# CHAOTIC TIME SERIES PREDICTION USING WAVELET DECOMPOSITION

Iulian B. Ciocoiu

*Technical University Iasi, Faculty of Electronics and Telecommunications*
*P.O. Box 877, Iasi, 6600, Romania, e-mail: iciocoiu@tuiasi.ro*

***Abstract* - A novel approach to chaotic time series prediction is proposed. It is based on the use of the Discrete Wavelet Transform for obtaining a proper decomposition of the original sequence and standard multilayer neural networks for performing the prediction of the individual components. Simulation results for the case of chaotic signals obtained by integrating the Lorenz equations are presented, and directions for further research are outlined.**

## I. INTRODUCTION

Artificial neural networks have already been used with significant success in time series analysis. Both purely feedforward and recurrent architectures have been tested and chaotic time series prediction was most frequently used as a benchmark. In this paper a novel approach to this problem is proposed, which is based on a multiscale decomposition of the original sequence using the Discrete Wavelet Transform (DWT) [1], followed by a prediction step on each component. Finally, the global predicted output is obtained by summing up the individual results on each scale.

In the following, a basic introduction to wavelet analysis is given. Wavelets represent a special class of functions which can generate bases in functional vector spaces, when certain mathematical requirements are fulfilled. In this respect, wavelet analysis is similar to the well-known Fourier analysis, whose basic elements are sines and cosines. Anyway, they differ in a major point: wavelet signals are *both* time and frequency (in fact, scale) localized, as opposed to sines and cosines, which are not time localized. This makes wavelets ideally suited for dealing with signals containing sharp spikes or sudden changes. The main idea behind the wavelet analysis is to implement some kind of *zoom* effect on the data, that is to successively decompose the original sequence in a sum of individual series containing progressively more information. As a result, one may obtain both a coarse (low-frequency) view on the analyzed data and more detailed (high-frequency) components.

Although the mathematical background of this research topic is quite difficult, we only point out that all the elementary basis functions are scaled and translated versions of a generic function called the *mother wavelet* $\phi(x)$. This function can be obtained by solving a so-called system of dilation equations of the form:

$$\begin{aligned} \phi(x) &= \sum_{k \in Z} c_k \phi(2x-k) \\ \Psi(x) &= \sum_{k \in Z} (-1)^k c_{1-k} \phi(2x-k) \end{aligned} \qquad (1)$$

where $\psi(.)$ designates the associated *scaling function* (which enters in the recovery of the low-frequency smooth component of the data), and $c_k$ are a set of real coefficients. These coefficients verify regularity, orthogonality and normalization requirements and are specific to the chosen wavelet basis [1].

Wavelets have been used in image compression, noise removal, solution of differential equations, and an enormous amount of literature related to the subject is available. Software implementations have also been elaborated, as stand alone products or as components of more general programs. For example, we have performed our simulations using a MATLAB toolbox called Wavelab, available from Stanford University [2].

## II. THE PROPOSED APPROACH

The use of orthogonal transforms in signal processing applications has been widely considered as an option in order to simplify the task at hand. In data compression problems for example, the Karhunen-Loeve Transform (KLT) [5] (also known as Principal Component Analysis) or a close approximation of it offered by the Discrete Cosine Transform DCT have been usually employed. In filtering applications, the Discrete Fourier Transform (DFT) or its finite-length versions Windowed Fourier Transform and Short-Time Fourier Transform represent classical approaches. Anyway, the end-effects due to the length limitation are oftenly very important and careful data handling is required in order to obtain proper results. Wavelet analysis of discrete sequences is based on the use of the Discrete Wavelet Transform (DWT), which is characterized by the following matrix:

$$C=\begin{bmatrix} c0 & c1 & c2 & c3 & & & & & & \\ c3 & -c2 & c1 & -c0 & & & & & & \\ & & c0 & c1 & c2 & c3 & & & & \\ & & c3 & -c2 & c1 & -c0 & & & & \\ \vdots & \vdots & & & & \ddots & & & & \\ & & & & & & c0 & c1 & c2 & c3 \\ & & & & & & c3 & -c2 & c1 & -c0 \\ c2 & c3 & & & & & & & c0 & c1 \\ c1 & -c0 & & & & & & & c3 & -c2 \end{bmatrix}$$

If $Y=[y_0, y_1, \ldots, y_M]^T$ is a data vector, the DWT is obtained by repeatedly multiplying matrix C with the (reordered) intermediary results, according to the scheme depicted in Fig.1. The final result is a sequence of numbers $S_1$, $S_2$, … representing the low-frequency (smooth) information contained in vector Y, and a concatenation of detail (high-frequency) information $D_1$, $D_2$, $D_3$, $D_4$, $d_1$, $d_2$, ... .

