Wavelet-GA-ANN Based Hybrid Model for Accurate Prediction of Short-Term Load Forecast

Nidul Sinha[1] *Member IEEE* and Loi Lei Lai[2] *Fellow IEEE*, Palash Kumar Ghosh[1], Yingnan Ma[2]

Abstract — **This paper proposes a hybrid model developed through wiser integration of wavelet transforms, floating point GA and artificial neural networks for prediction of short-term load. The use of wavelet transforms has added the capability of capturing of both global trend and hidden templates in loads, which is otherwise very difficult to incorporate into the prediction model of ANN. Auto-configuring RBF networks are used for predicting the wavelet coefficients of the future loads. Floating point GA (FPGA) is used for optimizing the RBF networks. The use of GA optimized RBF network has added to the model the online prediction capability of short-term loads accurately. The performance of the proposed model is validated using Queensland electricity demand data from the Australian National Electricity Market. Results demonstrate that the proposed model is more accurate as compared to RBF only model.**

Index Terms— **Wavelet transforms, Genetic Algorithm, ANN, RBF networks, Load forecast, Short-term load forecast.**

I. INTRODUCTION

Deregulation of power utility industry being a reality today, which has resulted into the competition in every aspects in power systems; be it in power generation, or in transmission or in energy consumption, professional management of electric energy is of utmost importance. Many power systems not only are being pushed to their limits to meet their customers' demands, but also spend a lot of resources in their operation scheduling. Furthermore, power systems need to operate at even higher efficiency in a deregulated electricity market whereby the generating companies (Gencos) and distribution companies (Discos) have to compete in order to maximize their profits. Accurate prediction of load consumption pattern is becoming very important function to a utility company, as it is needed to support for wiser management decisions. A forecast that exceeds the actual load may lead to extra power being generated and therefore may result in excessive investment in a power plant that is not fully utilized. On the other hand, a forecast that is too low may lead to some revenue loss due to loss of opportunity of selling power to neighboring utilities. Hence, accurate electricity load forecasting (LF), including very short-term, short-term, mid-term, and long-term, plays a vital role in ensuring adequate electricity generation to meet the customer's demands in the future. LF also helps to build up cost effective risk management plans for the participating companies in the electricity market. Consequently, good operational, planning and intelligent management decision making, such as, economic scheduling of generation capacity, scheduling of fuel purchase, ability to avoid unnecessary start-ups of generating units, planning the scheduling of peaking power, buying or selling electricity at best price, and scheduling of ancillary services, all of them can be carried out based on accurate LF, which forecasts the load of a few minutes, hours, days, weeks, months ahead. The aim of LF is to predict future electricity demand based on historical load data, and currently available data.

To facilitate accurate load-forecasting analysis, a robust noise filtering and trend analysis algorithm must be used to enable effective eventual automation of the analysis of large volumes of data generated by the monitoring and recording of load consumption readings in any particular system. Currently, several forecasting schemes utilize Artificial Intelligence (AI) methods like ANN and GA to perform load-forecasting tasks. The common problem with such a method is that an AI scheme is only as intelligent as the program that trains it. This in turns depends heavily on the reliability of the training data collected. If such training data were in the first place corrupted by noise, it would mean that pre-processing of such data would be necessary. All these add to the implementation cost and set-up time. A good trend analysis scheme should be able to de-noise the electrical noise inherent in the data, and disregard portions of data where monitoring devices might have failed, giving lower resolution readings as a result of, and be able to take a macro view of the trend while preserving temporal information. The analysis of non stationary signals like load consumption data often involves a compromise between how well important transients can be located and how finely evolutionary behaviors can be detected. Extremely noisy data poses a problem to the operator as how to ascertain the amount of noise in the retained high frequency transient data [19].

The interest in applying neural networks to electric load forecasting began more than a decade ago. Artificial neural networks based methods for forecasting have shown ability to give better ability in dealing with the nonlinearity and other difficulties in modeling of the time series data. ANNs have been applied recently in the area of time-series forecasting due to their flexibilities in data modeling [1-2]. Most of the approaches reported since are based on the use of an MLP network as an approximator of an unknown nonlinear relation. There have been some pioneering works on applying wavelet techniques together with ANN to time series forecasting, [3-7]. Among ANN based forecasting

[1]Nidul Sinha and Palash Kumar Ghosh are with the Department of Electrical Engineering, NIT, Silchar, Assam, India-788010, (email:nidulsinha@rediffmail.com).
[2]Loi Lei Lai and Yingnan are with the Energy Systems Group, City University, London, UK. (email: l.l.lai@city.ac.uk).

methods, Radial Basis Function (RBF) networks have been widely used primarily because of their simple construction and easier training is as compared to Multi-layer Perceptrons (MLPs) in addition to their capability in inferring the hidden relationship between input and desired target patterns. This capability is attributed to its property that it can approximate any continuous function to any degree of accuracy by constructing localized radial basis functions. From the standpoint of preserving characteristics of different classes, this local approximation approach has the advantage over the global approximation approach of multi-layer perception networks.

