

Load Forecasting using Fuzzy Wavelet Neural Networks

Mahdi Amina

School of Electronics & Computer Science
University of Westminster
London, United Kingdom
mahdiamina@gmail.com

Vassilis S. Kodogiannis

School of Electronics & Computer Science
University of Westminster
London, United Kingdom
kodogiv@wmin.ac.uk

Abstract—Load forecasting is an important component for power system energy management system. Precise load forecasting helps the electric utility to make unit commitment decisions, reduce spinning reserve capacity and schedule device maintenance plan properly. This paper presents the development of a novel fuzzy wavelet neural network model and validates its prediction on the short-term electric load forecasting of the Power System of the Greek Island of Crete. In the proposed scheme, a wavelet neural network has replaced the classic TSK model in the consequent part, while subtractive clustering has been applied to the definition of fuzzy rules. The forecasting error statistical results, corresponding to the minimum and maximum load time-series, indicate that the proposed load forecasting model provides more accurate forecasts, compared to conventional neural networks models.

Keywords- Load forecasting; power systems, fuzzy wavelet networks neural networks.

I. INTRODUCTION

Load forecasting is an important component for power system energy management system. Precise load forecasting helps the electric utility to make unit commitment decisions, reduce spinning reserve capacity and schedule device maintenance plan properly. Besides playing a key role in reducing the generation cost, it is also essential to the reliability of power systems. The system operators use the load forecasting result as a basis of off-line network analysis to determine if the system might be vulnerable. If so, corrective actions should be prepared, such as load shedding, power purchases and bringing peaking units on line. Since in power systems the next days' power generation must be scheduled every day, day-ahead short-term load forecasting (STLF) is a necessary daily task for power dispatch. Its accuracy affects the economic operation and reliability of the system greatly. Under-prediction of STLF leads to insufficient reserve capacity preparation and, in turn, increases the operating cost by using expensive peaking units. On the other hand, over-prediction of STLF leads to the unnecessarily large reserve capacity, which is also related to high operating cost.

In spite of the numerous literatures on STLF published since 1960s, the research work in this area is still a challenge to the electrical engineering scholars because of its high complexity. How to estimate the future load with the historical data has remained a difficulty up to now, especially for the load forecasting of holidays, days with extreme weather and other

anomalous days. With the recent development of new mathematical, data mining and artificial intelligence tools, it is potentially possible to improve the forecasting result. With the recent trend of deregulation of electricity markets, STLF has gained more importance and greater challenges [1].

The research approaches of short-term load forecasting can be mainly divided into two categories: statistical methods and computational intelligence (CI) methods. In statistical methods, equations can be obtained showing the relationship between load and its relative factors after training the historical data, while CI methods try to imitate human beings' way of thinking and reasoning to get knowledge from the past experience and forecast the future load. Regression is one of most widely used statistical techniques. For load forecasting regression methods are usually employed to model the relationship of load consumption and other factors such as weather, day type and customer class. Engle *et al.* [2] presented several regression models for the next day load forecasting. Their models incorporate deterministic influences such as holidays, stochastic influences such as average loads, and exogenous influences such as weather. Time series approach is based upon the assumption that the data has some internal structure such as autocorrelation, trend or seasonal variation. Most commonly used classical time series methods are ARMA (autoregressive moving average), ARIMA (autoregressive integrated moving average), ARMAX (autoregressive moving average with exogenous variables), and ARIMAX (autoregressive integrated moving average with exogenous variables). Fan and McDonald [3] and Cho *et al.* [4] described implementations of ARIMAX models for load forecasting.

The interest in applying neural networks to electric load forecasting began in 1990. Neural Networks (NN) have parallel and distributed processing structures. As for the STLF problem, the multilayer perceptron (MLP) network is the most widely used one. With the ability to approximate any continuous nonlinear function, the MLP network has very good mapping (forecasting) abilities. Peng *et al.* presented a search procedure for selecting the training cases for ANNs to recognize the relationship between the weather changes and load shape. It utilizes a neural network algorithm that includes a combination of linear and non-linear terms which map past load and temperature inputs to the load forecast output. The single forecast output is the total daily load [5]. Lu *et al.* used a similar feed-forward neural network incorporating the previous

load demands, day of week, hour of day and temperature information for load forecasting [6]. Papalexopoulos *et al.* proposed the inclusion of additional input variables such as a seasonal factor and a cooling/heating degree into a single neural network [7]. An alternative technique for load forecasting using recurrent high order neural networks was considered in Kariniotakis *et al.* [8]. This type of neural network is supposed, in theory, to approximate very accurately any non-linear function, with exponential error convergence to zero [9]. Based on an interpolation procedure, the method gives superb results in training; in testing, however, the results are rather disappointing. The lack of generalization can be explained by the fact that the specific method needs an enormous number of training patterns in order to create a satisfactory input/output mapping. In this work, the training data set was created from the whole year of 1994 and the testing set for the first four months of 1995. The relatively small training data set had, as a consequence, the poor performance of the recurrent high order neural network model, despite its perfect training phase.

