Proceedings of the 2007 IEEE Symposium on
Computational Intelligence and Data Mining (CIDM 2007)

Fuzzy Wavelet Modeling Using Data Clustering

Nasser Sadati, *Member, IEEE* and Bahram Marami

***Abstract*— In this paper, a novel approach for tuning the parameters of fuzzy wavelet systems which are used for modeling of nonlinear and complex systems is proposed. In fuzzy inference system, each fuzzy rule is analogous to a wavelet basis function multiplied by a coefficient. Using clustering techniques, the center of these basis functions are located in the detected center of clusters. In this way, not only the approximation accuracy is increased, but also the number of unknown parameters is decreased. The feasibility of the proposed method is shown by modeling two highly nonlinear functions. The comparison of the results using the proposed approach, with the previous schemes, shows the effectiveness and superiority of this algorithm.**

I. INTRODUCTION

FUZZY systems, due to their universal approximation property, constitute a good framework for modeling complex and nonlinear systems. However, fuzzy systems are difficult to be constructed by specialists' knowledge or experience when complexity and desirable accuracy of identified or controlled systems are increased, and they require formal synthesis techniques that guarantee the global stability and acceptable performance [1], [2]. Many researchers have done plenty of study in combining the fuzzy systems with other intelligent systems or beneficial classic theories to construct accurate inference systems. One of these theories is wavelet transform that in recent years has become a very active subject in many scientific and engineering research areas. Wavelet transform decomposes a signal into a set of special basis wavelet functions that give a time-frequency localization of the signal. In [3], [4], it is shown that the fuzzy models with some restrictions and modifications can be functionally equivalent to discrete wavelet transforms. Therefore, the fuzzy models can not only offer a framework for combining linguistic information and numerical data in a unified fashion, but also take advantage of the rigorous approximation theory of wavelet basis function expressions. In order to benefit from the discrete wavelet transform (DWT), the general fuzzy models, such as Takagi-Sugeno's model can be modified such that each fuzzy rule can be viewed as a wavelet basis function multiplied by a real coefficient. But major parameters such as the dilation and translation parameters of these basis functions should also be tuned in order to construct the fuzzy system which can approximate a given system accurately. In some studies [5], [6] these parameters, including the multiplied real coefficients, are tuned by particular learning algorithms using checking data. But because of dilation and translation parameters, which are located nonlinearly in the approximation function, these tuning algorithms may not result in optimal solutions. Therefore in this paper, data clustering algorithms are used for setting the translation parameters. So by selecting the appropriate dilation parameters, only multiplied real coefficients remain which they could also be tuned by tuning algorithms. Hence only linear parameters are tuned. Therefore, as is shown, besides increasing the accuracy, computational volume is reduced effectively.

The authors are with the Department of Electrical Engineering, Intelligent Systems Laboratory, Sharif University of Technology, Tehran, Iran (e-mail: sadati@sina.sharif.edu , marami@ee.sharif.edu).

II. CLUSTERING ALGORITHM

The objective of the clustering analysis is the classification of objects according to their similarities, and organizing the data into several groups [7]. Clustering techniques which can be classified into unsupervised methods, do not use prior class identifiers. The main potential of clustering is to detect the underlying structure in data, not only for classification and pattern recognition, but for model reduction and optimization. In this paper, the Gustafson-Kessel [8] algorithm is used for data clustering. The algorithm is the extended form of standard fuzzy c-means which employs an adaptive distance norm, in order to detect the clusters of different geometrical shapes in one data set. In this way, each cluster has its own norm-inducing matrix A_i, which yields the following inner-product norm

$$D_{ikA}^2 = (x_k - v_i)^T A_i (x_k - v_i); \quad 1 \le i \le c, \quad 1 \le k \le N. \tag{1}$$

where $V = [v_1, v_2, \ldots, v_c], v_i \in R^n$ is a vector of cluster centers, which have to be determined. The clustering algorithm is based on the minimization of an objective function defined as

$$J(X;U,V,A) = \sum_{i=1}^{c}\sum_{k=1}^{N}(\mu_{ik})^m D_{ikA_i}^2 \tag{2}$$

where $U = [\mu_{ik}]$ represent the fuzzy partitions and its conditions given by

$$\mu_{ij} \in [0,1], 1 \le i \le N, 1 \le k \le c, \tag{3}$$

$$\sum_{k=1}^{c} \mu_{ik} = 1, 1 \le i \le N, \tag{4}$$

$$0 < \sum_{i=1}^{N} \mu_{ik} < N, 1 \le k \le c \tag{5}$$

Also m is the weighting exponent that determines the fuzziness of the clusters and must be defined as a scalar greater than one. The objective function is minimized using the Lagrange multiplier method having the above constraints and allowing the matrix A_i to vary with its determinant fixed corresponding to optimizing the cluster's shape, while its volume remains constant: $\|A_i\| = \rho_i; \quad \rho_i > 0,$ where ρ_i is fixed for each cluster. Therefore, the algorithm is implemented according to the following procedure:

Given the data set X, choose the number of clusters $1 \le c \le N$, the weighting exponent $m > 1$, the termination tolerance $\varepsilon > 0$ and the norm-inducing matrix A. Also initialize the partition matrix randomly, such that $U^{(0)} \in M_{fc}$.

Repeat for $l = 1, 2, \ldots$

Step1: Calculate the cluster centers

$$v_i^{(l)} = \frac{\sum_{k=1}^{N} (\mu_{ik}^{(l-1)})^m x_k}{\sum_{k=1}^{N} (\mu_{ik}^{(l-1)})^m}; \quad 1 \le i \le c \tag{6}$$

Step2: Compute the cluster covariance matrices

$$F_i^{(l)} = \frac{\sum_{k=1}^{N} (\mu_{ik}^{(l-1)})^m (x_k - v_i^{(l)})(x_k - v_i^{(l)})^T}{\sum_{k=1}^{N} (\mu_{ik}^{(l-1)})^m}; \quad 1 \le i \le c \tag{7}$$

Add a scaled identity matrix

$$F_i := (1-\gamma) F_i + \gamma (F_0)^{1/n} I; \quad 0 \le \gamma \le 1 \tag{8}$$

Extract eigenvalues λ_{ij} and eigenvectors ϕ_{ij}. Find $\lambda_{i,\max} = \max_j \lambda_{ij}$, and set

$$\lambda_{i,\max} = \lambda_{ij} / \beta; \forall j \text{ for which } \lambda_{i,\max} / \lambda_{ij} \ge \beta \tag{9}$$

Reconstruct F_i by

$$F_i = \left[\phi_{i,1} \ \ldots \ \phi_{i,n}\right] diag(\lambda_{i,1} \ \ldots \ \lambda_{i,1}) \left[\phi_{i,1} \ \ldots \ \phi_{i,n}\right]^{-1} \tag{10}$$

Step3: Compute the distances

$$D_{ikA_i}^2(x_k, v_i) = (x_k - v_i^{(l)})^T \left[(\rho_i \det(F_i))^{1/n} F_i^{-1}\right](x_k - v_i^{(l)}) \tag{11}$$

Step4: Update the partition matrix

$$\mu_{ik}^{(l)} = \frac{1}{\sum_{j=1}^{c} \left(D_{ikA_i}(x_k, v_i) / D_{jk}(x_k, v_i)\right)^{2/(m-1)}} \tag{12}$$

$$; \quad 1 \le i \le c, 1 \le k \le N$$

until $\|U^{(l)} - U^{(l-1)}\| \le \varepsilon$.

Now use the center of clusters; $V = [v_1, v_2, \ldots, v_c]$ to construct the fuzzy system described in next section.

$$D_{ikA}^2 = (x_k - v_i)^T A_i (x_k - v_i); \quad 1 \le i \le c, \quad 1 \le k \le N$$

III. Discrete Wavelet Transform

Wavelet system is a set of building blocks to construct or represent any function $f(x)$ that could be written as

$$f(x) = \sum_{k=-\infty}^{\infty} c_k a^{j_0/2} \varphi(a^{j_0} x - k) + \sum_{k=-\infty}^{\infty} \sum_{j=j_0}^{\infty} d_{jk} a^{j_0/2} \psi(a^{j_0} x - k) \tag{13}$$

where it is a series expansion in terms of the scaling function $\varphi_k(x)$ and wavelets $\psi_{kj}(x)$ [9]. In this expansion, c_k's are the coefficients that are referred to as approximation coefficient at scale j_0 and it could also be zero. The set of d_{jk} coefficients represents the details of the signal $f(x)$ at different scales. The DWT coefficients consist of both c_k's and d_{jk}'s. For a given wavelet $\psi \in L^2(R^n)$, the sequence function $\{\psi_{kj}\}$ is defined by dilating and translating mother wavelet function ψ as follows