As large amounts of historical load patterns are needed in a typical load-forecasting algorithm, even low sampling rates of 1 sample per minute generates a huge amount of data. Hence, the effective compression of large data and faithful reconstruction of original signal from compressed data is a major challenge for time series data. Also, when an ANN, especially RBF network, is trained with huge data (with noise), it may result into not only a big network model and very time consuming training but also that the network may fail to capture the true features in the data. With the development of wavelet transforms, the difficulty of effective data compression and faithful retrieval of original data can be well tackled. This tempted researchers to try RBF networks model combined with wavelet transformed data for capturing useful information on various time scales. These strategies approximate a time-series at different levels of resolution using multi-resolution decomposition. Recent works [10] stresses on the use of shift invariant wavelet transforms, which is an auto correlation shell representation technique, for making the analysis of time series data easier. This technique is employed to reconstruct singles after wavelet decomposition. With the help of this technique, a time series can be expressed as an additive combination of the wavelet coefficients at different resolution levels. These data are then applied to build Neural-Wavelet based forecasting models to predict electricity demand as from the data obtained from a real electricity market.

In this work, autocorrelation shell representation base wavelet transform is used to approximate Short-term Load Forecast (STLF) at different levels of resolution using multi-resolution decomposition. This decomposed data is used for training the RBF network for predicting the wavelet coefficients of future loads. RBF networks optimized with the help of FPGA. This technique is then applied to build Neural-Wavelet based forecasting models to predict electricity demand as from the data obtained from a real electricity market.

In view of the above, the main objectives of the present work are:

(i) To develop a wavelet based RBF network model for accurate prediction of short-term load forecast. The use of neural network will enable to online prediction of STLF in an effective and efficient way.

The rest of the paper is organized as follows: In Section II describes the concept of wavelet transforms and their use for STLF and section III presents the concept of RBF network and development of hybrid wavelet-RBF model. Section IV presents the experimental results together with discussions. Conclusions are drawn in Section V.

II WAVELET TRANSFORMS IN LOAD FORECAST

Wavelet transforms [26, 27] though known previously has gained much attention only recently. It has been exploited in many fields like seismic studies, image compression, signal processing processes and mechanical vibrations. The flexible time-scale representations of wavelet transform has found its place in many applications that traditionally used modified forms of Fourier Transforms (FT) like Short Time FT (STFT) and the Gabor Transforms. Its impressive temporal content and frequency isolation features have tempted researchers to use them in the area of power systems analysis.

Wavelet transforms provide a useful decomposition of a signal, or time series, so that faint temporal structures can be revealed and handled by nonparametric models. They have been used effectively for image compression, noise removal, object detection, and large-scale structure analysis, among other applications.

2.1 TIME SERIES AND WAVELET DECOMPOSITION IN LOAD FORECASTING

2.1.1 Á TROUS WAVELET DECOMPOSITION

The continuous wavelet transform of a continuous function produces a continuum of scales as output. On the other hand, input data is usually discretely sampled, and furthermore a dyadic or two-fold relationship between resolution scales is both practical and adequate. The latter two issues lead to the discrete transform. Fig.1 shows the wavelet decomposition.

Wavelet decomposition provides a way of analyzing a signal in both time and frequency domains. For a suitably chosen mother wavelet function ψ a function f can be expanded as:

$$f(t) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} w_{jk} 2^{j/2} \psi(2^{j} t - k) \qquad (1)$$

where the functions $\psi(2^j t - k)$ are all orthogonal to each other. The coefficients w_{jk} gives information about the behavior of the function f concentrating on the effects of scale around 2^{-j} near time $t \times 2^{-j}$. This wavelet decomposition of a function is closely related to a similar decomposition (the discrete wavelet transform, DWT) of a signal observed in discrete time.

It is well known that DWT has many advantages in compressing a wide range of signals observed in the real world. However, in time series analysis, DWT often suffers from a lack of translation invariance. This means that DWT based statistical estimators are sensitive to the choice of origin.

The output of a discrete wavelet transform can take various forms [21]. Traditionally, a triangle (or pyramid in the case of *2*-dimensional images) is often used to represent all that is worth considering in the sequence of resolution scales. Such a triangle comes about as a result of decimation or the retaining of one sample out of every two. The major advantage of decimation is that just enough information is retained to allow exact reconstruction of the input data. Therefore decimation is ideal for effective compression. However, it can be easily shown that the storage required for the wavelet-transformed data is exactly the same as is required by the input data. The computation time for many wavelet transform methods is also linear in the size of the input data, i.e. $O(n)$ or *n*-length input time series. Also, with the decimated form of output it is less easy to visually or graphically relate information at a given time point at different scales. More problematic is their lack of shift invariance. This means that if the last few values of the input time series are deleted, then the wavelet transformed, decimated output data will be quite different from heretofore. One way to solve this problem at the expense of greater storage requirements is by means of a redundant or non-decimated wavelet transform.

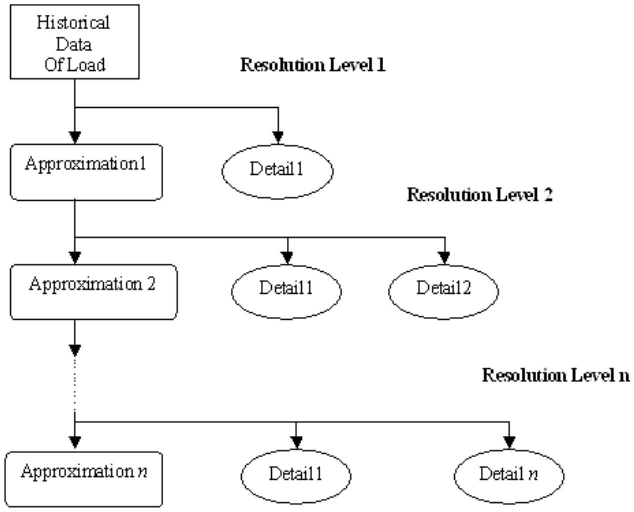


Fig. 1 Wavelet Decomposition Process

A non-decimated wavelet transform based on an *n*-length input time series, then, has an *n*-length resolution scale for each of the resolution levels of interest. Therefore, information at each resolution scale is directly related at each time point. This results in shift invariance. Finally, the extra storage requirement is by no means excessive.