Many attempts by researchers have been made to improve load-forecasting process in many worldwide regions. Khan and Abraham [10] used a hybrid of neural network and fuzzy logic to forecast the load in Czech Republic. They found that hybrid fuzzy neural network and radial basis function networks are the best candidates for the analysis of the load in Czech Republic. Another study by Kodogiannis and Anagnostakis [11] discussed the development of improved neural-network-based short-term electric load-forecasting models for the power system of the Greek island of Crete. The performance was evaluated through a simulation study, using metered data provided by the Greek Public Power Corporation. Their results indicated that the load-forecasting models developed provided more accurate forecasts than the conventional methods.

NN models basically use the sigmoid activation function in neurons. However, the sigmoid function is not orthogonal, and the energy of the sigmoid function is limitless, and this leads to slow convergence. Wavelet function is a waveform that has limited duration and an average value of zero. The integration of the localization properties of wavelets and the learning abilities of NN shows advantages of Wavelet neural networks (WNN) over NN in complex nonlinear system modeling. A WNN that uses wavelet functions have been proposed by researchers for solving approximation and classification problems [12]. In this paper we present a methodology that follows the time series approach. In our case a novel fuzzy wavelet neural network traces previous load patterns and predicts a load pattern using recent load data without using weather information for modeling. The developed system identifies the load model, which reflects the stochastic behavior of the hourly load demand for the island of Crete in Greece. The balance of this paper contains also a comparative study of various other prediction techniques used to develop the STLTF for the power system of Crete.

II. THE SHORT-TERM LOAD FORECASTING PROBLEM

This section is dealing with the application of the developed STLTF models on the power system of the island of Crete.

Public Power Corporation of Greece is the main provider of the electric power in Crete. Although the island has a population of around 600.000, during summer its total population is almost doubled due to the high tourist season. The island's electric system can be considered as small-scale one; hence serious damages could be suffered by small disturbances-overload. Thermoelectric power stations in two areas, Chania and Heraklion provide the necessary power. The later in Chania is composed of 4 (wind-turbine) units with total power of 202.5 MW. The former in Heraklion is composed of 12 (gas-turbine) units with total power of 192.9 MW. The objective of STLTF is to predict the 24 hourly loads of the next day. Here, a modular-constructed forecasting system is proposed, where 24 neural blocks with a single output have to be developed and trained separately to represent the 24 hourly loads respectively. The outline of the proposed architecture is illustrated in Fig. 1.

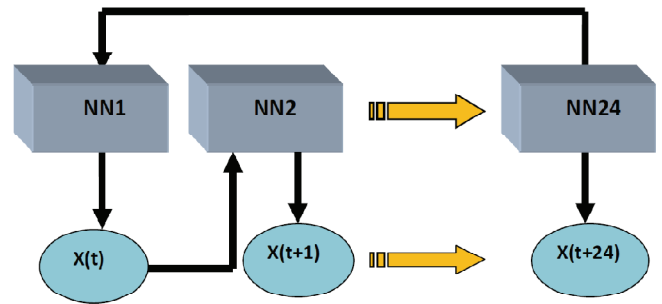


Figure 1. Proposed modular architecture for the STLTF problem

Each neural block is fed by its previous. Hence, step-by-step, 24-hour load prediction can be obtained. In this work, the training data set was created from the whole year of 1994 and the testing one for the first four months of 1995. Both training and testing sets were classified into 24 time series, each one corresponds to a different daily hour. Therefore, for each neural block, the following non-linear load model is proposed for one-hour ahead forecasting:

$$x(t) = F \{ x(t-24l), x(t-k), x(t-24l-k) \} \quad (1)$$

$$l = 1, \dots, p \quad k = 1, \dots, q$$

where p and q indicate the number of previous days and hours respectively. The main objective of the proposed system is the development of sufficiently accurate blocks representing the individual hourly loads. The investigation of a novel fuzzy wavelet neural network (FWNN) for system modelling has been carried out for the prediction of all 24h, each one trained separately, one for each hour. In the proposed scheme, a wavelet neural network has replaced the classic TSK model in the consequent part usually found in neuro-fuzzy schemes, while subtractive clustering has been applied to the definition of fuzzy rules. An assumption has been made in the case of black-outs, which occurred during the whole year. All the zero load values have been removed from both training and testing sets, and were replaced by the mean value of the preceding and subsequent load value. Eq. (1) is used to provide the input variables to the individual blocks. According to Eq. (1), the

requested load is forecast using not only the load data of previous days, but also the forecast load data for the same day at previous time steps. In the following sections, only results that correspond to hours with the maximum (14:00 h) and minimum (02:00 h) load consumption are illustrated.

