Time series forecasting using a TSK fuzzy system tuned with simulated annealing

ALMARAASHI, Majid, JOHN, Robert, COUPLAND, Simon and HOPGOOD, Adrian

Published version

ALMARAASHI, Majid, JOHN, Robert, COUPLAND, Simon and HOPGOOD, Adrian (2010). Time series forecasting using a TSK fuzzy system tuned with simulated annealing. In: IEEE International Conference on Fuzzy Systems (FUZZ), 2010. IEEE xplore, 1-6.

Time Series Forecasting Using a TSK Fuzzy System tuned with Simulated Annealing

Majid Almaraashi, *Student Member, IEEE*, Robert John, *Senior Member, IEEE*, Simon Coupland, *Member, IEEE*, Adrian Hopgood

Abstract— **In this paper, a combination of a Takagi-Sugeno fuzzy system (TSK) and simulated annealing is used to predict well known time series by searching for the best configuration of the fuzzy system. Simulated annealing is used to optimise the parameters of the antecedent and the consequent parts of the fuzzy system rules. The results of the proposed method are encouraging indicating that simulated annealing and fuzzy logic are able to combine well in time series prediction.**

I. Introduction

One of the features of fuzzy systems is the ability to be hybridised with other methods such as neural networks, genetic algorithms and other search methods. These approaches have been proposed because of the lack of learning capabilities in fuzzy systems [1]. Fuzzy systems are good at explaining how they reached a decision but can not automatically acquire the rules or membership functions to make a decision [2, p.2]. On the other hand, learning methods such as neural networks can not explain how a decision was reached but have a good learning capability [2, p.2]. Hybridisation overcomes the limitations of each method in one approach such as neuro-fuzzy systems or genetic fuzzy systems.

Soft Computing is a branch of computer science described as "a collection of methodologies aim to exploit the tolerance for imprecision and uncertainty to achieve tractability, robustness and low solution cost" [3]. Among Soft Computing techniques combinations, we are interested in the combination of fuzzy logic with simulated annealing to design a high-level performance and low-cost system. When designing a simple fuzzy system with few inputs, the experts may be able to give their knowledge to provide efficient rules but as the complexity of the system grows, the optimal rule base and membership functions become difficult to acquire. So, researchers often use some automated tuning and learning methods and evaluate their solutions by some criterion [4]. Although optimisation search algorithms such as genetic algorithms and simulated annealing are not specifically designed for learning, they offer some advantages for machine learning as many machine learning methodologies are based on a search of a good model among a space of possible models such as the space of rule sets allowing these types of methodologies to model the learning process as a search problem [5]. From an optimisation perspective, the task of finding a good knowledge base (KB) for a problem is equivalent to the task of parametrising the fuzzy knowledge base (KB) and equivalent to the task of finding the parameters values that are optimal based on the criteria of the problem design [1].

Simulated annealing has been used in some fuzzy systems to learn or tune the fuzzy system. For example, simulated annealing has been applied to optimise fuzzy logic controllers and showed that the optimality of the solutions is proved in probability if the optimisation process is infinite and provides efficient solutions in finite cases [6]. Garibaldi and Ifeachor have applied simulated annealing to tune a medical fuzzy expert system and showed that the fuzzy expert system tuned by simulated annealing outperformed the fixed fuzzy expert system and the crisp expert system [7]. Also, they comment on their experience that simulated annealing can be applied easily to either discrete or continuous variables which allowing the tuning of the structure and the parameters of the model at the same time. Liu and Yang [4] presented a study of using simulated annealing for learning and tuning the membership functions showing the efficiency of this algorithm as providing promising results. Compared to genetic algorithms, the main strength of simulated annealing is its wide applicability [4]. In this paper, a forecasting method is proposed by combining (TSK) fuzzy system with simulated annealing algorithm to predict two time series, Mackey-Glass time series and Henon Time series. The rest of the paper starts by describing the time series data sets in section II followed by a review of fuzzy systems (section III) and simulated annealing (section IV). The methodology and the results of this paper are detailed in section V where the conclusion is drawn in section VI.