An *à trous* algorithm is used to realize the shift-invariant wavelet transforms. Such transforms are based on the so-called auto-correlation shell representation [10] by dilations and translations of the auto-correlation functions of compactly supported wavelets.

By definition, the auto-correlation functions of a compactly supported scaling function ϕ(x) and the corresponding wavelet Ψ(x) are as follows:

$$\phi(x) = \int_{-\infty}^{\infty} \phi(y)\phi(y-x)dy$$
$$\psi(x) = \int_{-\infty}^{\infty} \psi(y)\psi(y-x)dy \qquad (2)$$

The set of functions $\{\overline{\psi}_{j,k}(x)\}_{1\le j\le n_0, 0\le k\le N-1}$ and $\{\overline{\phi}_{n_0,k}(x)\}_{0\le k\le N-1}$ is called an auto correlation shell, where:

$$\overline{\psi}_{j,k}(x) = 2^{-j/2}\psi(2^{-j}(x-k))$$
$$\overline{\phi}_{n_0,k}(x) = 2^{-n_0/2}\phi(2^{-n_0/2}(x-k)) \qquad \text{……………(3)}$$

A set of filters $P = \{p_k\}_{-L+1\le k\le L-1}$ and $Q = \{q_k\}_{-L+1\le k\le L-1}$ can be defined as:

$$\frac{1}{\sqrt{2}}\phi\left(\frac{x}{2}\right) = \sum_{k=-L+1}^{L-1} p_k \phi(x-k)$$
$$\frac{1}{\sqrt{2}}\psi\left(\frac{x}{2}\right) = \sum_{k=-L+1}^{L-1} q_k \phi(x-k) \qquad \text{………………………..(4)}$$

Using the filters *P* and *Q*, the pyramid algorithm for expanding into the auto-correlation shell can be obtained as:

$$c_j(k) = \sum_{l=-L+1}^{L-1} p_l c_{j-1}(k+2^{j-1}l) \qquad \text{…………………..…..(5)}$$

$$w_j(k) = \sum_{l=-L+1}^{l=L-1} q_l c_{j-1}(k+2^{j-1}l) \qquad \text{……………………..(6)}$$

These shell coefficients obtained from Eqs. (5) and (6) can then be used to directly reconstruct the signals. Given smoothed signal at two consecutive resolution levels, the detailed signal can be derived as:

$$w_j(k) = \sqrt{2}c_{j-1}(k) - c_j(k) \qquad \text{…………………………(7)}$$

The process of generating wavelet coefficient series is further illustrated with the block diagram as shown in Fig.2. Then the original signal $c_0(k)$ can be reconstructed from the coefficients $\{w_j(k)\}_{1\le j\le n_0, 0\le k\le N-1}$ and residual $\{c_{n_0}(k)\}_{0\le k\le N-1}$:

$$c_0(k) = 2^{-n_0/2}c_{n_0}(k) + \sum_{j=1}^{n_0} 2^{-j/2} w_j(k) \qquad \text{……………(8)}$$

for *k*=0,…, *N*−1, where $c_{n_0}(k)$ is the final smoothed signal. To make more precise predictions the most recent data shall

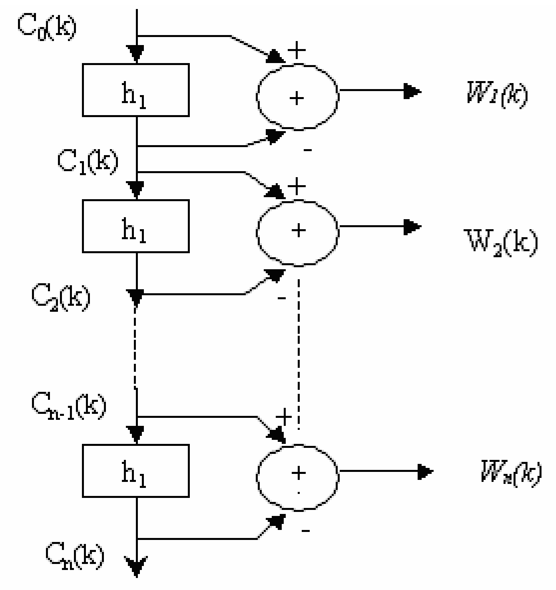


Fig.2. À Trous Wavelet Transform of a Time-Series Signal

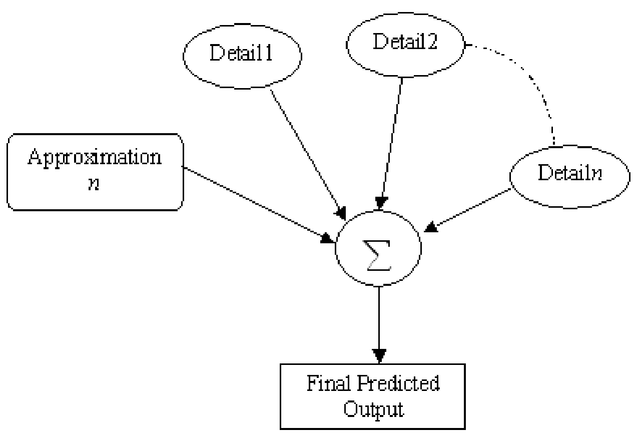


Fig. 3: Wavelet Recombination Process

be used. In case of adaptive learning, the previous data is penalized with forgetting factors. The time-based à trous filters similar to that of are used to deal with the boundary condition. Fig.3 shows the wavelet recombination process.