It is worth noting that the Discrete Wavelet Transform is less sensitive to finite-length effects than the Discrete Fourier Transform, since it contains only a *small number of non-zero coefficients.*

Prediction applications should take into account that ***no future information*** can be introduced at the input of the predictive system. This means that no global processing on the available data base is allowed. Instead, we could apply local processing in a sliding window and retain all or a part of the resulting information to be used for prediction purposes on the subsequent time instances. The present approach obeys this rule and works as

follows: we get a discrete sequence of 500 points obtained by integrating the well-known Lorenz

follows: we get a discrete sequence of 500 points obtained by integrating the well-known Lorenz equations, which exhibit chaotic behavior for certain values of the parameters. We consider a sliding window whose length equals 128 samples and perform a 4-scale DWT decomposition employing the Daubechies_4 mother wavelet [3]. In order to avoid the end-effects we used an approach presented in [4] (which is implemented in the Wavelab toolbox), and which sligthly modifies the border wavelets. We retain only the *last point* from each scale and then move the sliding window one sample at a time and repeat the above procedure. Finally, a set of distinct resolution individual sequences results, each point of which represents the last point of the corresponding sliding window.

## III. SIMULATION RESULTS

Prediction is performed on each of the individual series and the final predicted sequence is obtained by simply summing up the results. We have used multilayer perceptron (MLP) networks with 15 inputs, 2 hidden layers with 5, respectively 3 neurons, and 1 linear output neuron. The inputs were made by 15 tapped-delay samples of the corresponding resolution scale. The training algorithms were: a) back-propagation, the momentum version; b) a back-propagation type algorithm based on the use of the Extended Kalman Filter (EKF) [5],[6],[7]. For comparison purposes we also give the results obtained with a simple FIR network. The training set was made by 300 input-output pairs, and a test set of 30 pairs was used for early stopping using crossvalidation.

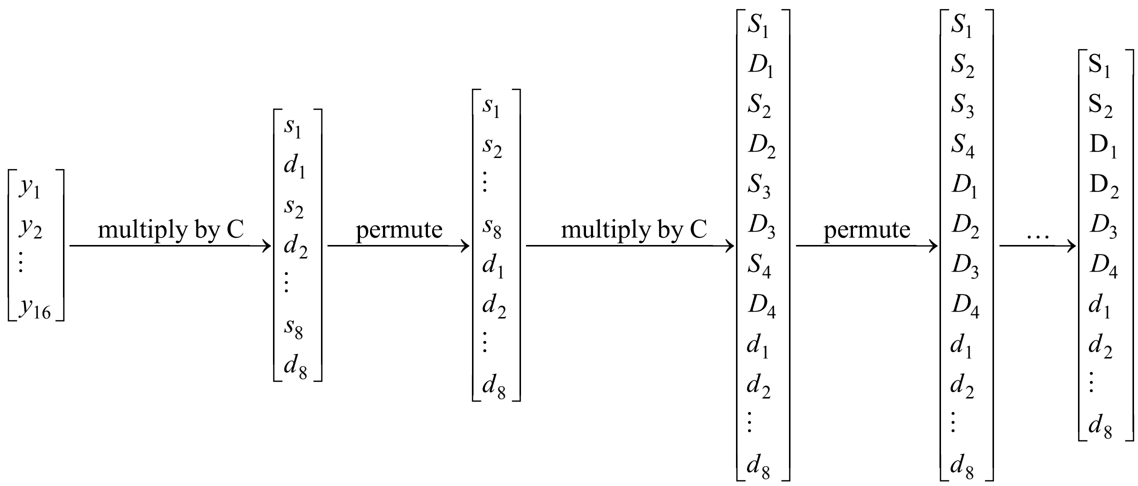



Fig. 1. Computational scheme for obtaining the Discrete Wavelet Transform of vector y

