AN ADAPTIVE NEURAL NETWORK APPROACH TO ONE-WEEK AHEAD LOAD FORECASTING

T. M. Peng
non-member

N. F. Hubele G. G. Karady
member fellow

Pacific Gas & Electric Company
San Francisco, CA 94106

Arizona State University
Tempe, AZ 85287-5706

Abstract A new neural network approach is applied to one-week ahead load forecasting. This approach uses a linear adaptive neuron or adaptive linear combiner called "Adaline." An energy spectrum is used to analyze the periodic components in a load sequence. The load sequence mainly consists of three components: base load component, and low and high frequency load components. Each load component has a unique frequency range. Load decomposition is made for the load sequence using digital filters with different passband frequencies After load decomposition, each load component can be forecasted by an Adaline. Each Adaline has an input sequence, an output sequence, and a desired response-signal sequence. It also has a set of adjustable parameters called the weight vector. In load forecasting, the weight vector is designed to make the output sequence, the forecasted load, follow the actual load sequence; it also has a minimized Least Mean Square error. This approach is useful in forecasting unit scheduling commitments. Mean absolute percentage errors of less than 3.4 percent are derived from five months of utility data, thus demonstrating the high degree of accuracy that can be obtained without dependence on weather forecasts.

Keyword: adaptive signal processing, neural network, adaline, spectrum analysis, load decomposition.

1. INTRODUCTION

System load forecasting is an essential function in any power system control center. The accuracy of the forecasted load influences decision-making in unit commitment, hydro-thermal coordination, fuel allocation, and off-line network analysis. A great saving potential exists for electric utilities. A lot of research has addressed accurate and efficient load forecasting methods. Many approaches have been applied to this area, including linear regression, exponential smoothing, stochastic process, state space methods, and expert systems [1-9]. Some significant issues remain in this field. It is difficult to model the relationships between the loads and the variables that influence the loads, such as weather and holiday activity. These factors complicate the modeling process. Another difficulty lies in estimating and adjusting the model parameters. These parameters are estimated from the historical data, which may become obsolete or may not reflect short-term load pattern changes. Besides, these models highly depend on the particular utility environments; it is not easy to transfer them from one company to another.

Recent progress in the application of neural network to load forecasting [10-15] provides a potential technique to deal with the above difficulties. This is because of two key features of the neural networks. First, the neural network does not rely on the explicitly expressed relationship between input variables and forecasted load. When using neural networks for load forecasting, one need only consider the selection of variables as the network input variables. The relationship between the input variables and the predicted load will be formulated by a training process. This avoids difficulties in the modeling process. The adaptive algorithm is another appealing feature of neural networks. New training cases can be selected and system parameters estimated each time a new forecast is needed. Park et al. [10] demonstrated the usefulness of recent load observations as training cases. A completely different approach to training case selection was presented by Peng et al. [14] and illustrated the use of minimum-distance criteria for selecting the most appropriate training cases. While all these prior works focus on the application of back-propagation networks to update and retrain the system, Peng et al. [14] explored the effect of using a different network structure.

We continue to investigate alternate forms of neural networks and training case selection. This approach does not explicitly rely on training cases but rather on model development that more closely resembles the early work started by Vemuri, Hill and Balasubramanian [7] in the application of autoregressive moving-average Box and Jenkins times series models to load forecasting. The parameters of this model are updated using an adaptive neural network consisting of linear adaptive neurons called an Adaline (ADAptive LINear Neuron). This network structure is based on the early work of B. Widrow in the 1960s [16, 17] and has been widely applied in neural networks, signal processing and many other areas [18, 19].

