India

**Step 1.** The historical time series data of production are obtained and further normalized to bring all the values in the range of 0 and 1 (Table 2 & Table 3)

**Step 2.** The proposed Neural Network has three layers. Input layer, hidden layer and output layer. The hidden layer has 2 neurons and output layer has one neuron only.

**Step 3.** Back propagation algorithm is used as training algorithm.

**Step 4.** Transfer function of hidden layer and output layer are Sigmoid Function.

**Step 5. Base SAS** is used as tool to write the program for proposed ANN

**Step 6.** Actual production obtained from historical data have been used for the training in the development process of the ANN model and forecasted production obtained by **Chen and Hsu [6]** method through fuzzy time series have been used as desired output for validation of developed model outside the training set in the Neural Network Method I and Forecasted inland fish production obtained by **Chen and Hsu[6]** method through fuzzy time series have been used for the training of ANN model development and actual production have been used for the validation of developed model outside the training set in the method II

**Step 7.** After De- normalization we get the forecasted value (Table 2 & Table 3).

**Step 8** The forecasted values for the inland fish production have been obtained by using the above said computational algorithm. Comparative studies of forecasts by various methods is presented in table 4.

**Step 9.** We have done the error analysis for all the methods used for the comparative study. Table 4 Shows the MSE and average error for various forecasted methods used in this paper. The algorithm has been implemented through Base SAS programming language, considering two hidden layers and computations have been made by various iterations levels like: 25, 50, 100, 150, 100, 200 & 300. Out of these, the best suitable forecasted values have been obtained by model with 25 iterations. The results so obtained have been illustrated in Table. 2 and Table. 3.

### RESULTS & DISCUSSION

In time series forecasting, the forecasting accuracy of a model is commonly measured in terms of Mean Square Error (MSE) or in terms of Average Error. Lower the MSE or average error, better the forecasting method. MSE is defined as

Mean Square Error =

$$\frac{\sum_{i=1}^n (\text{actual value}_i - \text{forecasted value}_i)^2}{n}$$

and forecasting error as

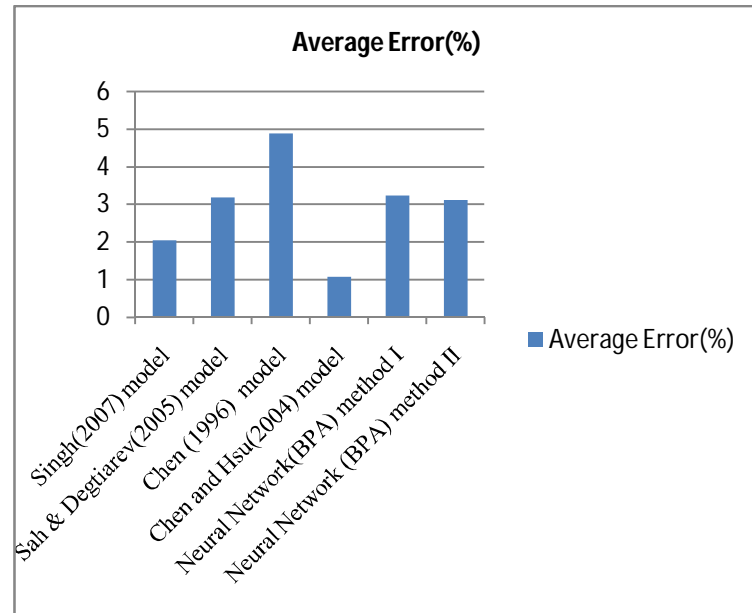
Forecasting error (in percent) =

$$\frac{|\text{forecasted} - \text{actaul value}|}{\text{actual value}} \times 100$$

Average forecasting error (in percent)

$$= \frac{\text{sum of forecasting error}}{\text{numbers of errors}}$$

With the above comparison of actual production of inland fish production of India with the forecasted production by [6], one can conclude that the forecasted results are very close to that of actual result. The MSE and Average Error (%) of the forecasting results of the [6] method is smaller than that of [3][4][5] and Neural Network (BPA method- I and II) methods.(fig.2) Hence, the [6] method can get a higher forecasting accuracy rate for forecasting inland fish production of India than the , other above said fuzzy and Neural Network methods studied.