II. The Predicted Time Series

A. Mackey-Glass Time Series

Mackey-Glass Time Series is a chaotic time series proposed by Mackey and Glass [8]. It is obtained from this non-linear equation :

$$\frac{dx(t)}{dt} = \frac{a * x(t-\tau)}{1 + x^n(t-\tau)} - b * x(t)$$

Where a, b and n are constant real numbers and t is the current time where τ is the difference between the current time and the previous time $t - \tau$. To obtain the simulated data, the equation can be discretised using the Fourth-Order Runge-Kutta method. In the case where $\tau > 17$, it is known to exhibit chaos and has become one of the

The authors are with the Centre for Computational Intelligence, Faculty of Technology, De Montfort University, Leicester LE1 9BH, U.K. (e-mail: almaraashi@dmu.ac.uk).

benchmark problems in soft computing [9, p.116]. Mackey-Glass equation with $\tau = 30$ is considered as a good example of low-dimensional chaos [10].

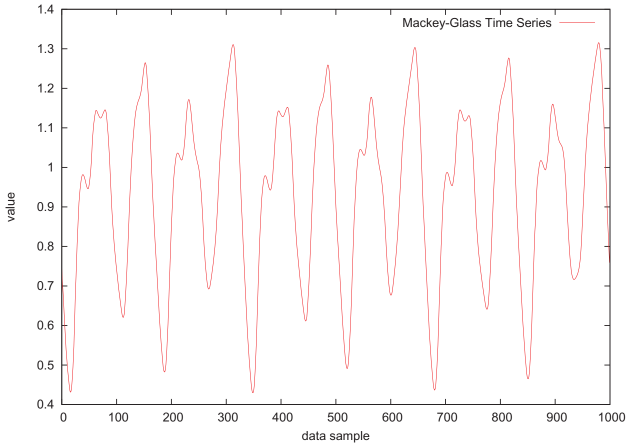

Fig. 1. Mackey-Glass time series when Tau=17

B. Henon Time Series

The Henon map proposed by [11] is a dynamical system that exhibits chaotic behaviour. It is a simple model that has the same essential properties of Lorenz system [12] of differential equations. Henon mapping is based on the following equations :

$$x(n+1) = y_n + 1 - a * x^2$$

$$y(n+1) = b * x_n$$

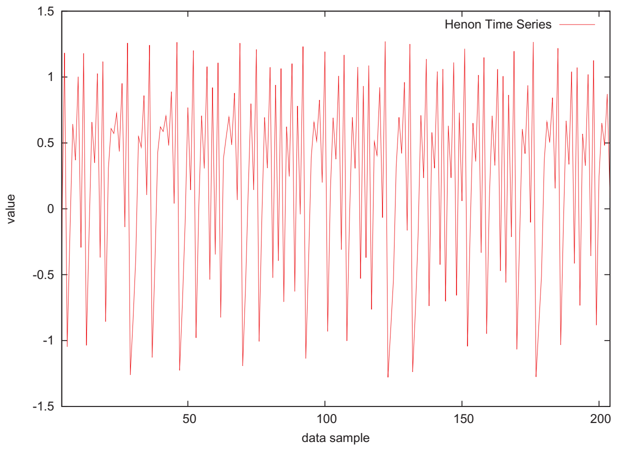

Fig. 2. Henon Time Series

III. Type-1 Fuzzy Systems

In this research, we are interested in the role of simulated annealing in fuzzy systems. A type-1 fuzzy system is a rule based system which can be viewed as a process that maps crisp inputs to crisp outputs by using the theory of fuzzy logic [13, p. 106]. The Mamdani fuzzy model contains four components : fuzzifier, rules, inference engine and output processor or defuzzifier [9, p. 6]. The TSK fuzzy model is different from Mamdani in the inference engine and the output processor. The fuzzifier maps crisp inputs to fuzzy sets by evaluating the crisp inputs $x = (x_1, x_2, \ldots, x_n)$ based on the antecedents part of the rules and assigns each crisp input to its fuzzy set $A(x)$ in X with its degree of membership in each fuzzy set. A fuzzy rule is a conditional statements in the form of IF-THEN where it contains two parts, the IF part called the antecedent part and the Then part called the consequent part. To acquire these rules, many methods can be used such as getting them from experts or using statistical data. The inference and defuzzification processes in Takagi and Sugeno model [14] are different from the known Mamdani model as the rules in TSK are based on input fuzzy sets in the antecedent part and a mathematical linear function in the consequent part. The ith rule in the first-order TSK is described as follows:

$$R^i : \text{ IF } x_1 \text{ is } A_1^i \text{ and } x_2 \text{ is } A_2^i \ldots \text{ and } x_p \text{ is } A_p^i$$

$$\text{THEN } y^i(x) = c_0^i + c_1^i * x_1 + c_2^i * x_2 \ldots + c_p^i * x_p$$

Where i represents the rule number and c_0^i, c_1^i, c_2^i and c_p^i are the the coefficients of the consequent part of the fuzzy system rules. The final control value Y is computed as follows:

$$Y = \frac{\sum_{i=1}^{n} \alpha_i * y_i}{\sum_{i=1}^{n} \alpha_i}$$

Where α_i is the firing level for the ith rule which is derived by using a t-norm operator such as minimum or product.

IV. Simulated Annealing Algorithm

The concept of annealing in the optimisation field was introduced by Kirkpatrick et al in 1982 [15]. This concept is based on the analogy between the annealing process in metallurgy where heating and controlled cooling of materials is used to recrystallize metals by increasing the temperature to the maximum values until the solids almost melted then decreasing the temperature carefully until the particles are arranged and the system energy becomes minimal. Simulated annealing is a powerful randomized local search algorithm that has shown great success in finding optimal or nearly optimal solutions of combinatorial problems [16]. We now define the simulated annealing algorithm. Let s be the current state and $N(s)$ be a neighbourhood of s that includes alternative states. By selecting one state $s' \in N(s)$ and computing the difference between the current state cost and the selected state energy as $D = f(s') - f(s)$, s' is chosen as the current state in two cases [17, p.124]

- If $D < 0$ then s' is chosen as the current state as downhills always accepted.
- If $D > 0$ and the probability of accepting s' is larger than a random value Rnd such that $e^{-D/T} > Rnd$ then s' is chosen as the current state where T is a control parameter known as $Temperature$ which is

gradually decreased during the search process making the probability of accepting uphill moves decreasing over time. Rnd is a randomly generated number where $0 < Rnd < 1$. In this case uphill moves might be accepted to avoid being stuck in a local minima.

In the last case where $D > 0$ and the probability is lower than the random value $e^{-d/T} <= Rnd$, no moves are accepted and the current state s continues to be the current solution. In the original proposed version of simulated annealing by Kirkpatrick, Gelatt and Vecchi, the probability of accepting s' equals 1 when $f(s') <= f(s)$. Simulated annealing can be implemented to find the optimal annealing by allowing infinite number of transitions or can be implemented to find a closest possible optimal value within a finite time. When starting with a large cooling parameter, large deteriorations are accepted. Then, as the temperature decreases, only small deteriorations are accepted until the temperature approaches zero when no deteriorations are accepted. Therefore, adequate temperature scheduling is important to optimise the search. In the case of finite-time implementation, the cooling schedule of finite values of temperature and a finite number of transitions of each value of temperature is specified by four components [16]:

1) Initial value of temperature.
2) A function to decrease temperature value gradually.
3) A final temperature value.
4) The length of each Homogeneous Markov chains which is a sequence of trials where the probability of the trial outcome depends on the previous trial outcome [18, p.33].

One of the methods used to determine the initial temperature value proposed by [19] is to choose the initial temperature value within the standard deviation of the mean cost. Markov chains are not used in this experiment.