Our objective is to study the potential of using this alternate neural network. Until now, the most current results of neural networks perform about as well as any of the prior methods. The appealing features of neural networks may present a slight advantage over other methods, but the newness of the methodology may minimize the advantages to the user community. In this paper a familiar model structure based on the Box and Jenkins models is incorporated into a neural network framework. Instead of off-line estimation, Adalines are used to update the model parameters. Simulation results indicate that one-week ahead hourly forecasts can be generated with mean absolute percentage errors of less than 3.4 percent. Furthermore, this model performance is not dependent on forecasted temperature inputs or the errors contained therein. These results suggest that this direction of using Adalines for model development is worthy of further investigation.

2. OVERVIEW OF THE PROPOSED METHOD

In this approach, we consider a series of load data, obtained by sequentially sampling, as a signal sequence. This load sequence is comprised of several components having different periods and magnitudes. The usual method of analyzing different frequency components is by a power spectrum analysis, in which the signal energy corresponding to each frequency is calculated. Through spectrum analysis, a load sequence displays three main components: base load component, and low and high frequency load components. Each load component has a unique frequency range; it can be extracted from the original load sequence by a digital filter with a corresponding passband frequency.

After load decomposition, each load component can be forecasted by an Adaline. An Adaline can operate as a linear predictor which has an input sequence, an output sequence, and a desired response signal sequence. It also has a set a adjustable parameters called a weight vector. When used for prediction, the weight vector is designed to make the output sequence, produced from the present and previous signal values, follow the desired sequence and have a minimized Least Mean Square error.

Additional modifications to the forecasts are generated using a residual model based on the periodic components of dry-bulb temperatures. This subsystem model is also based on the Adaline linear predictor structure. A total of five Adalines are used to predict each hour of a 24-hour load curve, one-week ahead.

92 SM 407-7 PWRS A paper recommended and approved by the IEEE Power System Engineering Committee of the IEEE Power Engineering Society for presentation at the IEEE/PES 1992 Summer Meeting, Seattle, WA, July 12-16, 1992. Manuscript submitted January 21, 1992; made available for printing April 6, 1992.

Many researchers have followed the direction of Box and Jenkins models for load forecasting started by Vemuri [7]. Typically, the autocorrelation function is used to identify the model. In this paper we use the spectral density, which is merely the Fourier transform of the autocorrelation function. The most significant difference between our work and that of prior researchers is the incorporation of the weather-sensitive component. In prior works, transfer functions mapping load to weather readings have been used. In this proposed model, a spectral decomposition of the residuals of the load model and dry-bulb temperature readings is used to identify a residual model. This residual model represents a novel approach and alternative to the classical transfer function models proposed by prior researchers.

This paper presents a basic description of the adaptive linear neuron and the signal spectral decomposition. Then, a more detailed description and simulation results for the proposed approach are addressed.

3. POWER SPECTRUM ANALYSIS AND LOAD DECOMPOSITION

This approach decomposes a load sequence into three different frequency components. A load forecast is made by predicting each load component. The following procedure is used to perform this task:

- Identifying different frequency components by power spectrum analysis;
- Filtering load components by the use of digital filters;
- Forecasting each load component;
- Forecasting residual and temperature-sensitive load.

3.1. Power Spectrum Analysis

The concept of the spectral decomposition of a load results in three components: a long-term trend, a component that varies with the day of the week, and a random component. In our method, a spectral decomposition concept, based on power spectrum analysis is used. In the power spectrum analysis, a Fast Fourier Transformation (FFT) was taken of a series of load-sample data, then the signal energy corresponding to each frequency component was calculated.

In Figure 1, a 400-hour load sequence for Wednesdays is presented. This load sequence contains a large DC component. For the spectrum analysis, a constant load (3500 MW) was subtracted from the initial load. The resulting power spectrum is given in Figure 2.

In Figure 2, the scaled sampling frequency fs=1.0 corresponds to half of the sample rate. Its period, Ts=2 hours, is twice of the sampling period T=1 hour. The time-domain period of any frequency component is found as follows. For example, there is a large component at $f = 0.0833$ fs. Its period is $T = 1/f = 1/0.0833fs = Ts/0.0833 = 24$ hours. In the same way, when $f = 0.166$, its time period is 12 hours.