III. MODELLING USING FWNN

A. Wavelet Neural Networks

Wavelet techniques can offer added insight and performance in data analysis situations where Fourier techniques have previously been used. The idea of using wavelets in neural networks has been proposed by Zhang and Benveniste [13]. The fundamental of WNN is looking for a series of appropriate wavelet basis functions in wavelet space. This process can be realised by iteration computation of wavelet basis function, which is to make the energy function minimized. To design a WNN, the main work is to design network structure, to determine the number of nodes in the middle layer, and to choose the wavelet basis function and the learning algorithm of weights training. Compared to the multilayer perceptron neural network (MLP), WNNs provide some advantages, such as the learning efficiency and the structure transparency. However, the problem of basis function networks is also led by it. Due to the crudeness of the local approximation, a large number of basis function units have to be employed to approximate a given system. In order to take advantage of the local capacity of the wavelet basis functions while not having too many hidden units, in this paper we propose an alternative type of WNN which has been influenced from related neuro-fuzzy architectures. The architecture of the proposed multiplication wavelet neural network-linear coefficients weights (LCW-MWNN) is shown in Fig. 2 [14].

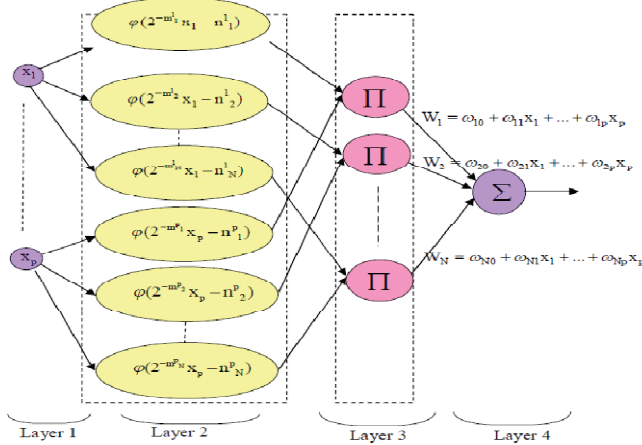


Figure 2. LCW-MWNN architecture

Referring to Fig. 2, Layer 1 accepts the input variables which are in form of $\mathbf{X}_j = [x_{j1}, x_{j2}, \dots, x_{jp}]^T$, while Layer 2 is used to calculate the wavelet “membership” values. In this layer, each node performs a membership function and acts as an element for membership degree calculation, where a wavelet

function is adopted as the membership function. Generally, a WNN approximates any desired signal $y(t)$ by generalising a linear combination of a set of daughter wavelets $\varphi_{m,n}(t)$ which are generated by step sizes dilation and translation m and n from a mother wavelet with either of the forms below:

$$\varphi_{m,n} = \varphi\left(\frac{t-n}{m}\right) \quad (2)$$

$$\varphi_{m,n} = \varphi(2^{-m}t - n)$$

where $m > 0$. The mother wavelet in the proposed Layer 2 is in the form of $\varphi_{m,n} = \varphi\left(\frac{t-n}{m}\right)$. The wavelet function adopted in this layer, is a modified differentiable version of Morlet wavelet

$$\varphi(x) = \cos(2\pi\beta x) e^{-\frac{x^2}{v}} \quad (3)$$

Therefore, the mother wavelet of Eq. 2 from the i^{th} wavelet node connection to the p^{th} input data is expressed as

$$\varphi_{m^p_i, n^p_i}(x^p_j) = 2\pi\beta\left(\frac{x^p_j - n^p_i}{m^p_i}\right) e^{-\frac{\left(\frac{x^p_j - n^p_i}{m^p_i}\right)^2}{v}} \quad (4)$$

This wavelet is derived from a function that is proportional to the cosine function and Gaussian probability density function. The nodes of Layer 3 are regarded as the “wavelet” rules in association to the fuzzy rules in a neuro-fuzzy architecture. The number of the “wavelet” membership functions for each input variable is equal to the number of “wavelet” inference rules. These units are fixed, meaning that no modifiable parameter is associated with them. The multiplicative inference (Larsen product operator) has been used, thus the output of this inference layer is given by

$$\phi_i(x_j) = \prod_{p=1}^p \varphi_{m^p_i, n^p_i}(x_j), \quad p=1,2,\dots,P \quad j=1,2,\dots,N \quad i=1,2,\dots,C \quad (5)$$

The proposed approach differs from the conventional fuzzy rule table approaches. In those models, an input space is divided into $K_1 \times K_2 \times \dots \times K_n$ fuzzy subspaces, where K_p ,