$$\psi_{jk}(x) = \det D_j^{1/2} \psi(D_j x - \Lambda_k k) \tag{14}$$

where $j = [j_1, \ldots, j_n]^T \in Z^n, k \in Z^n$, $D_j = diag(a^{j_1}, \ldots, a^{j_n})$ is the dilation matrix and $\Lambda_k = diag(b_1, \ldots, b_n); a > 1, a \in R$ $b = (b_1, \ldots, b_n) \in R^n$ is the translation matrix. The conditions on ψ, a and b to guarantee a multi-scaling wavelet frame for $L^2(R^n)$, are obtained and given by [10].

Now let the multidimensional wavelet functions be the generalized one-dimensional wavelet functions, i.e.,

$$\psi(x) = \psi_1(x_1) \ldots \psi_n(x_n) \tag{15}$$

That is applying one-dimensional wavelet transform separately in each of n orthogonal direction. As is given in [10], for satisfying the sufficient conditions, a given function $\psi_i(x_i)$ with adequate decay at infinite, should have zero mean, i.e. ($\int \psi_i(x_i)\, dx_i = 0$).

In this paper, the sequence functions $\{\psi_{kj}\}$ are a set of wavelet basis functions to constitute a frame for a class of functions $f \in L^2(R^n)$, which are going to be approximated by $f = \sum_{j,k} c_{jk} \tilde{\psi}_{jk}$, where $\tilde{\psi}_{jk} = A^{-1} \psi_{jk}$ and $A\|f\|^2 \le \sum_{j,k} |< f, \psi_{jk} >|^2 \le B\|f\|^2$.

Also the coefficients are expressed as $c_{j,k} = < f, \tilde{\psi}_{jk} >$. Therefore, f can be reconstructed by the expansion of $\{\psi_{kj}\}$ exactly. Now a variety of wavelets can be used for our purpose. Here, the mother wavelet of the form $\psi_i(x_i) = g(x_i) e^{-\alpha_i x_i^2/2}$ is used, where the $g(x_i) = \alpha_i (1 - \alpha_i x_i^2)$ satisfies $\int \psi_i(x_i) dx_i = 0$. $\psi_i(x_i)$ is the so called 'Mexican Hat' mother wavelet function. Therefore, the multidimensional wavelet function can be constructed as

$$\psi_{jk} = a^{-j_1/2} \alpha_1 \left[1 - \alpha_1 (a^{-j_1} x_1 - b_1 k_1)^2\right] e^{-\alpha_1 (a^{-j_1} x_1 - b_1 k_1)^2/2} \ldots \tag{16}$$
$$a^{-j_n/2} \alpha_n \left[1 - \alpha_n (a^{-j_n} x_n - b_n k_n)^2\right] e^{-\alpha_n (a^{-j_n} x_n - b_n k_n)^2/2}$$

IV. Takagi–Sugeno Fuzzy System Using Discrete Wavelet Transform

The fuzzy model proposed by Takagi and Sugeno is suitable to model a large class of nonlinear systems and is described by fuzzy IF-THEN rules as follows

R_j : IF x_1 is A_{j1} and … and x_n is A_{jn}

$$\text{THEN } y = w_{j0} + w_{j1}x_1 + ... + w_{jn}x_n \tag{17}$$

where R_j represents the jth fuzzy inference rule, $A_{j1}, A_{j2}, \dots$ and A_{jn} are the input fuzzy sets, y is the output and $x = [x_1, ..., x_n]^T$ is the input state vector. The THEN part of this model is a linear combination of premise variables to represent the local linear input-output relation of nonlinear system. A fuzzy model, as proposed in [10], has the following form