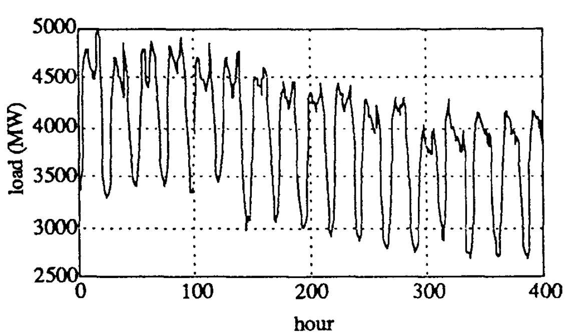

Figure 1. Hourly Load for Sixteen Wednesdays

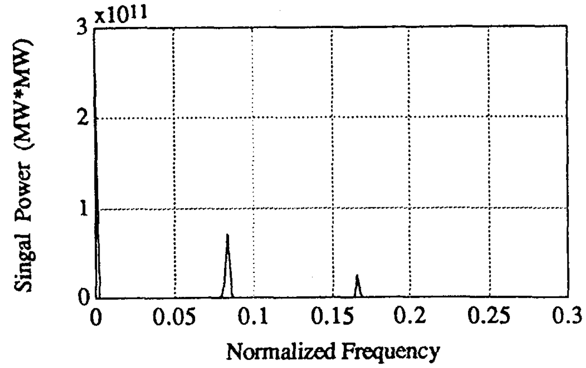

Figure 2. Power Spectrum of the Load Sequence in Figure 1.

The following observations can be made from Figure 2.

- The load signal sequence contains a large, slowly changing component. In the spectrum, this component appears in the frequency range from f=0 to f=0.015. We define this component as the base load component and it represents the seasonality in the daily shapes. This component can be useful in capturing the changes occurring between seasons, reflected from week to week.
- A large periodic component with a period of T=24 hours (f=0.083) can be observed. We define this component as a low-frequency component that represents the similarity of load shapes from week to week.
- A load component with a period of T=12 hours (f=0.166) can also be observed. We define this component as a high-frequency component that represents the relationship of the levels of hourly load within a given day's curve.
- Compared with the above components, other load components, like the noise component, are much smaller. Therefore, these components are not considered in load decomposition. However, when making load forecasts, its influence will be considered as a residual component. A detailed discussion will be given in Section 5.3.

In the spectrum, each load component can be easily distinguished from the others because it has a unique frequency range. This factor suggests that digital filters could extract each load component from the load sequence. For example, the base load component can be obtained using a low-pass filter with passband frequency f<0.02; the low-frequency component can be obtained by a bandpass filter with passband frequency 0.07<f<0.1; the high-frequency component can be obtained by a high-pass filter with a passband frequency f>0.15.

3.2 Load Decomposition

As discussed above, the load can be decomposed into three main frequency components in the following way

$$x(t) = x'(t) + x''(t) + x'''(t), \tag{1}$$

where x'(t) is the base load component, x''(t) is the low-frequency component, and x'''(t) is the high-frequency component. The load components with different frequencies can be obtained using digital filters with different critical passband and stopband frequencies. In our studies, Butterworth filters and Chebyshev I filters [19] have been investigated. The use of the different type of filters made only slight difference on load-forecasting accuracies.

Frequency-selective filters can pass signals undistorted in one frequency band and attenuate or totally eliminate signals in the remaining frequency bands. This feature could be used to decompose different frequency components in a load sequence. In the time domain, the input-output relationship of a digital filter can be described by

$$y(k) = a_0\, x(k) + a_1\, x(k\text{-}1) + \cdots + a_n\, x(k\text{-}n) - b_1\, y(k\text{-}1) - b_2\, y(k\text{-}2) - \cdots - b_n\, y(k\text{-}n)' \tag{2}$$