$p=1,2,\dots,n$ is the number of fuzzy subsets for the p^{th} input variable. There is a fuzzy rule for each of these subspaces. The main drawback of that approach is that the number of fuzzy rules increases exponentially with respect to the number of inputs n . The structure of the output Layer 4, has been designed in the form of a local model $f(x_1, x_2, \dots, x_p)$, identical to the output layer appeared in *Takagi-Sugeno-Kang* (TSK) defuzzification schemes. For modelling of dynamic processes, TSK schemes possess an excellent interpretation, which is superior to most, if not all, alternative approaches. Usually $f(x_1, x_2, \dots, x_p)$ is a polynomial in the input variables x_p but it can be any function as long as it can appropriately describe the output of the model within the fuzzy region specified by the antecedent of the rule. When $f(x_1, x_2, \dots, x_p)$ is a first-order polynomial, the resulting fuzzy inference system is called a

first-order Sugeno fuzzy model. Thus, the output of Layer 4 is given as

$$y = \sum_{i=1}^C (w_{i0} + w_{i1}x_1 + \dots + w_{ip}x_p) \phi_i(x) \quad (6)$$

where, the network output is calculated as a weighted sum of the outputs of the local linear models and ϕ_i is interpreted as the operating point dependent weighting factor (i.e. output of Layer 3). Low time consumption is an advantage of this system. The Sugeno-type system is usually used when we operate on numerical (crisp) data. This system can be viewed as switching regression model or mixture of local experts. The scale and translation parameters and local linear model parameters are randomly initialised at the beginning and are optimized by using the Extended Kalman Filter algorithm.

B. Fuzzy EM-based Clustering

An effective approach to the identification of complex non-linear systems is the partitioning of the available data into subsets and approximation each subset by a simple model. This aim is furnished by gathering the objects are more similar to each other in one cluster. The term “similarity” in many cases considered as a *distance norm* from a data vector to a prototype object called as centre. While most classical clustering algorithms assign each datum to exactly one cluster, thus forming a crisp partition of the given data, fuzzy clustering allows for degrees of membership, to which the transitions of the subsets are gradual rather than abrupt (soft membership). Here clustering is based on Expectation Maximization in which the similarity should be understood as the probability that a data belongs to a specific Density. The Density parameter estimation is sought by the algorithm along with partitioning of the data.

Clustering algorithms typically require the user to pre-specify the number of cluster centers and their initial locations; the locations of the cluster centers are then adapted in a way such that these can better represent a set of data points covering the range of data behavior. Fuzzy Subtractive approach is a fast, one pass algorithm for estimating the number of clusters and clusters centers in a set of data. The subtractive clustering method assumes each data point is a potential cluster centre and calculates a measure of the likelihood that each data point would define the cluster centre, based on the density of surrounding data points. For better results it is recommended that normalization of each point into a unit hyper-box would make each dimension to be identical. The algorithm starts by finding the first large cluster, and then goes to find the second, etc [15]. The algorithm is represented in the following lines:

- Selects the data point with the highest potential to be the first cluster centre. It is done by introducing a potential measure at data point x_j , defined as

$$\text{Pot}_j = \sum_{k=1}^N \exp\left(-\frac{\|x_j - x_k\|^2}{(r_a/2)^2}\right) \quad j=1, \dots, N \quad (7)$$

Where r_a is a positive constant representing a neighborhood radius and $\|x_j - x_k\|^2$ is the square of Euclidean distance between x_j and x_k . Hence, a data point will have a high density value if it has many neighbouring data points. After calculating the potential for each vector, the one with the higher potential is selected as the first cluster centre x_{c1} .

- Removes all data points in the vicinity of the first cluster centre (as determined by radii of all data points to the newly selected cluster centre), in order to determine the next data cluster and its centre location. The first cluster centre x_{c1} is chosen as the point having the largest density value Pot_{c1} . Next, the density measure of each data point x_j is revised as follows:

$$\text{Pot}_j = \text{Pot}_j - \text{Pot}_{c1} \exp\left(-\frac{\|x_j - x_{c1}\|^2}{(r_b/2)^2}\right) \quad (8)$$

Usually the r_b variable is taken to be as $1.5r_a$. The process of acquiring new cluster centre is based on potential value in relation to an acceptance threshold $\bar{\mathcal{E}}$, rejection threshold $\underline{\mathcal{E}}$, and the relative distance criterion. A data point with the potential greater than the acceptance threshold is directly accepted as a cluster centre. The acceptance of a data point with a potential between the upper and the lower thresholds depends on the relative distance equation, defined as

$$\frac{d_{\min}}{r_a} + \frac{\text{Pot}_k}{\text{Pot}_1} \geq 1 \quad (9)$$

where $d_{\min}$ is the shortest distance between the candidate cluster centre and all previously found cluster centers. If Eq. 9 is greater than 1, then the associated x will be considered as a new cluster centre.

- Iterates on this process until a sufficient number of clusters is attained. Since the algorithm is fixed and does not rely on any randomness, the results are fixed. However, we can test the effect of the four parameters, namely, acceptance ratio $\bar{\mathcal{E}}$, reject ratio $\underline{\mathcal{E}}$, cluster radius r_a and squash factor r_b . These parameters have influence on the number of clusters and error performance measures. Large values of $\bar{\mathcal{E}}$ and $\underline{\mathcal{E}}$ will result in small number of rules. Conversely small values of $\bar{\mathcal{E}}$ and $\underline{\mathcal{E}}$ will increase the number of clusters. A large value of r_a generally result in fewer clusters that lead to a coarse model. A small value of r_a can produce excessive number of clusters that may result in an over-defined system.