V. Methodology and Results

The experiment can be divided into three steps : generating time series, constructing the initial fuzzy system and optimising the fuzzy system parameters. Firstly, the time series are generated. Two Mackey-Glass time series are initialised with the following parameters : $alpha = 0.2$, $Beta = 0.1$, $\tau = 17$ for one experiment and $\tau = 30$ for the second experiment and initial condition $x(0) = 1.2$. A Henon time series is generated with the following values : alpha=1.4 and beta = 0.3. From each time series, input-output samples are extracted in the form $x(t-18), x(t-12), x(t-6)$ and $x(t)$ where $t = 118$ to $t = 1117$ using a step size of 6. Then the generated data are divided into 500 data points for training and the remaining 500 data points for testing. Four initial input values $x(114)$ and $x(115)$ and $x(116)$ and $x(117)$ are used to predict the first four training outputs.

The fuzzy system is a four-inputs one-output first-order Takagi-Sugeno fuzzy model and consists of four input fuzzy sets $A_1, A_2 A_3$ and A_4. Gaussian membership functions were chosen to define the fuzzy sets. However, any other types of membership functions can be chosen. The training procedure aims to optimise the the parameters of the antecedent parts and the coefficients of the consequent parts of the fuzzy system rules. Then, the found parameters are used to predict the next 500 testing data points. For the sake of comparison of the performance with other methods, all settings of Mackey-Glass examples were chosen as close as possible to [20] when $tau = 17$ and to [21] and [22] when $tau = 30$. Using a step size of 6, the input values to the fuzzy system are the previous data points $x(t-18), x(t-12), x(t-12)$ and $x(t)$ while the output from the fuzzy system is the predicted value $x(t+6)$. The parameters of the Gaussian membership functions are the mean m and the standard deviation σ. All the means and standard deviations are initially chosen for all the input fuzzy sets as following [22] : $m = m_l - 2\sigma_l$ and $m = m_l + 2\sigma_l$ respectively where m_l and σ_l are the mean and standard deviation of all 500 inputs. The standard deviation of all input fuzzy sets is chosen as $\sigma = 2\sigma_l$. The initial fuzzy sets are depicted in figure 3. The fuzzification process is based on the minimum t-norm. By using four inputs and two fuzzy sets for each input, we end up with 16 rules and 8 input fuzzy sets representing all possible combinations of input values with input fuzzy sets. While each rule is linked with 5 coefficients c_0, c_1, c_2, c_3, c_4, there are 8 means and 8 standard deviations are linked with all these rules. The total number of optimised parameters is $8 + 8 + (5 * 16) = 96$. The objective is to find the best set of parameters for all the rules.

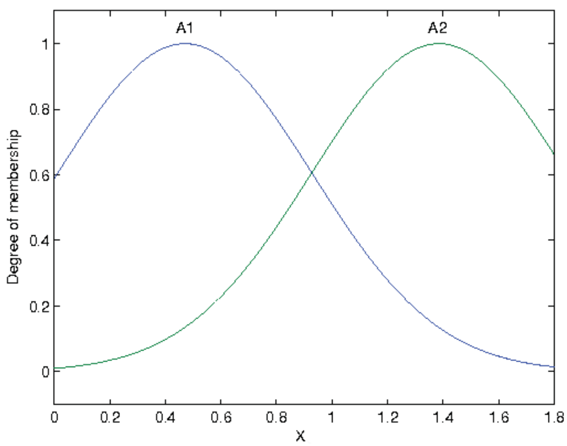

Fig. 3. The initial state for all input fuzzy sets for Mackey-Glass time series when tau=17

The optimisation process is done using simulated annealing that searches for the best configuration of the parameters by trying to modify one parameter each time and evaluate the cost of the new state which is measured by Root Mean Square Error (RMSE). The simulated algorithm is initialised with a temperature that equals to the standard deviation of mean of RMSE's for 100 runs for the 500 training points. The cooling schedule is based on a cooling rate of 0.99990 which allows a slow cooling process. The neighbouring states for a current state are chosen randomly by adding a small