R_c : IF x_1 is $\tilde{A}_{c1}$ and … and x_n is $\tilde{A}_{cn}$

$$\text{THEN } y_i = d_c g_{c1}(x_1)...g_{cn}(x_n) \tag{18}$$

where $\tilde{A}_{c1}, \dots, \tilde{A}_{cn}$ are the fuzzy sets with Gaussian membership functions, d_c's are the real coefficients, and $g_{ci}(x_i)$ is the shifted and scaled functions of $g(x_i)$. Now using the min-product inference process, the output y_i can be obtained as

$$y_i = \frac{\sum_c \left[d_c \prod_{i=1}^{n} \mu_{\tilde{A}_{ci}}(x_i) g_{ci}(x_i) \right]}{\sum_c \prod_{i=1}^{n} \mu_{\tilde{A}_{ci}}(x_i)} \tag{19}$$

where $\mu_{\tilde{A}_{ci}}(x_i)$ is the membership function of $\tilde{A}_{ci}$. As mentioned in [3], [5], the main purpose of proposed modification is to link the fuzzy model and the wavelet theory such that the fuzzy model can share the advantages of wavelet transform. Beside the modification of THEN part of fuzzy rules, multi-resolution wavelet function is needed for obtaining the accurate approximation which is considered by more than one fuzzy rule base. In [5], [6] any function f is approximated by $f = \sum_{j,k} c_{jk}\tilde{\psi}_{jk}$, where all unknown parameters such as c_{jk}, α, a and b, for any resolution j, are tuned using checking data. This is done by various procedures such as Kalman filter algorithm, least square (LS) and back-propagation law. Accordingly, a great set of parameters should be tuned which not only a large amount of computations is needed, but also for multiplicity of linear and nonlinear parameters, desired accuracy can not be obtained. Therefore, in this paper we use the data clustering techniques to specify the translation parameters. For a given resolution $j = [j_1, ..., j_n]^T$, we assume that the sequence of basis vectors is denoted as $\{\vec{e}_1, ..., \vec{e}_2\}$, and the center of clusters is represented by

$$v_k = k_1 a^{j_1} b_1 \vec{e}_1 + ... + k_n a^{j_n} b_n \vec{e}_n \tag{20}$$

Now for constructing the fuzzy system, the centers of membership functions can be located on the center of clusters. This procedure is depicted in Fig.1, for a two dimensional input data with two clusters. It should be noted

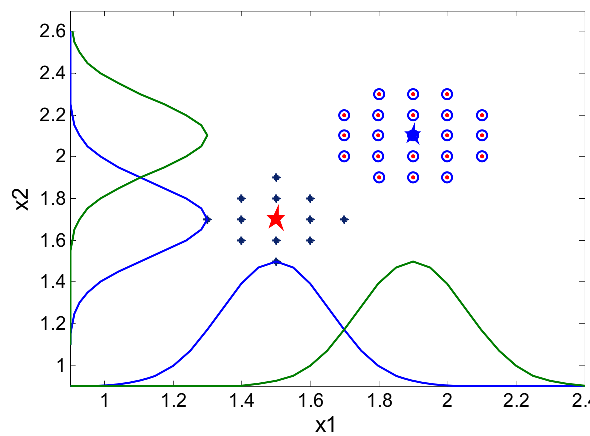

Fig. 1. Membership functions located on the center of clusters.

that in [3], [4], it is assumed that the center of membership function of linguistic terms are equally spaced and the center of fuzzy basis functions are located in these lattice points. Thus in the proposed approach presented in here, each cluster's center is a fuzzy basis function, ϕ_{jk}, defined as

$$\phi_{jk} = \frac{e^{-(x-v_k)^T Q_j (x-v_k)}}{\sum_{K \in Z^n} e^{-(x-v_k)^T Q_j (x-v_k)}} \tag{21}$$

where $Q_j = diag\left\{\alpha_1/(a^{j_1})^2, ..., \alpha_n/(a^{j_n})^2\right\}$. Hence, the output of each fuzzy rule base can be given by

$$y_j = \sum_k d_{jk}\psi_{jk} \tag{22}$$

where $\psi_{jk} = g_{jk1}(x_1) \dots g_{jkn}(x_n)\phi_{jk}$ and d_{jk}'s are the real coefficients. Also y_j represents the output of the jth fuzzy rule base. Therefore, the multi-resolution approximation of a given function f needs more than one fuzzy rule base. So it can be written as

$$f = \sum_j \sum_k d_{jk}\psi_{jk} \tag{23}$$

This expression represents the equivalent functional behavior of the multi-resolution wavelet transform and the proposed fuzzy model.