III. RADIAL BASIS NETWORKS

An RBF is a function which has in-built distance criterion with respect to a center [29]. A typical RBF neural network consists of three layers (input, hidden, output). The activation of a hidden neuron is determined in two steps: The first is to compute the distance (usually the Euclidean norm) between the input vector and a center c_i that represents the i^{th} hidden neuron; second, a function, that is usually bell shaped, is applied, using the obtained distance to get the final activation of the hidden neuron. In the present case the well known Gaussian function G(x) is used.

$$G(x) = \exp\left(-\frac{\|x - c_i\|^2}{2\sigma^2}\right) \tag{9}$$

The parameter σ is called unit width (spread factor) and is determined using the GA. All the widths in the network are fixed to the same value and this results in a simpler training strategy. The activation of a neuron in the output layer is determined by a linear combination of the fixed nonlinear basis functions, i.e.

$$F(x) = w_0 + \sum_{i=1}^{M} w_i \phi_i(x), \tag{10}$$

where $\phi_i(x) = G(\|x - c_i\|)$ and w_i are the adjustable weights that link the output nodes with the appropriate hidden neurons and w_0 is the bias weight. These weights in the output layer can then be learnt using the least-squares method.

The present work adopts a systematic approach to the problem of centre selection. Because a fixed center corresponds to a given regressor in a linear regression model, the selection of RBF centres can be regarded as a problem of subset selection. The orthogonal least squares (OLS) method [8] can be employed as a forward selection procedure that constructs RBF networks in a rational way. The algorithm chooses appropriate RBF centres one by one from training data points until a satisfactory network is obtained. Each selected centre minimizes the increment to the explained variance of the desired output, and so ill-conditioned problems occurring frequently in random selection of centres can automatically be avoided. In contrast to most learning algorithms, which can only work if a fixed network structure has first been specified, the OLS algorithm is a structural identification technique, where the centres and estimates of the corresponding weights can be simultaneously determined in a very efficient manner during learning. OLS learning procedure generally produces an RBF network smaller than a randomly selected RBF network. Due to its linear computational procedure at the output layer, the RBF is shorter in training time compared to its back propagation counter part.

A major drawback of this method is associated with the input space dimensionality. For large numbers of inputs units, the number of radial basis functions required, can become excessive. If too many centres are used, the large number of parameters available in the regression procedure will cause the network to be over sensitive to the details of the particular training set and result in poor generalization performance (overfit).

The present work uses a floating point GA based algorithm for optimizing the centers and spread factors.

3.1 A HYBRID NEURAL-WAVELET MODEL FOR SHORT-TERM LOAD PREDICTION

The proposed hybrid neural-wavelet model for short-term load prediction is shown in Fig. 4. Given the time series *f*(*n*), *n*=1,…, *N*, the aim is to predict the *l*-th sample ahead, *f*(*N*+*l*), of the series. As a special case, *l*=1 stands for single step

prediction. For each value of l a separate prediction architecture is trained accordingly. The hybrid scheme basically involves three stages [3]. At the first stage, the time series is decomposed into different scales by autocorrelation shell decomposition; at the second stage, each scale is predicted by a separate RBF network; and at the third stage, the next sample of the original time series is predicted by another RBF network using the different scale's prediction.
For time series prediction, correctly handling the temporal aspect of data is one of the primary concerns. The *time-based à trous* transform as described above provides a simple but robust approach. Here we introduce an *à trous* wavelet transform based on the autocorrelation shell representation for the prediction model usage. This approach is realized by applying Eqs. (7) and (8) to successive values of t. As an example, given an electricity demand series of 1008 values, we hope to extrapolate into the future with 1 or more than 1 subsequent values. By the *time-based à trous* transform, we simply carry out a wavelet transform on values x_1–x_{1008}. The last values of the wavelet coefficients at time-point t=1008 are kept because they are the most critical values for

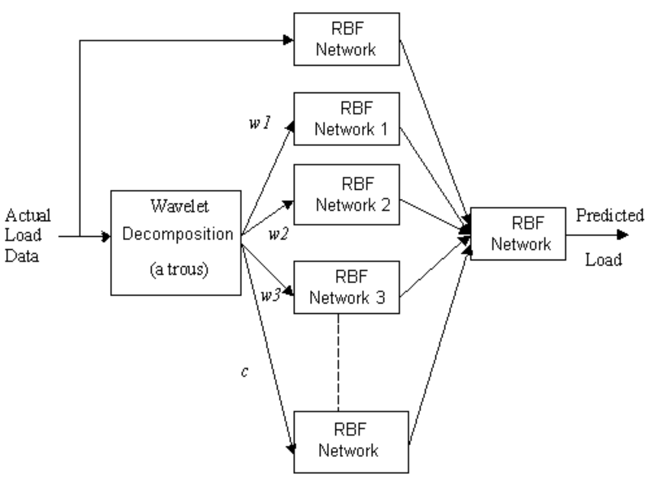


Fig. 4: Overview of the neural-wavelet multiresolution forecasting system. $w1$, …,wk are wavelet coefficients, c is the residual coefficient series.

prediction. Repeat the same procedure at time point t=1009, 1010… repeatedly. We empirically determine the number of resolution levels J, mainly depending on the inspection of smoothness of the residual series for a given J. Much of the high-resolution coefficients are noisy. Prior to forecasting, we get an over complete, transformed dataset.