IV. CB-FWNN SCHEME

This paper presents a clustering based FWNN (CB-FWNN) that follows an “ANFIS structure” and its details are illustrated in Fig. 3. However significant differences against ANFIS have been considered in the proposed modeling structure. In many fuzzy and neuro-fuzzy models, the kernel of the fuzzy system

is the fuzzy knowledge base that consists of the input–output data points of the system interpreted into linguistic fuzzy rules. The consequent parts of TSK-type fuzzy IF–THEN rules are represented by either a constant or a function. Generally, the function is a linear combination of the input variables plus a constant term. TSK-type systems cannot, in general, model complex processes with desired accuracy using a certain number of rules. These systems do not have localizability; instead model the global features of the process, and their convergence is generally slow. The consequent part of these TSK-types fuzzy networks do not provide full mapping capabilities, and, in the case of modeling of complex nonlinear processes, may require a high number of rules in order to achieve the desired accuracy. Increasing the number of the rules leads to an increase in the number of neurons in the hidden layer of the network. In addition, in such TSK-schemes, the gradient decent method is usually used to train the non-linear antecedent parameters. However in the proposed structure instead of a linear function, a novel Linear Combination Multiplication Weight Wavelet Neural Network (LCW-MWNN) has been utilized. The specific model integrated two learning schemes; Weighted Least Square as well as the classic gradient descent (GD) learning algorithms. However, GD scheme is characterized by slow convergence training speed. Hence, in this study, the scale and translation parameters and local linear model parameters which have been randomly initialized at the beginning are optimized by using the Extended Kalman Filter algorithm.

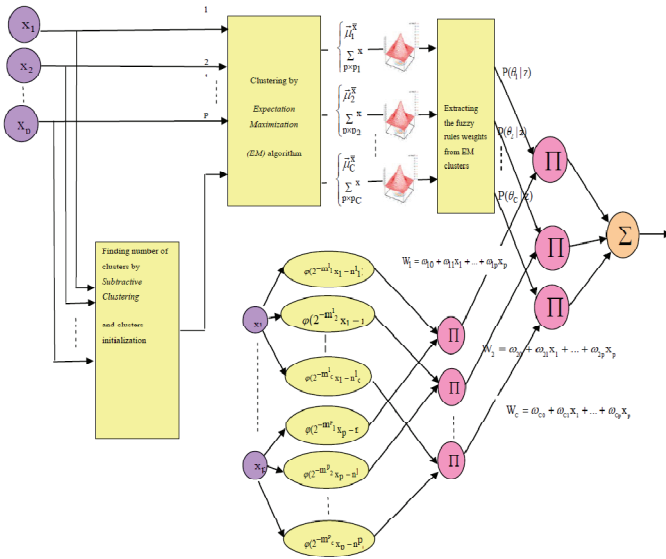


Figure 3. CB-FWNN proposed structure

In order to design the CB-FWNN structure, we assume initially, that one cluster in the input–output space corresponds to one potential fuzzy logic rule. Then the first step in structure learning is the determination of the number of fuzzy sets in the universal of discourse of each input variable [16]. In this research, the number of clusters are determined by subtractive clustering based on the training data as described

earlier and remain constant throughout. Once the number of clusters defined, then the layout of the desired CB-FWNN can be sketched. At the consequence section, different scales of wavelet-neurons $\phi_{x_p}(m_i^p, n_i^p)$ assigned for every dimension of the input. Although the number of different scales/translation allocated to each dimension is fixed, each of inputs can be associated with different scales/translations. Knowing the optimum number of clusters, the number of different scales which each input dimension is going to be decomposed then can be determined, while the known number of clusters could assist in the determination of the number of fuzzy rules C as well as to the number of unknown parameters at the antecedent section. CB-FWNN's output is formulated as,

$$\begin{aligned}\hat{y}_j &= \sum_{i=1}^C \gamma_i(X_j) W_i(X_j) \phi_i(X_j) \\ &= \sum_{i=1}^C \gamma_i(X_j) W_i(X_j) \prod_{p=1}^P \phi_{m_i^p, n_i^p}(x_j) \\ &= \sum_{i=1}^C W_i(X_j) \gamma_i(X_j) \left(\prod_{p=1}^P \cos(2\pi\beta \left(\frac{x_j^p - n_i^p}{m_i^p} \right)) e^{-\frac{(x_j^p - n_i^p)^2}{m_i^p}} \right)\end{aligned}\quad (10)$$