Now based on (22), for approximation of a given function f, using the proposed formulation, we only need to tune the real coefficients, i.e., d_{jk}'s. The center of membership functions which are set on the center of clusters are also obtained by Gustafson-Kessel data clustering algorithm. In each resolution, α and a are also set appropriately for obtaining the accurate approximation. Therefore, the number of unknown parameters depends on the number of clusters and also the number of used resolutions. Now for more accuracy, the final equation is added to another unknown parameter which adds a dc value to function evaluation generated by fuzzy wavelet. Hence, the approximation function can be written as

$$f = \sum_j \sum_k d_{jk}\psi_{jk} + d_0 \tag{24}$$

It should be noted that the tuning of linear coefficients, d_{jk}, using the training data, is almost simple and there exist some methods such as least mean square (LMS), Kalman filtering algorithm, back-propagation law, etc., which do the job for us. In this study, the least square (LS) algorithm is used to fit the evaluated function by the training data. That is, given input data $Xdata$, and the observed output $Ydata$, the coefficients d_{jk} is found such that the following equation is minimized.

$$\min_{d_{jk}} \frac{1}{2}\sum_{i=1}^{n}(F(d_{jk}, Xdata) - Ydata_i)^2 \tag{25}$$

$Xdata$ and $Ydata$ are the vectors of length n (length of training data), and $F(d_{jk}, Xdata)$ is a vector-valued function.

V. Numerical Experiments

In this section, the effectiveness of the proposed fuzzy wavelet modeling algorithm is demonstrated by some examples. The results for the proposed approach are obtained by using a model which gives better performance index and less number of parameters than others. If the performance of the selected model is not adequate, then we increase the number of unknown parameters, which is possible by increasing the number of data clusters or using more resolutions. The performance index that is used for comparing with other works is Non-Dimensional Error Index (NDEI). It is the root mean square error divided by the standard deviation of checking data.

$$J = \sqrt{\frac{\sum_{j=1}^{n}(y_j - y_j^d)^2}{\sum_{j=1}^{n}(y_j^d - \overline{y})^2}} \tag{26}$$

where y_j^d is the desired output for the *j*th sample of data, y is the corresponding fuzzy system output and $\overline{y}$ is the average value of desired output.

In presented examples we have used a single fuzzy rule based (single resolution), where the output of fuzzy rule based is given by

$$f = \sum_k d_k \psi_k + d_0 \tag{27}$$

and the number of unknown parameters is one more than the cluster's number.

Example1: This example addresses the approximation of a single variable piecewise function given by

$$f(x) = \begin{cases} -2.186x - 12.864 & ; -10 \le x < -2 \\ 4.246x & ; -2 \le x < 0 \\ 10e^{-(0.5x+0.5)}\sin(x(0.03x+0.7)); & 0 \le x < 10 \end{cases} \tag{28}$$

This piecewise function is continuous and analyzable. However traditional analytical tools become inefficient and often fail due to following reasons; 1) The wide-band information hidden at the turning points, 2) The coexistence of linearity and nonlinearity. The sampled data which was selected for checking and training is distributed uniformly over [-10, 10] and 200 input-output pairs, as described below, are considered for each set.

TABLE I
Comparison of the Proposed Approach with FWN and WNN

Method	Number of unknown parameters	NDEI
Proposed fuzzy wavelet system	20	0.00101
First FWN [9]	27	0.00228
Second FWN [9]	37	0.00957
FWN [5]	28	0.021
WNN[11]	22	0.05057
WNN[12]	23	0.0480

$$\left\{X^d = \left[x(t), x(t-1), x(t-2), y(t-1)\right]; y^d = y(t)\right\} \tag{29}$$

As is shown in Table I, the performance index (NDEI) of the proposed approach is better than the others, yet the number of unknown parameters is decreased. In Table I, the NDEI is the average of ten simulations. In Fig. 2, excellent performance of the proposed fuzzy wavelet system is shown. For this example, the parameters are selected as $j = 1, b = 1, a = 115, \alpha = 38$

Example2: This example illustrates the prediction of a high dimensional chaotic time series. The considered chaotic series is the Mackey-Glass differential delay equation defined as

$$\dot{x}(t) = \frac{0.2x(t-\tau)}{1 + x^{10}(t-\tau)} - 0.1x(t) \tag{30}$$

where $\tau = 17$ and $x(0) = 1.2$. The prediction of future values of this time series is a benchmark problem. 1000 input-output pairs $\{X^d, Y^d\}$ are extracted from the Mackey-Glass time series of the following form, while it is also available in MATLAB documents.

$$\begin{Bmatrix} X^d = \left[x(t-18), x(t-12), x(t-6), x(t)\right] \\ Y^d = x(t+6) \end{Bmatrix} \tag{31}$$

where $t = 118$ to 1117. For comparison, similar to what the others have done [4], [5], the first 500 pairs were used as the training data set, and the remaining 500 pairs were the checking data set for validation of the learned fuzzy wavelet system. The performance index (NDEI) of the new proposed approach in comparison with the other approaches, for this example, is shown in Table II.