where n is the order of the filter, a_i (i=0,...,n), b_j (j=1,...,n) are the constant coefficients of the filter; and k is the hourly time index. This is also known as a Box-Jenkins ARMA(n,n) model. This relationship can also be described by a frequency domain transfer function form given by

$$Y(z) = H(z)\, X(z), \tag{3}$$

where z is the complex frequency in Z-transform form; Y(z) and X(z) are z-transform of y(k) and x(k) respectively; H(z) is the transfer function of the digital filter in (2) given by

$$H(z) = \frac{\sum_{i=0}^{n} a_i\, z^{-i}}{1 + \sum_{j=1}^{n} b_j\, z^{-i}}. \tag{4}$$

The transform function of a filter is the function of a complex frequency variable z. The frequency response of the filter can be used to decompose a load sequence. To obtain a load component, e.g., a base load component that has a frequency range f < 0.02, the parameters of the filter, a_i and b_j (i=0,1,...,n; j=1,2,...,n), can be designed so that for f < 0.02, the magnitude of H(z), denoted by |H(z)|, approximates a unit. While for the other components, with a frequency range of f>0.07, |H(z)| approximates zero. The following procedure is used in the decomposition of a load sequence as given in Figure 1.

- The base load component, x'(t), can be obtained by a Chebyshev I low-pass filter with the passband frequency f= 0.03. Figure 3 shows the base load component compared with the original load sequence.

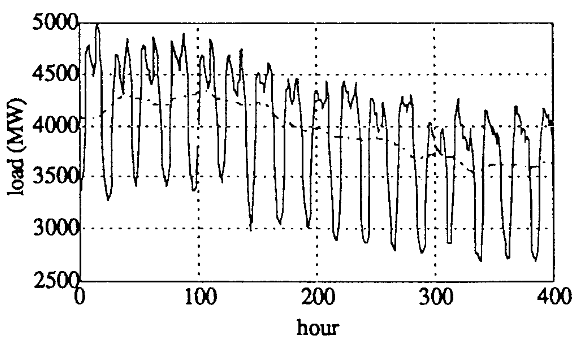

Figure 3. Base Load Component and Load Sequence

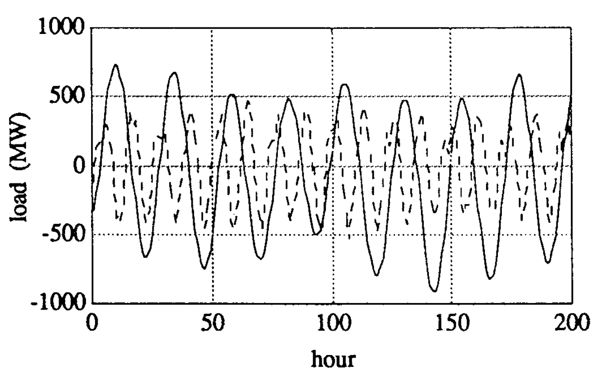

Figure 4. Low and High Frequency Load Components

- The high-frequency load component, x'''(t), can be obtained by a high-pass filter with passband frequency of f > 0.15.
- The low-frequency component can be obtained by x''(t) = x(t) - x'(t) - x'''(t).

Figure 4 shows the low-frequency and high-frequency load components filtered from the load sequence.

4. ADALINE

Adaline, proposed by Widrow [16, 17], has a structure shown in Figure 5. In the figure, k is the sampling time index k=0,1,... ; x(k) is the input signal sequence; y(k) is the output sequence; and d(k) is the desired response sequence. The output of Adaline can be expressed as a linear combination of the input signal

$$y(k) = \sum_{i=0}^{L} w(i)\, x(k\text{-}i) = X^T_k W \tag{5}$$

where L is the order of the adaptive linear combiner; superscript ^{T} means transpose of a matrix or vector; $W=[w(0),w(1), \ldots , w(L)]^T$ is a set of adjustable weight vector of the combiner; $X_k = [x(k),x(k\text{-}1), \ldots , x(k\text{-}L)]^T$ is input signal vector at time k.