TABLE I

THE FORECASTING RESULTS FOR MACKEY-GLASS TIME SERIES BY TWO TAU VALUES

Experiment	Tau	RMSE
Training Results	17	0.00364
Training Results	30	0.00417
Testing Results	17	0.00366
Testing Results	30	0.00649

TABLE II

THE FORECASTING RESULTS FOR HENON TIME SERIES

Experiment	RMSE
Training Results	0.01308
Testing Results	0.01465

number between -0.1 and 0.1 to one of the 80 coefficients in the current state or adding a small number between -0.02 and 0.02 to one of the 16 antecedent parameters. Then, the new state is evaluated by examining the 500 data points outputs. The proposed method shows a small RMSE for the testing data in Mackey-Glass time series which is 0.00366 for $\tau = 17$ and 0.00649 for $\tau = 30$ as shown in Table I. For Henon time series, the RMSE was 0.01465 for the testing data as shown in Table II. The results of predicting of Mackey-Glass data are depicted in figures 5 and 6 while the results of predicting Henon time series are depicted in figure 7. Figures 8, 9 and 10 show the the absolute prediction error for results. To show the final state of the membership functions, Figure 4 shows the final two input fuzzy sets after optimisation for the first input only to predict Mackey-Glass time series when tau=17. Other fuzzy sets for each input have different values for their means and standard deviations.

For comparison purpose, The results obtained by other methods for Mackey-Glass time series with $\tau = 17$ taken from [20] are shown in Table III. These results are measured by non-dimensional error index (NDEI) which is defined as the root mean square error divided by the standard deviation of the target time series [20] which is in our experiment equals to 0.016. We see that our result of (NDEI)= 0.016 is the closest result to the best result which was obtained by ANFIS. When $\tau = 30$ the result is the best among the two methods shown in Table V while the prediction of Henon time series is quite accurate and better than the LS-SVM with GA method proposed by [23] which is RMSE=0.024.

TABLE III

RESULTS COMPARISON FOR PREDICTING MACKEY-GLASS TIME SERIES WHEN TAU=17

Method	None-Dimensional Error Index (NDEI)
ANFIS	0.007
Auto Regressive Model	0.19
Cascaded-Correlation NN	0.06
Back-Propagation NN	0.02
Sixth-Order Polynomial	0.04
This Model	0.016

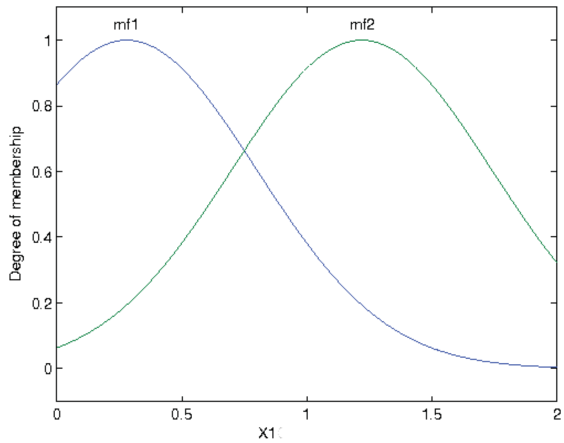

Fig. 4. The final two input fuzzy sets after optimisation for the first input only to predict Mackey-Glass time series when tau=17

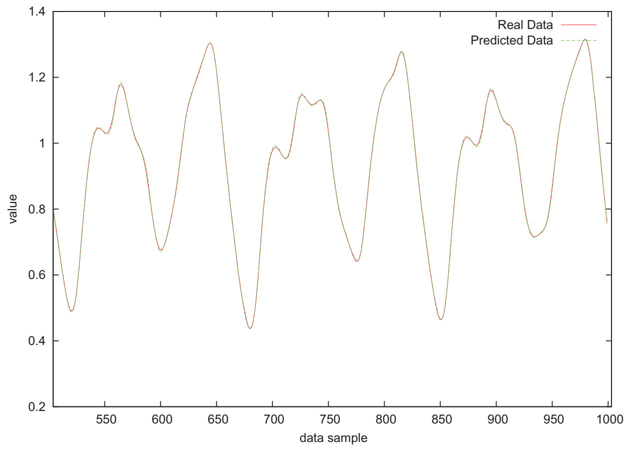

Fig. 5. The actual and predicted data of Mackey-Glass time series when tau=17 for the training and testing data showing the precision of the method where the real data and predicted data are indistinguishable