The conditional density $p(y | x)$ is a mixture of Gaussians functions at the output of clustering sub-unit therefore the regression problem is re-formulated also as

$$\begin{aligned}\hat{y}_j &= f(x) = E[y_j | x_j] \\ &= \int y_j p(y_j | x_j) dy_j = \frac{\int y_j p(x_j, y_j) dy_j}{p(x_j)} \\ &= \frac{\sum_{i=1}^C [W_i \phi_i(x_j)] p(x_j | \theta_i) p(\theta_i)}{p(x_j)} = \sum_{i=1}^C W_i \times P(\theta_i | x_j) [\phi_i(x_j)]\end{aligned}\quad (11)$$

where $\phi_i(x_j)$ is related to the i^{th} from C LCW-MWNN outputs, and the j^{th} input vector. $W_i(X_j)$ is the linear combination weight and $P(\theta_i | X_j)$ is the probability that the i^{th} Gaussian component is generated by the input vector X_j as depicted in the following equation, which replaces γ in eq. 10.

$$P(\theta_i | x_j) = \frac{P(\theta_i)}{\sum_{i=1}^C \frac{P(\theta_i)}{(2\pi)^{p/2} \sqrt{|\Sigma_i|}} \exp(-\frac{1}{2} (x_j - \mu_i)^T (\Sigma_i)^{-1} (x_j - \mu_i))} \exp(-\frac{1}{2} (x_j - \mu_i)^T (\Sigma_i)^{-1} (x_j - \mu_i)) \quad (12)$$

The final output based on dynamics of the proposed structure is derived by merging Eqs. 10 and 11 and have the following form

$$\hat{y}_j = \sum_{i=1}^C W_i(X_j) \times P(\theta_i | X_j) \left(\prod_{p=1}^P \cos(2\pi\beta \left(\frac{x_j^p - n_i^p}{m_i^p} \right)) e^{-\frac{(x_j^p - n_i^p)^2}{m_i^p}} \right) \quad (13)$$

The cluster parameters in this proposed scheme are estimated with the use of Expectation Maximization algorithm, while the Wavelet Neural Network parameters have been modified by Extended Kalman Filter and Weighted Least Square. The

choice of EM algorithm is justified by the fact that this approach avoids numerical instabilities encountered in Gradient Decent algorithm and presents an improved learning convergence.

V. RESULTS

This section presents the results and the statistics of forecasts obtained from the application of the developed STLF models on the power system of the island of Crete. Case studies for the proposed methods were carried out for a 24-hour load forecasting. The complete results for the STLF problem, for the hours with minimum and maximum load consumption, are illustrated in Table 1. Training has been conducted using power load data for 1994 (365 data points), while testing has been evaluated using data from the 4 first months of 1995.

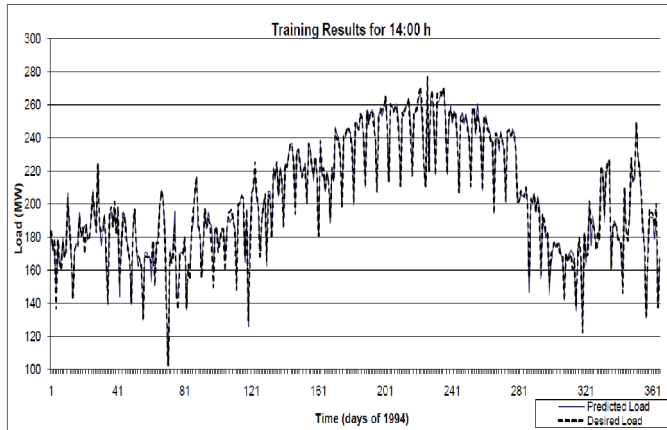


Figure 4. Training performance for max load

An obvious advantage of the proposed modular architecture is that, since the complete system consists of 24 neural blocks, each one with a single output, training is easier and faster compared to traditional neural approaches, which treat the output as a 24x1 vector. After many trials, it has been found that only two previous time load parameters are necessary for the proposed CB-FWNN to achieve an acceptable performance.

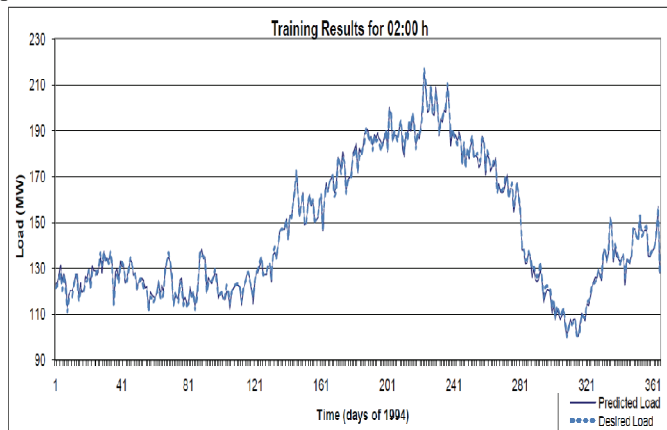


Figure 5. Training performance for min load

Fig. 4 and 5 illustrate the training performances on both minimum and maximum power consumption cases. Table I summarizes the various performance indices. The regression coefficient (R^2) is often used as an overall measure of the prediction attained. It measures the fraction of the variation about the mean that is explained by a model. The higher the value ($0 \leq R^2 \leq 1$), the better is the prediction by the model. The FWNN scheme developed herein was found to yield high level agreement with experimental observations for the test data set. The values of the coefficient of determination (R^2), as shown in Table I, indicate a very good fit of the experimental data from the FWNN-based approach.