In Fig.5, we show the behavior of the four-wavelet coefficients over 1008 points for a load series. Note that the data have been normalized for wavelet analysis. Normalization of data is an important stage, for training the neural network. The normalization of data not only facilitates the training process but also helps in shaping the activation function. It should be done such that the higher values should not suppress the influence of lower values and the symmetry of the activation function is retained. The input load data is normalized between the minimum value, -1 and the maximum value, +1 by using the formula.

$$\left(\frac{Actualvalue-Minimum}{Maximum-Minimum}\right)\times(Maximum-Minimum)+Minimum \qquad (11)$$

The load data should be normalized to the same range of values. The original time series and residual are plotted at the top and bottom in the same figure, respectively. As the wavelet level increases, the corresponding coefficients become smoother. As we will discuss in the next section, the ability of the network to capture dynamical behavior varies with the resolution level.

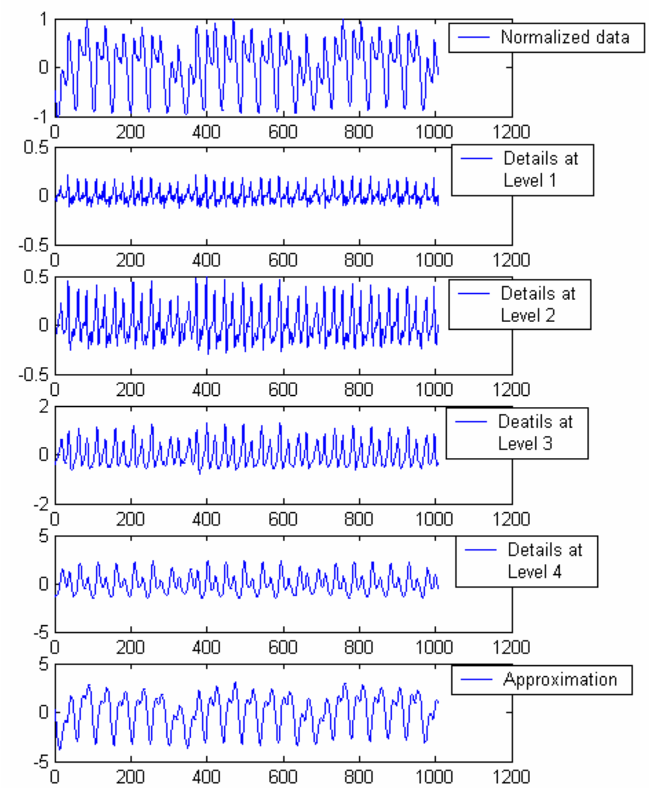


Fig. 5: Illustrations of the *à trous* wavelet decomposition of a series of electricity demand.

At the second stage, a predictor is allocated for each resolution level and the following wavelet's coefficients $w_i^j(t)$; j=0,…, J; i=1,…, N are used to train the predictor. All networks used to predict the wavelets' coefficients of each scale are of similar feed forward RBF perceptrons with D input units, one hidden layer with *radial basis function as an* activation function, and one linear output neuron. Each unit in the networks has an adjustable bias. The D inputs to the j-th network are the previous samples of the wavelets' coefficients of the j-th scale. In the proposed model implementation, each network is trained by the orthogonal least squares (OLS) method, which can be employed as a forward selection procedure that constructs RBF networks in a rational way. The procedure for designing neural network structure essentially involves selecting the input, hidden and output layers. At the third stage, the predicted results of all the different scales $\hat{w}^{j}_{N+i}(t)$, j=0,...,J are appropriately combined. Here we discussed and compared three methods of combination. In the first method, we simply applied the linear additive reconstruction property of the à *trous*, see Eq. (8). The fact that the reconstruction is additive allows the predictions to be combined in an additive manner. For comparison purpose, a plain RBF was also trained and tested for original time series, denoted as RBF, without any wavelet preprocessing involved.

The target selection is an important issue in applying neural networks to time series forecasting. A neural network, whose output neurons are reduced from two to one, will have half the number of network weights required. It also carries with important consequences for the generalization capability of the network. A single output neuron is the ideal case, because the network is focused on one task and there is no danger of conflicting outputs causing credit assignment problems in the output layer. Accordingly, it is preferred to have a forecasting strategy, which proceeds separately for each horizon in the second stage.

The target selection is an important issue in applying neural networks to time series forecasting. A neural network, whose output neurons are reduced from two to one, will have half the number of network weights required. It also carries with important consequences for the generalization capability of the network. A single output neuron is the ideal case, because the network is focused on one task and there is no danger of conflicting outputs causing credit assignment problems in the output layer. Accordingly, it is preferred to have a forecasting strategy, which proceeds separately for each horizon in the second stage.

IV RESULTS AND DISCUSSION

The proposed model is tested with two sets of historical data containing the electricity load for the month of July 2005 and month of July 2006, on a half-hourly basis; both sets of electricity load data of Queensland. The sets of electricity load data are downloaded from the NEMMCO website [30].