**Fig. 2.** Comparison of average error of inland fish production forecasted values



### Acknowledgement

The authors are grateful to the Director CIFE, Mumbai for encouragement and providing the facilities under the Institutional project (Project code: CIFE-2011/9)

### CONCLUSIONS

The motivation of using different fuzzy time series model is that the historical inland fish production data are collected through various sampling techniques involving the vagueness. Further the production forecast has also been obtained through ANN (Artificial Neural Network) using Back propagation algorithm (BPA). A comparison of forecasted productions obtained through these models has been compared in table-4. It is also observed that ANN produces less accurate results in comparison of fuzzy time series methods given by Chen and Hsu [6]. It is found that when the dependence is linear (in case of inland fish production), the performance of ANN model is similar or less performer as comparison to fuzzy time series methods. In such cases, uses of ANN only lead to increase in complexity without significant improvement in the performance.

### REFERENCES

- [1] Zadeh L A, Fuzzy sets. *Information and Control*, 8 (1965) 338-353.  
 [2] Song Q & Chissom B S, Fuzzy time series and its models. *Fuzzy Sets and Systems*, 54 (1993).

- [3] Chen S M, Forecasting enrollments based on fuzzy time series, *Fuzzy Sets and Systems*, 81 (1996) 311-319.  
 [4] Sah M. and Degitiarev Y. (2005). Forecasting Enrollment Model Based on First –Order Fuzzy Time Series. *World Academy of Science, Engineering and Technology*  
 [5] Singh S R, A Simple method of forecasting based on fuzzy time series, *Applied mathematics and computation*, 186 (2007) 330-339.  
 [6] Chen S M & Hsu C C, A new method to forecasting production using fuzzy time series. *International*  
 [7] Song Q & Chissom B S, Forecasting enrollments with fuzzy time series - Part II. *Fuzzy Sets and Systems*, 62 (1994) 1-8.  
 [8] Song Q, A note on fuzzy time series model selection with sample autocorrelation functions. *Cybernetics and Systems: An International Journal*, 34 (1996) 93-107.  
 [9] Chen S M & Hwang J R ,Temperature prediction using fuzzy time series, *IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics*, 30 (2000) 263-275.  
 [10] Hwang J R, Chen S M & Lee C H, Handling forecasting problems using fuzzy time series. *Fuzzy Sets and Systems*, 100 (1998) 217-228.  
 [11] Huarng K, Heuristic models of fuzzy time series for forecasting. *Fuzzy Sets and Systems*, 123 (2001) 369-386.  
 [12] Chen S M, Forecasting enrollments based on high-order fuzzy time series, *Cybernetics and Systems: An International Journal*, 33 (2002) 1-16. ng LST images obtained from the ANN model are presented in the Figures 6 to 9.

**Table 1- A comparison of the forecasting results of different fuzzy time series forecasting methods**

Year	Actual production (Lakh Kg)	Forecasted Production by Singh(2007)	Forecasted Production by Chen (1996)	Forecasted Production by Chen and Hsu(2004)	Forecasted Production By Sah & Degitiarev
1995-96	22420				
1996-97	23810		25000		
1997-98	24380		25000	24750	24360
1998-99	26020	25907	25000	26500	27030
1999-2000	28230	28830	29500	28500	27503
2000-2001	28450	28774	31000	28500	29713
2001-2002	31200	32465	31000	30667	29000
2002-2003	32100	32632	34000	32000	31750
2003-2004	34580	35650	34000	34500	32650
2004-2005	35200	35238	37000	35500	35130
2005-2006	37550	38160	37000	37500	35750
2006-2007	38450	38259	40000	39000	38100
2007-2008	42070	41425	40000	43000	39000
2008-2009	46000	44240	41500	46667	46120
2009-2010	48700	46533	44500	49334	48650
Mean Square Error		1010453	5105175	226086.2	2053367
Average Error (%)		2.04	4.962999	1.05	3.18