TABLE I PERFORMANCE INDICES		
Statistical index	Testing Data sets	
	14:00	02:00
Coefficient of determination (R^2)	0.9804	0.9654
Root mean square error (RMSE)	4.9492	2.6664
Mean absolute relative error (MARE) (%)	2.0481	1.3029
Mean Square Error	24.4943	7.1099
Standard error of prediction (SEP) (%)	2.7021	2.0324
Bias factor (B_f)	1.0089	1.0052
Accuracy factor (A_f)	1.0205	1.0131
Deviation	4.9492	2.6664

However, R^2 is a suitable criterion for model comparison on the assumption that the error is normally distributed and not dependent on the mean value. In fact, the distribution of the error is not clearly known, so this term must be used with caution, particularly in non-linear regression models and hence additional indices must be employed for model comparison.

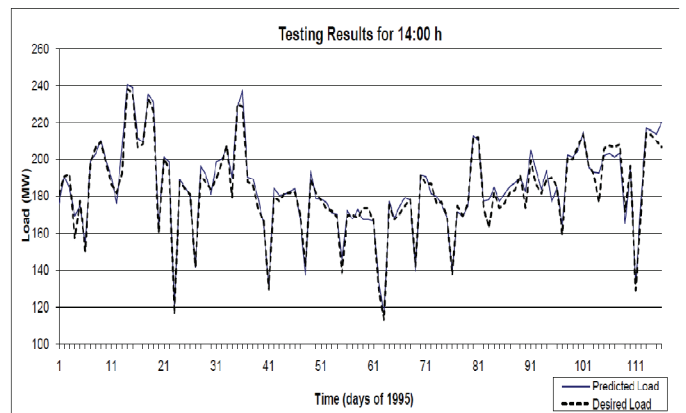


Figure 6. Testing performance for max load

The RMSE index is calculated between the desired and output values and then averaged across all data and it can be used as an estimation of the goodness of fit of the models. It can also provide information about how consistent the model would be in the long run. The MARE term provides information on how close forecasts or predictions are to the eventual outcomes. The SEP index is determined as the relative deviation of the

mean prediction values and it has the advantage of being independent on the magnitude of the measurements. Based on this index, the FWNN scheme achieved again a very good performance. The usual measures of goodness-of-fit for model comparison in many applications is performed by calculating in addition to squared correlation coefficient (R^2) the bias (B_f) and accuracy (A_f) indices [17]. Models describing predictions with B_f in the range of $\{0.9 - 1.05\}$ could be considered good, in the range of $\{0.7 - 0.9\} || \{1.06 - 1.15\}$ are considered acceptable, while for $\{< 0.7 || > 1.15\}$ are considered unacceptable. Bias factor is a multiplicative factor that compares model predictions and is used to determine whether the model over- or under-predicts the power consumption. A B_f greater than 1.0 indicates that a growth model is fail-dangerous.

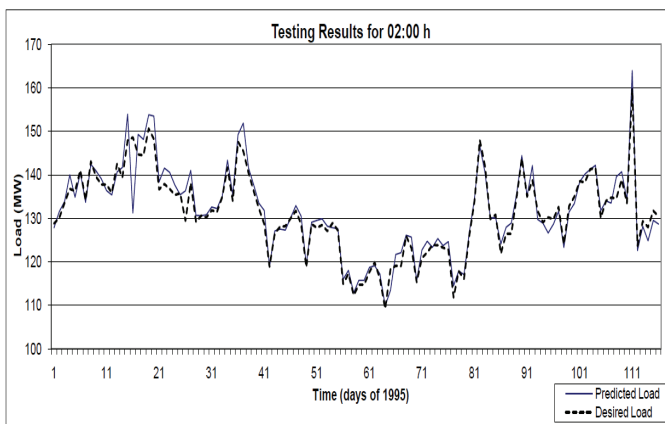


Figure 7. Testing performance for min load

Conversely, a B_f less than 1.0 generally indicates that a growth model is fail-safe, i.e. observed generation times were larger than predicted values, so that predicted values give a margin of safety. Perfect agreement between predictions and observations would lead to B_f of 1. The accuracy factor (A_f), is a simple multiplicative factor indicates the spread of results about the prediction. A value of 1 indicates that there is perfect agreement between all the predicted and measured values. Table I also shows the bias and accuracy factor values obtained for the two testing cases. Figs. 6 and 7 illustrate the testing performances for both cases. In order to evaluate the goodness of the current performance of the proposed FWNN scheme, a comparison against traditional neural networks using the same datasets has been made [11]. For the maximum load consumption, the MARE indices for MLP and RBF networks were 11.967 and 2.792 respectively, while for the minimum load case, the relevant values were 2.734 and 1.35. This comparison reveals the superiority of the proposed CB-FWNN scheme.