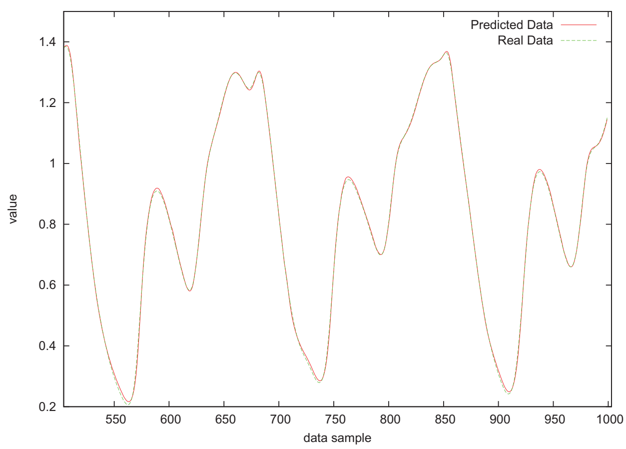

Fig. 6. The actual and predicted data of Mackey-Glass time series when tau=30 for the training and testing data showing the precision of the method

TABLE IV

RESULT COMPARISONS FOR PREDICTING MACKEY-GLASS TIME SERIES WHEN TAU=30

Method	RMSE
Back-Propagation with fuzzy system [22]	0.02
One-Pass Method with fuzzy system [22]	0.04
Gradient Descent with fuzzy system [21]	0.21
This Model	0.00649

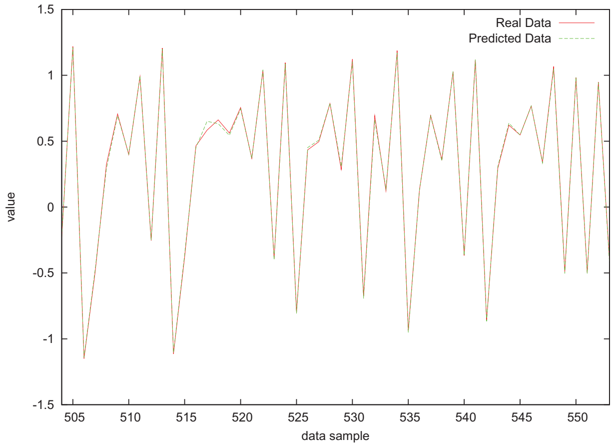

Fig. 7. The actual and predicted data of Henon time series for a sample of 50 testing data showing the precision of the method

VI. CONCLUSION

Simulated annealing is used to optimise Takagi-Sugeno fuzzy system by searching for the best parameters of the antecedent and the consequent parts of the fuzzy system. Two time series have been predicted by this method which are Mackey-Glass and Henon time series. The combination exhibited good performance. We are planning to extend the experiment to introduce more uncertainty in other data set mirroring much of the data in the real world. We believe that type-2 fuzzy logic will produce good performance as type-2 fuzzy sets are known to handle highly uncertain data well [22] and we will compare the performance of the approach between type-1 fuzzy systems and type-2 fuzzy systems.

REFERENCES

[1] O. Cordón, F. Gomide, F. Herrera, F. Hoffmann, and L. Magdalena, "Ten years of genetic fuzzy systems: current framework and new trends," *Fuzzy Sets and Systems*, vol. 141, no. 1, pp. 5–32, 2004.

[2] S. Goonatilake and S. Khebbal, *Intelligent Hybrid Systems*. John Wiley & Sons, 1995.

[3] L. Zadeh, "Soft computing and fuzzy logic," *IEEE Software*, vol. 11, no. 6, pp. 48–56, 1994.