The global results are given in Fig. 2. The waveforms corresponding to the individual scales are presented in Fig. 3. In Table 1 the final MSE values are indicated. These figures clearly show that the neural networks are

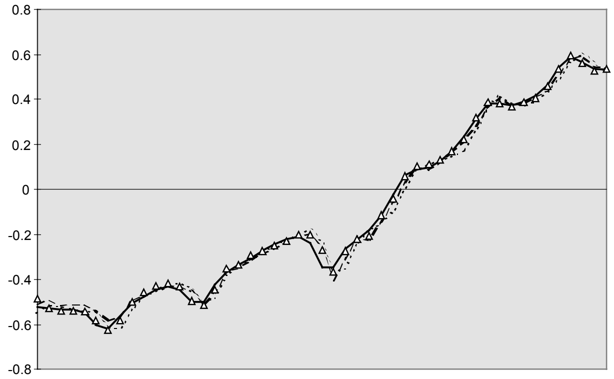
able to closely predict both the individual lower-resolution scales, and the original sequence.
We have tested the possibility of training a FIR filter to obtain the global predicted output instead of simply summing up the results on individual scales, but no clear improvement was obtained.

## IV. CONCLUSIONS

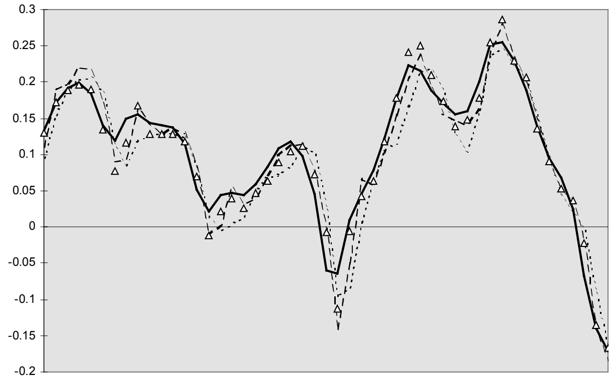
The wavelet decomposition approach represents a useful solution to chaotic time series prediction. Although it is more computationally demanding, it greatly simplifies the global task of the predictive system, by decomposing the original problem in a set of simpler ones. This approach should prove extremely useful in prediction applications involving complicated maps and/or significant noise levels superimposed on the true data. In this case, the analysis of higher resolution scales, which most probably contain mainly noise, could be entirely dropped.

***References:***

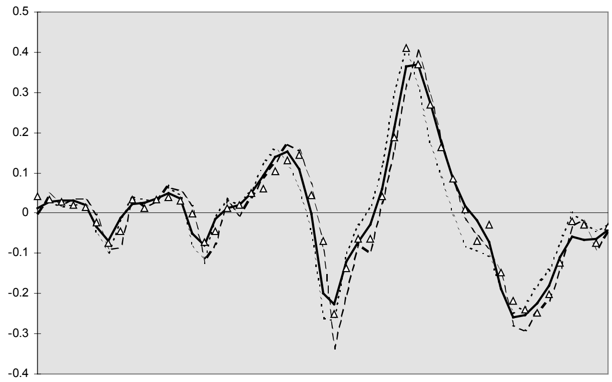

[1] Graps, A., “An Introduction to Wavelets”, *IEEE Comput. Science & Eng.*, Summer 1995, vol. 2, no. 2
[2] Wavelab, MATLAB Toolbox, available from Stanford University on Internet
[3] Daubechies, I., “Ten Lectures on Wavelets”, SIAM, Philadelphia, Pennsylvania, 1992
[4] Cohen A., I. Daubechies, and P. Vial, “Wavelets on the interval and fast wavelet transforms”, *Applied and Comput. Harmonic Analysis*, vol. 1, pp. 54-81, Dec. 1993
[5] Haykin, S. Neural Networks - A Comprehensive Foundation, *IEEE Press*, 1994
[6] Wan, E.A., and A.T. Nelson, "Dual Kalman Filtering Methods for Nonlinear Prediction, Smoothing, and Estimation", *Proc. NIPS'96*
[7] Lange, F., EKFNet software, available on Internet from DLR (German Aerospace Research Establishment)


a)