If the proposed fuzzy wavelet system is compared with one presented in [4], which considers the center of membership

TABLE II
Camparison of Generalization Capabilities.
The last 4 rows are from [4], [5].

Method	Training cases	NDEI
Proposed fuzzy wavelet system*	500	0.0061
Proposed fuzzy wavelet system**	500	0.0096
Fuzzy Wavelet Networks (FWN)[5]	500	0.0066
Fuzzy system proposed in [4]**	500	0.017
Back-Propagation NN	500	0.02
Six-order polynomial	500	0.04
Cascade-correlation NN	500	0.06
Linear predictive	2000	0.55

* Number of unknown parameters is 141.
** Number of unknown parameters is 81.

functions to be located on equally spaced lattice points, the performance index (NDEI) of the new approach is decreased noticeably, while the number of unknown parameters is equal. Also if the number of clusters is set to 140, the NDEI of the proposed method is less than the FWN approach which its unknown parameters are more than 150 [5]. For this example, the parameters are set as follows

$j = 1, \quad b = 1, \quad a = 0.4, \quad \alpha = 0.32$

Figure 3 shows the excellent performance of the proposed approach. In this figure the desired and the predicted values for checking data, and the difference between them is presented.

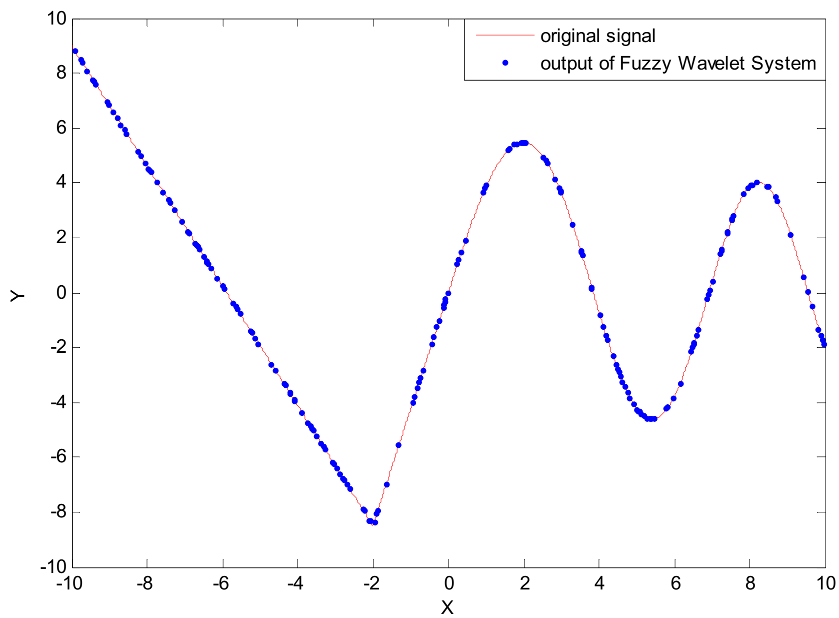

Fig. 2. Comparison between the original signal (solid line) and the output of the proposed Fuzzy Wavelet System (dotted line).