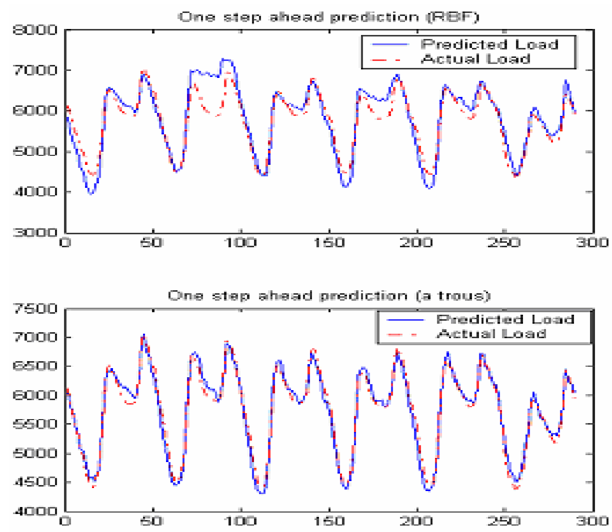
The simulation results are obtained through the use of four different programs. These programs were written in MATLAB command line in association with MATLAB toolboxes on wavelet, and neural network. Programs are run in a PC of Pentium IV, 256 MB RAM, 3.2 GHz.

Before the wavelet decomposition technique (*à trous*) is applied, the sets of historical load data are first normalized.

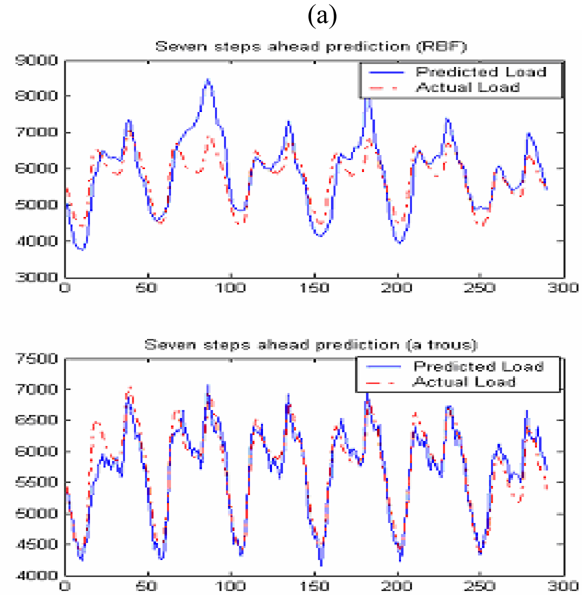
The model is evaluated based on it prediction errors. A successful model would yield an accurate time-series forecast. The performance of the model is hence measured using the absolute percentage error (APE), which is defined as

$$APE = \left(\frac{|x_i - y_i|}{x_i} \right) \times 100$$

where x_i is the actual values and y_i is the predicted values at time instance i. This error measure is more meaningfully represented as an average and standard deviation (S.D.) over the forecasting range of interests. Additional measure of the error is defined from the cumulative distribution function as the 90th percentile of the absolute percentage error, which provides an indication of the behavior of the tail of the distribution of errors and indicates that only 10% of the errors exceed this value.

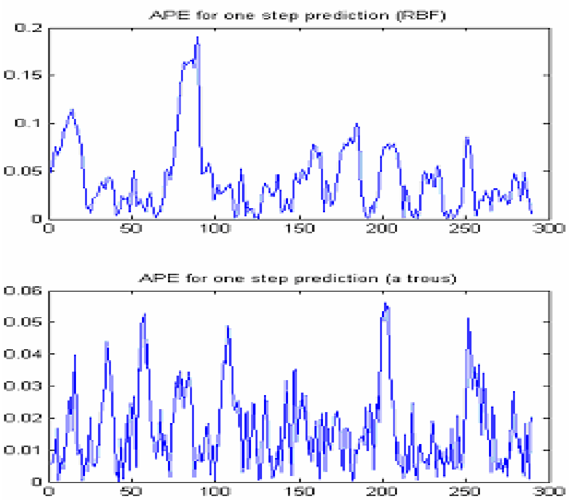


(a)

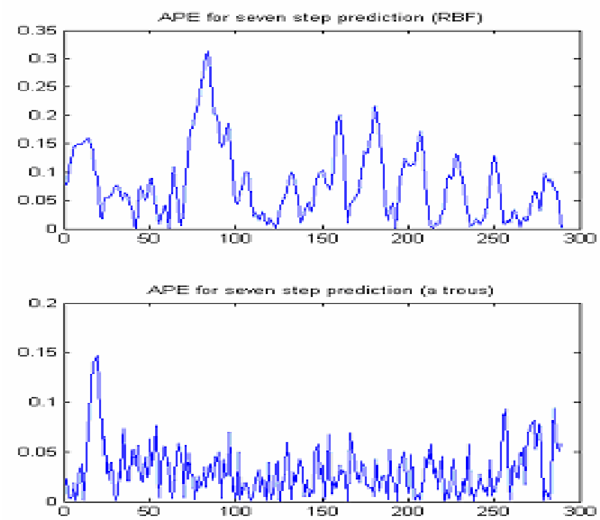


(b)

The forecasting results from the different forecasting schemes are presented in Table-1. The RBF network is optimized using floating point GA in terms of number of



(c)



(d)

Fig. 6: (a)-(d) Forecast results of the electricity demand, on a **testing data set**, by two schemes (i) RBF only and (ii) wavelet-RBF model for one step and seven steps ahead.

inputs, centers, and spread factor. The number of neurons in the hidden layer is auto-configured by the OLS algorithm. The Table-1 shows that the à trous wavelet transform system with adaptive combination coefficients for summing up the wavelet coefficients forecasting, is the best in seven step ahead forecasting for the testing data, with regards to the mean, variance and percentile over the absolute percentage error (APE).