**Table. 2** Forecasted production along with actual production by Neural Network (Method I)

Year	Actual production (Lakh Kg)	Actual production after normalization*	(Fuzzy output by Chen and Hsu (2004) after Normalization)	Neural Network output by method I	Neural Network Output after de-normalization** (by Method I)
1995-96	22420	0			
1996-97	23810	0.05	0		
1997-98	24380	0.07	0.04		
1998-99	26020	0.13	0.11		
1999-2000	28230	0.22	0.19		
2000-2001	28450	0.23	0.19		
2001-2002	31200	0.33	0.277		
2002-2003	32100	0.36	0.33		
2003-2004	34580	0.46	0.42		
2004-2005	35200	0.48	0.46		
2005-2006	37550	0.57	0.54	0.55	37709
2006-2007	38450	0.6	0.6	0.66	40550
2007-2008	42070	0.74	0.75	0.74	42617
2008-2009	46000	0.89	0.9	0.81	44426
2009-2010	48700	1	1	0.87	45976

\* Normalized new value= (original value-Input minimum value)/ (Input maximum value-Input minimum value)

\*\*De-Normalized value= Normalized value (Output maximum Value-Output Minimum value) +Output Minimum value

**Table. 3** Forecasted production along with actual production by Neural Network (Method II)

Year	Input(Fuzzy value obtained by Chen and Hsu (2004))	Input value after Normalization	Desired output (Actual Production after Normalization)	Neural Network output by method II	Neural Network Output after de-normalization(by Method II)
1995-96	23500				
1996-97	24750	0			
1997-98	26500	0.04			
1998-99	28500	0.1			
1999-2000	28500	0.18			
2000-2001	30667	0.18			
2001-2002	32000	0.262			
2002-2003	34500	0.30			
2003-2004	35500	0.4			
2004-2005	37500	0.43			
2005-2006	39000	0.50	0.57	.61	38451
2006-2007	43000	0.56	0.6	0.62	38714
2007-2008	46667	0.70	0.74	0.73	41604
2008-2009	49334	0.84	0.89	0.82	43970
2009-2010	23500	0.93	1	0.87	45284

**Table. 4** A comparisons of the forecasting results of different fuzzy time series & Neural Network forecasting methods

Year	Actual production (Lakh Kg)	Forecasted Production by Singh(2007)	Forecasted Production by Chen (1996)	Forecasted Production by Chen and Hsu(2004)	Forecasted Production By Sah & Degtiarev	Forecasted Production By Neural Network(BPA) method I	Forecasted Production By Neural Network (BPA) method II
1995-96	22420						
1996-97	23810		25000				
1997-98	24380		25000	24750	24360		
1998-99	26020	25907	25000	26500	27030		
1999-2000	28230	28830	29500	28500	27503		
2000-2001	28450	28774	31000	28500	29713		
2001-2002	31200	32465	31000	30667	29000		
2002-2003	32100	32632	34000	32000	31750		
2003-2004	34580	35650	34000	34500	32650		
2004-2005	35200	35238	37000	35500	35130		
2005-2006	37550	38160	37000	37500	35750	37709	38451
2006-2007	38450	38259	40000	39000	38100	40550	38714
2007-2008	42070	41425	40000	43000	39000	42617	41604
2008-2009	46000	44240	41500	46667	46120	44426	43970
2009-2010	48700	46533	44500	49334	48650	45976	45284
Mean Square Error		1010453	5105175	226086.2	2053367	2926428.4	3377722
Average Error (%)		2.04	4.962999	1.05	3.18	3.24	3.12