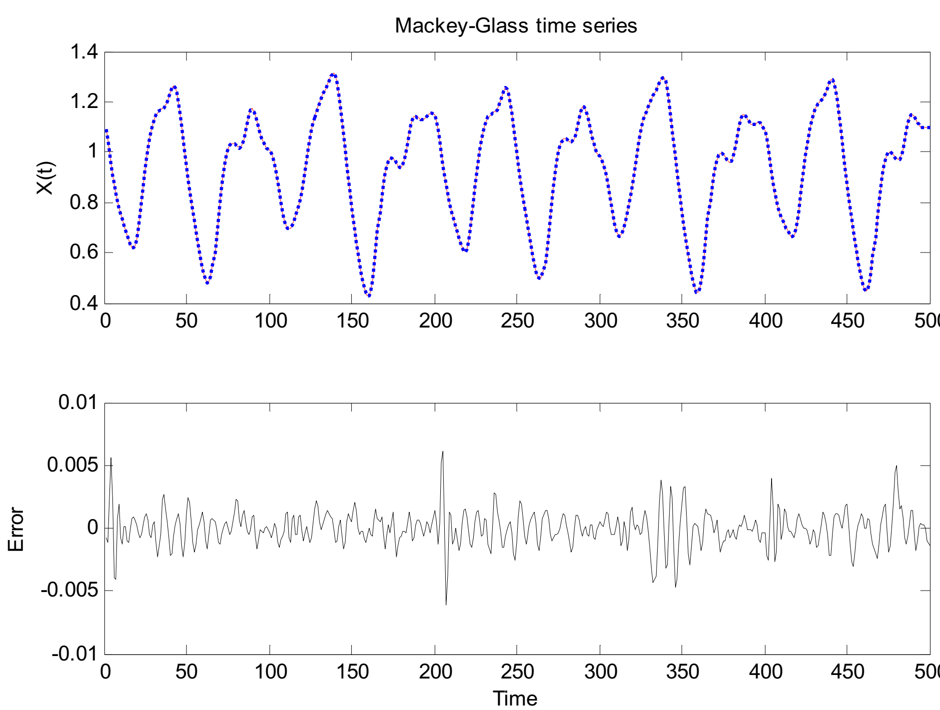

Fig. 3. a) Six step-ahead prediction by the proposed fuzzy wavelet system (dashed line) and Mackey-Glass time series (solid line), b) prediction error.

VI. CONCLUSION

Combining the rigorous approximation theory of wavelet basis functions and the linguistic information of fuzzy systems is a useful contraption for modeling the complex and nonlinear functions. In this paper, a new approach for tuning the parameters of fuzzy wavelet systems which are used for modeling of nonlinear systems is proposed. The Gustafson-Kessel algorithm is used for clustering the input data set and also applied for setting the nonlinear parameters of fuzzy basis functions. Consequently, by tuning the real linear coefficients of the proposed fuzzy system, it is shown that an accurate approximation of examined systems can be obtained.

REFERENCES

[1] L. X. Wang, "Fuzzy system: coexistence of challenge and opportunity for sentiment of ten years," Trans. of Automatic Control, vol. 27, no.4, pp. 585-590, 2001.

[2] L. X. Wang, "Stable adaptive fuzzy control of nonlinear systems," IEEE Trans. Fuzzy Systems, vol. 1, pp. 146-155, 1993.

[3] Ch.K. Lin, Sh. D. Wang, "Constructing a fuzzy model from wavelet transforms," IEEE Trans. Fuzzy Systems, pp. 394-399, 1996.

[4] Ch. K. Lin, Sh. D. Wang, "Fuzzy modeling using wavelet transform," Electronics Letters, vol. 32, no. 24, 1996.

[5] D.W.C. Ho, P.-A. Zhang, J. Xu, "Fuzzy wavelet networks for function learning," IEEE Trans. Fuzzy Systems, vol. 9, no. 1, pp. 200–211, 2001.

[6] H. Wang, J. Xiao, "T-S fuzzy system based on multi-resolution analysis and its function approximation," in Proc. 5^{th} Cong. on Intelligent Cont. & Automation, pp. 244-249, 2004.

[7] B. Balasko, J. Abonyi, B Feil, "Fuzzy clustering and data analysis toolbox, for use with MATLAB", available at: http://www.mathworks.com/matlabcentralfileexchange/loadFile.do?objectId=7473&objectType=file.

[8] D.E. Gustafson and W.C. Kessel, "Fuzzy clustering with fuzzy covariance matrix," in Proceedings of the IEEE CDC, San Diego, pp. 761-766, 1979.

[9] E. Karatepe, M. Alcı, "A new approach to fuzzy wavelet system modeling," International Journal of Approximate Reasoning, vol. 40, pp. 302–322, 2005.

[10] T. Kugarajah and Q. Zhang, "Multidimensional wavelet frames," IEEE Trans. Neural Networks, vol. 6, pp.1552- 1556, 1995.

[11] Q. Zhang and A. Benveniste, "Wavelet networks," IEEE Trans. Neural Networks, vol. 3, pp. 889–898, 1992.

[12] J. Chen and D. D. Bruns, "WaveARX neural network development for system identification using a systematic design synthesis," Ind. Eng. Chem. Res., vol. 34, pp. 4420–4435,1995.