[4] G. Liu and W. Yang, "Learning and tuning of fuzzy membership functions by simulated annealing algorithm," in *Circuits and Systems, 2000. IEEE APCCAS 2000. The 2000 IEEE Asia-Pacific Conference on*, 2000, pp. 367–370.

[5] F. Herrera, "Genetic fuzzy systems: Status, critical considerations and future directions," *International Journal of Computational Intelligence Research*, vol. 1, no. 1-2, pp. 59–67, 2005.

[6] E. Huyghe and Y. Hamam, "Simulated annealing for fuzzy controller optimization: Principles and applications," in *IEEE International Conference On Systems and Cybernetics*, vol. 5. Institute of Electrical Engineers INC (IEEE), 1995, pp. 4509–4514.

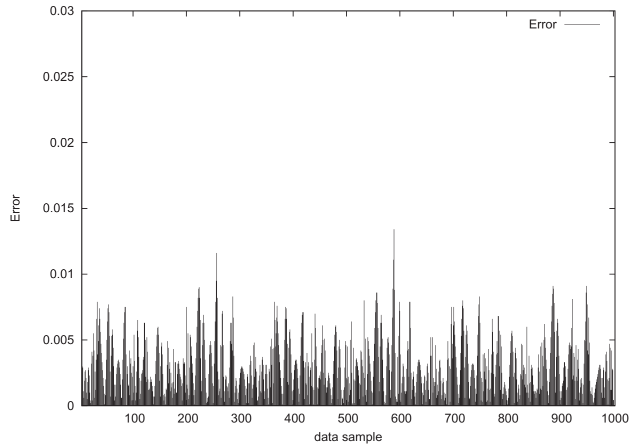

Fig. 8. The absolute prediction error for Mackey-Glass time series for the training and testing data points when tau=17

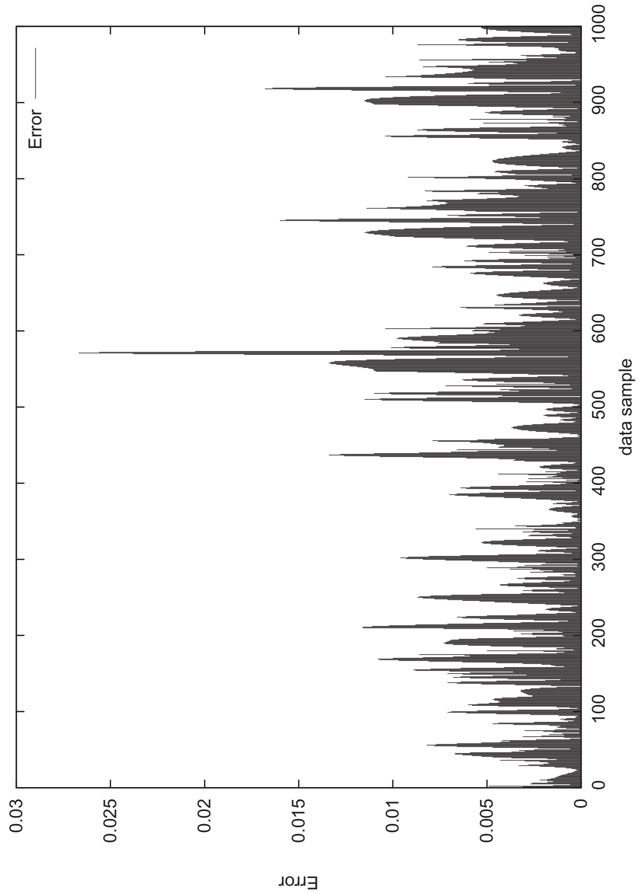

Fig. 9. The absolute prediction error for Mackey-Glass time series for the training and testing data points when tau=30