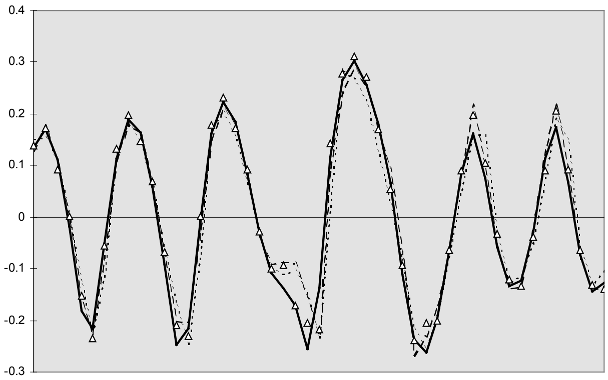



b)

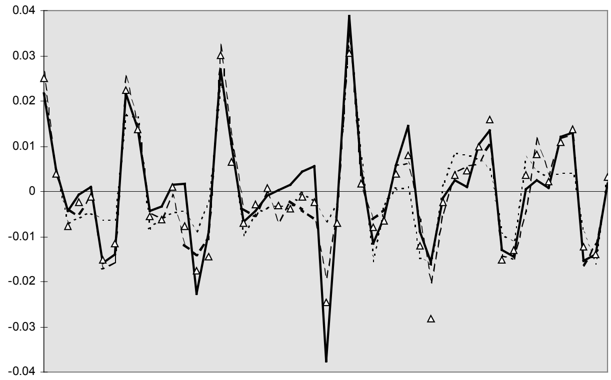



c)

Fig. 2: Real (solid line) and predicted components:
a) low frequency smooth component (m=0);
b)-e) detail components (m=1..4).
FIR: ····· ; MLP: — · · — ; EKF: Δ




d)




e)

Fig.2. (*continued*)

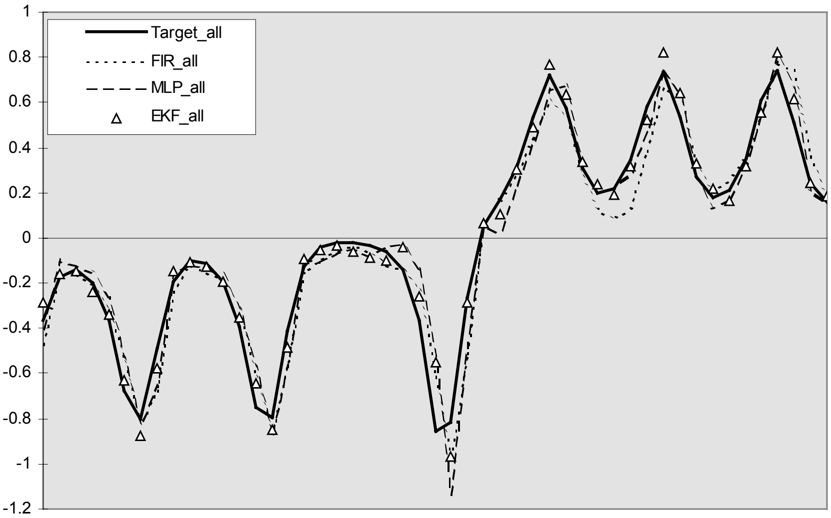



Fig. 3. Global real (solid line) and predicted chaotic time series: FIR: ····· ; MLP: —··— ; EKF: Δ

Table 1. MSE values for the individual scales and the global reconstructed sequence

| | Scale 0 | Scale 1 | Scale 2 | Scale 3 | Scale 4 | Global results |
|---|---|---|---|---|---|---|
| FIR | 0.461 | 0.581 | 1.171 | 0.668 | 0.02 | 4.551 |
| MLP | 0.259 | 0.223 | 0.49 | 0.253 | 0.0088 | 3.6 |
| EKF | 0.094 | 0.105 | 0.244 | 0.277 | 0.0062 | 1.72 |