To forecast a signal sequence y(k) with an Adaline, the desired signal sequence, d(k), can be set to x(k + lead time) while y(k) is the forecasted signal value at (k + lead time). The forecasting error is defined as the difference between the forecasted signal values y(k) and the actual signal d(k) = x(k + lead time). The error is given by

$$e(k) = d(k) - y(k) = d(k) - X^T_k W \tag{6}$$

The instantaneous squared error is defined as

$$e^2(k) = d^2(k) - W^T X_k X^T_k W - 2d(k)\, X^T_k W. \tag{7}$$

Forecasting performance is measured by Mean-Square-Error (MSE), which is defined as the expected value of the instantaneous square error given in (6). This MSE can be expressed as

$$MSE = \xi = E[e^2(k)] = E[d^2(k)] - W^T R W - 2P^T W, \tag{8}$$

where R is defined as the "input correlation matrix", and P is defined as the cross-correlation of the known desired response and the input components. These matrices are given by

$$R = E[X^T_k X_k]\,, \quad P = E[d_k X^T_k]. \tag{9}$$

The equation (8) is a quadratic function of the components of the weight vector W. There is a single globe point in which W=W* and MSE is minimized.

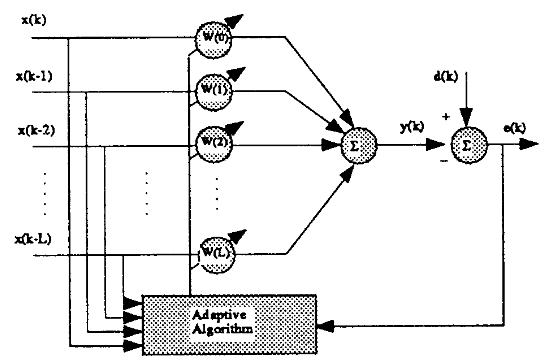

Figure 5. An Adaptive Linear Neuron

If the input correlation matrix is nonsingular, the optimum solution of the weight vector W*, Winer solution, can be obtained by the Widrow-Hoff equation

$$W^* = R^{-1} P \tag{10}$$

In general, the optimal weight vector can also be obtained by using an iterative process in which

$$W_{k+1} = W_k + \mu \nabla_k \tag{11}$$

where μ is a constant learning step length, and ∇_k is the gradient vector of MSE with respect to the W at the *k* th iteration

$$\nabla_k = [\frac{\partial\xi}{\partial W(0)}, \frac{\partial\xi}{\partial W(1)}, \ldots, \frac{\partial\xi}{\partial W(L)}]^T \tag{12}$$

One important iterative algorithm used in this application is called Least-Mean-Square (LMS) algorithm, or Widrow learning rules. Here the gradient vector is estimated by taking $e^2(k)$ as an estimate of ξ as follows

$$\nabla_k = \frac{d\xi_k}{dW} \cong \frac{de^2(k)}{dW} = -2e(k) X_k \tag{13}$$

With this estimated gradient, the weight vector is solved by an iterative process given by

$$W_{k+1} = W_k - \mu \nabla_k = W_k + 2\mu e(k) X_k \tag{14}$$

The iterative process will converge to the optimum W* if $0 < \mu < 1/\lambda_{max}$, where λ_{max} is the largest eigenvalue of the matrix R.

The forecasting error generated by this model can be used to improve the accuracy of the final forecast. This error is based on the model's most recent performance. For our purposes, the time index is denoted in hours. However, hour 25 is separated from hour 24 by an entire week, since each model corresponds to a particular day. Furthermore, the error depicted in Figure 5 is not known at the time of the forecast. Therefore, to incorporate the error corresponding to the same hour of the previous week for time index *k*, the weights are updated by the error e(k-24), which is defined as d(k-24)-y(k-24). This modified Adaline structure is given in Figure 6. The corresponding modification in the algorithm is that: when using equation (13) and (14) to update weight vector, the time index k of the error term is replaced by k-24.