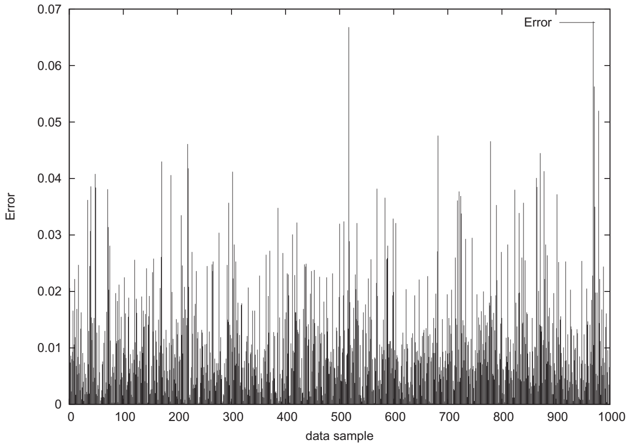

Fig. 10. The absolute prediction error for Henon time series for the training and testing data points

[7] J. Garibaldi and E. Ifeachor, “Application of simulated annealing fuzzy model tuning to umbilical cord acid-base interpretation,” *IEEE Transactions on Fuzzy Systems*, vol. 7, no. 1, pp. 72–84, 1999.

[8] M. Mackey and L. Glass, “Oscillation and chaos in physiological control systems,” *Science*, vol. 197, no. 4300, pp. 287–289, 1977.

[9] J. Mendel, *Uncertain rule-based fuzzy logic systems: introduction and new directions*. Prentice Hall, 2001.

[10] W. Mead, R. Jones, Y. Lee, C. Barnes, G. Flake, L. Lee, and M. O’rourke, “Prediction of chaotic time series using CNLS-net-example: The Mackey-Glass equation,” in *SFI STUDIES IN THE SCIENCES OF COMPLEXITY-PROCEEDINGS VOLUME-*, vol. 12. ADDISON-WESLEY PUBLISHING CO, 1992, pp. 39–39.

[11] M. Henon, “A two-dimensional mapping with a strange attractor,” *Communications in Mathematical Physics*, vol. 50, no. 1, pp. 69–77, 1976.

[12] E. Lorenz, “Deterministic Nonperiodic Flow.” *Journal of Atmospheric Sciences*, vol. 20, no. 2, pp. 130–148, 1963.

[13] M. Negnevitsky, *Artificial intelligence: a guide to intelligent systems*. Addison-Wesley, 2002.

[14] T. Takagi and M. Sugeno, “Fuzzy identification of systems and its applications to modeling and control,” *IEEE transactions on systems, man, and cybernetics*, vol. 15, no. 1, pp. 116–132, 1985.

[15] S. Kirkpatrick, C. Gelatt, and M. Vecchi, “Optimization by simulated annealing, 1983,” *Science*, vol. 220, pp. 671–680, 1983.

[16] E. H. L. Aarts and H. M. M. T. Eikelder, *Handbook of applied optimization*. Oxford University Press, 2002, ch. Simulated Annealing, pp. 209–220.

[17] B. L. Golden and E. A. Wasil, *Handbook of applied optimization*. Oxford University Press, 2002, ch. Metaheuristic, an Introduction, p. 123.

[18] E. Aarts and J. Korst, *Simulated annealing and Boltzmann machines: a stochastic approach to combinatorial optimization and neural computing*. John Wiley & Sons Inc, 1989.

[19] S. White, “Concepts of scale in simulated annealing,” in *American Institute of Physics Conference Series*, vol. 122, 1984, pp. 261–270.

[20] J. Jang, “ANFIS: Adaptive-network-based fuzzy inference systems,” 1993.

[21] S. Musikasuwan, T. Ozen, and J. M. Garibaldi, “An investigation into the effect of number of model parameters on performance in type-1 and type-2 fuzzy logic systems,” in *In Proc. 10 th Information Processing and Management of Uncertainty in Knowledge Based Systems (IPMU*, 2004.

[22] J. Mendel and R. John, “Type-2 fuzzy sets made simple,” *IEEE Transactions on Fuzzy Systems*, vol. 10, no. 2, pp. 117–127, 2002.

[23] X. Wang, H. Zhang, C. Zhang, X. Cai, J. Wang, and J. Wang, “Prediction of chaotic time series using ls-svm with automatic parameter selection,” dec. 2005, pp. 962 – 965.