Parameters for FPGA algorithm:
Population Size = 40
Maximum Iterations = 30
Operators for FPGA:

(i) Heuristic crossover
(ii) Uniform mutation
(iii) Normalized geometric select function

Table-1 Load forecast performance on testing data on APE measure for FGPA optimized spread factor and input.

	1	2	3	4	5	6	7
μ_R	5.00	5.11	5.66	6.14	6.73	7.20	7.85
σ^2_R	0.7	1.1	1.6	2.3	2.9	3.6	3.4
η_R	0.086	0.093	0.107	0.123	0.138	0.150	0.156
μ_w	1.43	1.13	1.19	1.53	2.22	1.55	1.14
σ^2_w	0.057	.0662	0.052	0.065	0.083	0.183	0.175
η_w	0.024	0.0216	0.022	0.025	0.033	0.036	0.034

$\sigma^2_R \times 10^{-3}$, $\sigma^2_w \times 10^{-3}$

Here, subscript R refers to the results with only RBF networks and subscript w refers to the results with hybrid wavelet-RBF model

.

V. CONCLUSION

A wavelet-GA-ANN based hybrid model is developed for accurate prediction of short-term load forecast in power systems. The structure of the auto-configuring RBF network is optimized using floating point GA. The short-term load data are transformed into wavelet coefficients using *à trous* wavelet transform before using them for training the RBF network. Use of wavelet transform ensures extraction of more hidden features (both in time and frequency) in a compressed form. The compression of data enables the RBF network to be more efficient. The results with different practical load data demonstrate that the proposed model is capable of accurately predicting the STLF for seven steps ahead.

VI.REFERENCES

1. D.K. Ranaweera, G.G. Karady, R.G. Farmer, "Effect of probabilistic inputs on neural network-based electric load forecasting," *IEEE Trans. Neural Works*, vol.7, No.6, pp.1528–1532. 1996.
2. J. N. Bastian, J. Zhu, V. Banunarayana, R. Mukerji, Forecasting energy prices in a competitive market, *IEEE Computer Appl. Power*, vol.13, No.3, pp. 40–45. 1999.
3. Amir B. Geva, "ScaleNet - Multiscale neural-network architecture for time series prediction," *IEEE Trans. Neural Networks,* Vol. 9, No. 5, pp. 1471-1482, 1998.
4. A. Oonsivilai and El-Hawary, "Wavelet Neural Network Based Short Term Load Forecasting of Electric Power System Commercial Load," *IEEE Proceedings*, pp.1223-1228, 1999.
5. O. Renaud, J. –L. Starck and F. Murtagh, "Wavelet-based Forecasting of Short and Long Memory Time Series," 2002..
6. Cheng-Jian Lin, "Wavelet Neural Networks with a Hybrid Learning Approach," *Journal of Information Science and Engineering* 22, pp. 1367-1387, 2006.
7. Chang-il Kim, In-keun Yu, Y.H. Song, "Kohonen neural network and wavelet transform based approach to short-term load forecasting," *Electric Power Systems Research* 63 pp.169-176, 2002.
8. S. Chen, C. F. N. Cowan, and P. M. Grant, "Orthogonal least squares learning algorithm for radial basis function networks," *IEEE Trans. Neural Networks,* vol. 2, pp. 302–309, 1991.