As an additional note, the structure in Figure 6 can be used to represent a multiple regression load forecasting model when the input vector in the figure represents a set of explanatory variables; y(k) represents a forecasted load; {W(k), k=0, ... , L} represents a set of regressive coefficients. The LMS algorithm in Equation (14) can be used to update the regressive coefficients adaptively. This approach is currently being studied to improve the existing multiple regression load forecasting model of the Pacific Gas & Electric Company [5]. Preliminary results indicate that the proposed model has a better performance than the original one.

5. FORECASTING LOAD COMPONENTS

5.1. A Network for Load Forecasting

In the previous section, a load sequence was decomposed into three load components. A load can also be forecasted by making forecasting for each of the components. Figure 7 shows an adaptive neural network used to perform this task that has a typical neural network structure. In the figure, the input load signal x(k) is decomposed into three components x'(k), x"(k), and x'"(k) by the digital filters H1, H2, and H3, respectively. These decomposed signal sequences reach the hidden neurons which are a set of Adalines. The outputs of the hidden neurons are the forecasted load components y'(k), y"(k), and y'"(k). Here y'(k) is the forecasted base load component; y"(k) is the forecasted low-

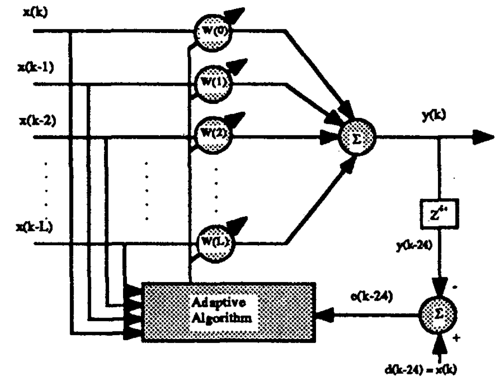

Figure 6. Modified Adaline Structure

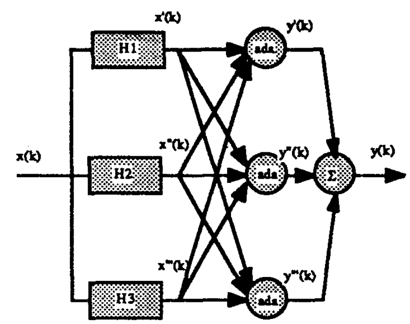

Figure 7. An Adaptive Neural Network Structure

frequency load component; y'"(k) is the forecasted high-frequency load component.

In the output neuron, a forecasted load with time index k, denoted by y(k), is obtained by summing the forecasted base load, low-frequency and high-frequency load components

$$y(k) = y'(k) + y''(k) + y'''(k) \tag{15}$$

5.2. Modified Network Structure

The network shown can be modified to address the following considerations. The influences between the different load components can be ignored because they are located in the different frequency ranges. Therefore the neural network has a decoupled structure. The temperature-sensitive load component is taken into consideration by an additional term $y_T(k)$. The forecasting residuals are also predicted by an adaptive neuron.

This modified neural network is shown in Figure 8. Both residual and temperature-sensitive load components will be discussed in the next section.

To make the system follow the changes in load pattern, the weight vectors of Adalines must be adjusted when the next forecasting error is known. The LMS algorithm in (14) is used to update the weight vectors. This is a learning process. However, it is much simpler and computationally efficient than learning by back-propagation algorithm in feedforward neural networks [11-13]. In the LMS learning, the weights will be updated each hour by (14) rather than be calculated again.