VI. CONCLUSIONS

This paper presented the development of a novel fuzzy wavelet neural network model and validated its prediction on the short-term electric load forecasting of the Power System of the Greek Island of Crete. In the proposed scheme, a wavelet neural network has replaced the classic TSK model in the consequent part, while subtractive clustering has been applied to the definition of fuzzy rules. The forecasting error statistical results, corresponding to the minimum and maximum load time-series, indicate that the proposed load forecasting model provides more accurate forecasts, compared to conventional neural networks models such as MLP and RBF. It should be noted that the proposed structures differ from the usual time-series prediction models since they include only loads for both previous times and previous days for hourly load forecasting. In a future work, the present approach will be enhanced by using additional load and weather information such as illuminations level, temperature, humidity, wind direction and industrial load.

REFERENCES

- [1] F. Gao, X. Guan, X.-R. Cao, A., Papalexopoulos, "Forecasting power market clearing price and quantity using neural network method", Power Engineering Society Summer Meeting, Vol. 4, 2000, pp. 2183-2188.
- [2] R.F. Engle, C. Mustafa and J. Rice, "Modelling Peak Electricity Demand", Journal of Forecasting, vol. 11, no. 3, 1992, pp. 241-251.
- [3] J.Y. Fan and J.D. McDonald, "A Real-Time Implementation of Short-Term Load Forecasting for Distribution Power Systems", IEEE Transactions on Power Systems, Vol. 9, 1994, pp. 988-994
- [4] M.Y. Cho, J.C. Hwang and C.S. Chen, "Customer Short-Term Load Forecasting by using ARIMA Transfer Function Model", Proceedings of the International Conference on Energy Management and Power Delivery, Vol. 1, No. 1, 1995, pp.317-322.
- [5] M. Peng, N.F. Hubele and G.G. Karady, "Advancement in the Application of Neural Networks for Short-Term Load Forecasting", IEEE Transactions on Power Systems, Vol. 7, No. 1, 1992, pp. 250-257.
- [6] Lu, C.N., Wu, H.T., Vemuri, S., "Neural network based short term load forecasting", IEEE Transactions on Power Systems, Vol. 8, No. 1, 1993, pp. 337-342.
- [7] Papalexopoulos, A., How, S., Peng, T.M., "An implementation of a neural network based load forecasting model for the EMS", IEEE/PES 1994 Winter Meeting, 1994, pp. 207-209.
- [8] Kariniotakis, G.N., Kosmatopoulos, E., Stavrakakis, G., "Load forecasting using dynamic high-order neural networks", Proceedings of IEEE-NTUA Joint International Power Conference, 1993, pp. 801-805.
- [9] Kosmatopoulos, E., Polycarpou, M., Christodoulou, M., Ioannou, P., "High-order neural network structures for identification of dynamical systems", IEEE Transactions on Neural Networks, Vol. 6, No. 2, 1995, pp. 422-432.
- [10] M.R. Khan, A. Abraham, "Short term load forecasting models in Czech Republic using soft computing paradigms", International Journal of Knowledge-Based Intelligent Engineering Systems, Vol. 7, No. 4, 2003, pp. 172-179.
- [11] V.S. Kodogiannis, E.M. Anagnostakis, "A study of advanced learning algorithms for short-term load forecasting", Engineering Applications of Artificial Intelligence Journal, Vol. 44, No. 9, 1999, pp. 159-173.
- [12] Y.C. Pati, P.S. Krishnaprasad, "Analysis and synthesis of feed-forward neural networks using discrete affine wavelet transformations", IEEE Transactions on Neural Networks, Vol. 4, 1993, pp. 73-85.
- [13] Q. Zhang, A. Benveniste, "Wavelet Networks", IEEE Trans. Neural Networks, Vol. 3, No. 6, 1992, pp. 89-98.
- [14] M. Amina, E.Z. Panagou, V.S. Kodogiannis, G.-J.E. Nychas, "Wavelet Neural Networks for modelling high pressure inactivation kinetics of *Listeria monocytogenes* in UHT whole milk", Chemometrics and

- intelligent laboratory systems, Elsevier, Vol. 103, No. 2, 2010, pp. 170-183.
- [15] A. Priyono, M. Ridwan, A. J. Alias, R. Atiq, K. Rahmat, A. Hassan, M. Ali, and M. Alauddin, "Generation Of Fuzzy Rules With Subtractive Clustering", *Jurnal Teknologi (D)*, Vol. 43, 2005, pp. 143-153.
- [16] C.-J. Lin and C.-H. Chen, "Identification and prediction using recurrent compensatory neuro-fuzzy systems", *Fuzzy Sets and Systems*, Vol. 150, 2005, pp. 307-330
- [17] T. Ross, "Indices for performance evaluation of predictive models in food microbiology", *J Appl Microbiol*, Vol. 81, 1996, pp. 501-508.