9. R.R. Coifman, D.L. Donoho, "Translation-invariant de-noising, in: A. Antoniades, et al. (Eds.), Wavelets and Statistics", Springer Lecture Notes, Springer, 1995.
10. N. Satio, G. Beylkin, "Multiresolution representations using the auto-correlation functions of compactly supported wavelets", *IEEE Trans. Signal Processing*, 41, No. 12 (1993) pp. 3584–3590.
11. Stephane *G.* Mallat, "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation," *IEEE Trans. Pattern Analysis and Machine Intelligence.* VolII.No.7, pp. 674-693, July 1989.
12. S.J. Yao, Y.H. Song, L.Z. Zhang, X.Y. Cheng, "Wavelet transform and neural networks for short-term electrical load," *Energy Conversion and Management*, 41: 1975-1988, 2000.
13. D. W. Bunn, "Forecasting loads and prices in competitive power markets," *Proc. of the IEEE,* Vol. 88, No.2, Feb. 2000, pp. 163-169.
14. A. D. Papalexopoulos and T. C. Hesterberg, "A regression based approach to short-term system load forecasting," *IEEE Trans. on Power System,* Vol. 5, No. 4, pp. 1535 - 1547, Nov. 1990.
15. U. Lotric, A. Dobnikar, "Predicting time series using neural networks with wavelet-based denoising layers," Neural Comput & Applic (2005) 14: 11–17.
16. Rong Gao and Lefteri H. Tsoukalas, "Neural-wavelet Methodology for Load Forecasting" *Journal of Intelligent and Robotic Systems* 31 149–157, 2001.
17. P. Hajto and M. Skrzypek "Wavelet-Neuronal Resource Load Prediction for Multi- processor Environment," LNCS 3019, pp. 119–124, 2004.
18. L. E. Kuo and S. S. Melsheimer "Using Genetic Algorithm to Estimate the Optimum Width Parameter in Radial Basis Function Networks" Proceedings of the American Control Conference, Baltimore, Maryland, pp.1368-1372, June 1994.
19. P. J. Santos, A. G. Martins, A. J. Pries, J. F. Martins, R. V. Mendes, "Short-Term Load Forecast Using Trend Information and Process Reconstruction," *International Journal of Energy Research*, pp.811-822, 2006.
20. B. Satish, K.S. Swarup, S. Srinivas, A. Hanumantha Rao, "Effect of temperature on short term load forecasting using an integrated ANN," Electric Power Systems Research 72 (2004) 95–101.
21. D. Benaouda, F. Muratagh, J.L. Starck, and O. Renaud. " Wavelet-Based Nonlinear Multiscale Decomposition Model for Electricity Load Forecasting," Neurocomputing 70 Elsevier, pp 139-154,2006.
22. Zunxiong Liu, Zhijun Kuang, Deyun Zhang, "Short-term Load Forecasting Method Based on Wavelet and Reconstructed Phase Space," IEEE conference, vol 8. pp. 4813- 4817, 2005.
23. A. K. Erkmen, Topalli, "Four methods for short-term load forecasting using the benefits of artificial intelligence," Electrical Engineering 85 pp. 229–233, Springer-Verlag 2003.
24. Tomonobu Senjyu, Hitoshi Takara, Katsumi Uezato, and Toshihisa Funabashi, "One-Hour-Ahead Load Forecasting Using Neural Network," *IEEE Trans on Power System*, Vol.17, No. 1, pp.113-118, Feb. 2002.
25. D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*. Reading, MA: Addison-Wesley, 1989.
26. Daubechies, I., ten Lectures on Wavelets, CBMS-NSF, SIAM, 1992.
27. K. P. Soman, and K. I. Ramachandran, "*Insight Into Wavelets From Theory to Practice*," Prentice-Hall of India, 2004.
28. Lorrin Philipson and H. Lee Willis "*Understanding Electric Utilities and De-regulation*" Marcel Decker, Inc.1999.
29. J. Park and I.W. Sandberg : "Universal approximation using radial basis function networks", Neural Computation, Vol.3(2), 1991, pp.246-257.
30. Australian National Electricity Market Management Company Limited (NEMMCO) website, http://www.nemmco.com.au/.

VII BIOGRAPHIES

Nidul Sinha received his B.E. degree in electrical engineering from Calcutta University, in 1984 and M.Tech. degree in power apparatus and systems from Indian Institute of Technology, New Delhi, in 1989. He received his Ph.D. degree in electrical engineering from Jadavpur University. His research interests include application of soft computing techniques to operation, control and economics of electrical power systems, deregulation and optimization. He is a reviewer of IEEE PWRS, PWRD, PESL, IEE part-c, EPSR, and DSP (Elsevier).

Loi Lei Lai (SM'92, F'07) received the B.Sc. (First Class Honours) and the Ph.D. degrees from Aston University in Birmingham, UK, in 1980 and 1984, respectively. In 2005, Dr. Lai was also awarded a D.Sc. degree by City University London, UK and he is its honorary graduate.

Currently he is Chair in Electrical Engineering and Head of Energy Systems Group at City University London, UK. He is also a Visiting Professor at Southeast University, Nanjing, China and Guest Professor at Fudan University, Shanghai, China. He has authored/co-authored over 200 technical papers. He has also written a book entitled *Intelligent System Applications in Power Engineering - Evolutionary Programming and Neural Networks*. In 2001, he edited a book entitled *Power System Restructuring and Deregulation - Trading, Performance and Information Technology*. He received a high-quality paper prize from the International Association of Desalination, USA in 1995 and the 2006 IEEE Power Engineering Society, Energy Development and Power Generation Committee Prize Paper Award. Among his professional activities are his contributions to the organization of several international conferences in power engineering and evolutionary computing, and he was the Conference Chairman of the IEEE/IEE International Conference on Power Utility Deregulation, Restructuring and Power Technologies 2000.

Professor Lai is a Fellow of the IEEE and IET, UK. He was awarded an IEE Prize in 1980; the IEEE Third Millennium Medal; 2000 IEEE Power Engineering Society UKRI Chapter Outstanding Engineer Award; 2003 IEEE Power Engineering Society Outstanding Large Chapter Award and 2006 IEEE Power Engineering Society, Technical Committee Working Group Recognition Award, for his contribution to the Working Group on New Technologies and Practical Applications, Power System Analysis, Computing and Economics Committee.

Palash Kumar Ghosh received his B.E. degree in electrical engineering from North Bengal University, in 2004 and M.Tech. degree in power and energy systems from National Institute of Technology, Silchar, India, in 2007. His reesearch interests include deregulation, wavelet transforms, soft computing techniques, and economics of electrical power systems. Now he is Reliance Energy.

Yingnan Ma received her B.Sc. degree in Industrial Automation from Shenyang University of China in 2000 and M.Sc. degree in Information Engineering from City University of London in 2003. Her research interests include energy management and intelligent weather forecast. Presently, she is working towards her Ph.D.