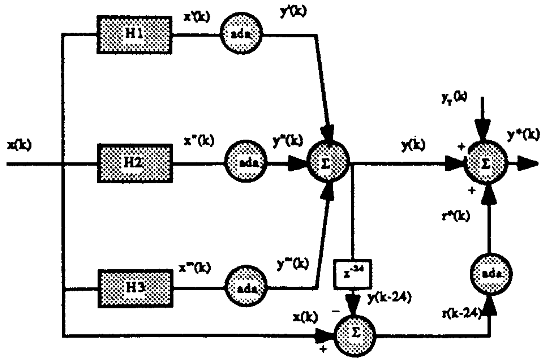

Figure 8. A Modified Adaptive Neural Network

5.3. Residual and Temperature-sensitive Load Components

A forecasting residual sequence can be obtained by comparing the load forecasted in time k-24, y(k-24), with the desired load in k-24, d(k-24). In Figure 8, this residual sequence can be expressed as

$$r(k-24) = d(k-24) - y(k-24) \qquad (16)$$

One way to improve the forecast expressed by equation (15) is to forecast the residual.

An Adaline is used to forecast residuals. In the Adaline, r(k-24) is the input sequence; r*(k), the forecasted residual for r(k), is the output sequence. When considering this additional residual component, a forecasting error sequence can be calculated by

$$e(k) = d(k)-y(k)-r^*(k) = d(k)-y'(k)-y''(k)-y'''(k)-r^*(k) \qquad (17)$$

Because the influence of temperature on load is not considered directly in (17), the error sequence is assumed to include a temperature-sensitive component which is denoted by $y_T(k)$. In this way, an error sequence can be expressed as

$$e(k) = y_T(k) + n(k) \qquad (18)$$

where n(k) is a zero-mean noise with unknown variance.

To identify the temperature-sensitive load component from sequence (18), the relationship between the load and temperature must be modelled. Most methods try to determine the relationship in time-domain; modelling the load change as a function of weather variables. This modelling approach is highly dependent on the utility's service environments. In multiple-days ahead load forecasting, the accuracy of forecasting will be influenced by the accuracy in weather prediction.

In our approach, the influence of temperature on load is addressed by considering its frequency feature. The frequency-domain relationship between load and temperature can be represented as

$$Y_T(f) = H_T(f)\, T(f), \qquad (19)$$

where T(f) is the frequence-domain representation of temperature sequence; $H_T(f)$ is a transfer function. The time-domain relationship of Equation (19) is described by a differential equation having a form

$$\frac{d^{(n)}y_T(t)}{d\,t^{(n)}} + a_{n-1}\frac{d^{(n-1)}y_T(t)}{d\,t^{(n-1)}} + \dots + a_0 y_T(t) = b_0 T(t) + b_1\frac{dT(t)}{dt} + \dots + b_m\frac{dT^{(m)}(t)}{d\,t^{(m)}} \qquad (20)$$

A conventional way to model load-temperature variations is using a set of algebraic equations. For example, a commonly used form is given by the "bathtub" curve.

$$y_T(t) = a\,[\,T(t) - c\,]^2, \qquad (21)$$

where a is a coefficient and c is a temperature threshold value. The proposed approach is essentially a piecewise-linear form, that approximates different parts of this "bathtub" curve depending on the season.

A basic feature of the frequency response of a linear system given in Equation (19) and (20) is that the system response has the same frequency as the system input. So, in the equation (19), if the frequency range of the temperature T(f) is known, then the frequency range of $Y_T(f)$ is also known. Based on this fact, an algorithm used to forecast a temperature-sensitive load component can be proposed as follows.

First, we determine temperature frequency range by spectrum analysis. In load-forecasting algorithms, the temperature parameters are usually maximum, minimum, and average temperatures. From the spectrum analysis of the temperature sequence, the temperature has a frequency range $0 < f < 0.03$. It is a low-frequency signal when compared with the hourly load. This observation can be verified by the fact that temperature changes at a slower rate than hourly load.

Second, there is a low-frequency component in the error sequence (18) with the frequency $0 < f < 0.03$, the temperature-sensitive load component $Y_T(f)$. A low-pass digital filter can be designed to extract this component from the error sequence. Finally, an adaptive neuron can be used to forecast $y_T(f)$, as forecasting base load, low-frequency and high-frequency load components. This is the fifth Adaline. The forecasted load, y*(k), is the sum of the forecasted load component (or autoregressive term in Box-Jenkins terminology), the forecasted residual and the temperature frequency component (analogous to the moving-average term of Box-Jenkins).

One aspect of this modeling approach is to estimate a temperature-sensitive load component from the error signal. In the traditional approaches, however, that component is predetermined by constructing a functional form between the temperatures and the load. When this functional form is not properly formulated, the error sequence still contains a low-frequency, temperature-sensitive component. For example, in a heatwave, the underestimated forecasts may occur over a period of several days because the influence of the temperature on load is not properly described.

To forecast the temperature-sensitive load component, only the frequency range of temperature is used. This range is set to a constant, e.g. $0 < f < 0.03$. Once this frequency range is known, no temperature readings are required in the forecasting algorithm. When there exists other low frequency weather variables, such as humidity or cloud cover, a check should be made to determine whether the frequency range is the same as the range of the temperature. If so, there is no need to alter the algorithm. However, if a different range is found for these other weather varaibles, then an additional filter may be incorporated.

6. SIMULATION RESULTS AND EVALUATION

6.1. Simulations

To test the performance of this proposed method, it was used to forecast both one-week ahead hourly load and 24-hour ahead hourly load. In the simulations, the data from a winter peak utility in the Northeast of USA are used; the same data set has been used in the references [5, 14]. A mathematical software, MATLAB, was used for numerical simulations running in a MAC II computer. In the simulation, the initial values of the weights are solved by applying Equation (10) to a set of training data. The weights are then updated by the LMS algorithm in the forecasting period, not used in the initial estimation.

The measurement of forecasting accuracy is accomplished by a absolute percentage error, which is defined as

| forecasted load - actual load | * 100 / actual load.

To forecast the load for the next week, we divided the load database into seven subsets corresponding to seven daily load patterns from Monday to Sunday. Each day-type load is forecasted by a network in Figure 8. In this way, when using a load sequence, e.g. a Monday load to forecast the load in the next 24 hours, we actually forecast the load of Monday in the next week (168 hour ahead). Table I gives the performance of the method in a 5-month period from Feb. 1.

In Figure 9, the absolute percentage error cumulative distribution is presented for Tuesday load forecasting. Significantly, nearly 90 percent of the forecasts are within 5 percent of the actuals.

Table I. One-Week (168 hours) Ahead Load Forecasting Results in a 5 Month Period

Day	average absolute percentage of forecast errors	standard deviation
Monday	4.0433	3.2438
Tuesday	3.4340	2.9422
Wednesday	2.7776	2.3181
Thursday	2.7964	2.3554
Friday	3.4334	3.1630
Saturday	3.2727	2.4672
Sunday	3.9673	3.3544
Average	3.38	2.8349

This proposed method can also be used to forecast a 24-hour ahead load. We made forecasts for a continous 1000 hour period (Tuesday-Friday) load beginning in January. The result shows an absolute percentage error of 2.09. Figure 10 displays actual and forecasted load in a 250-hour interval during this period. In the figure, the solid line '—' stands for the real load while for the dashed line '---' stands for the forecasted load.

Figure 11, 12, and 13 show the actual and forecasted load components. These forecasted load components are close to the actual load components.

6.2. Evaluation

The approach has two significant benefits. The algorithm is simple. The key equations in this algorithm are the filter equation (2), the Adaline output equation (5), the output error equations (6), and the weight vector update equation (14). All of these are easily programmed in a computer. Also, the algorithm is highly efficient from a computational point of view. To quantify this, we use the notion of *flop*. A flop is a floating operation point, which may be a floating-point addition or multiplication operation. For each time k, the number of flops in the forecasting load $y^*(k)$ and updated weight vectors can be calculated.

In decomposing load $x(k)$ from equation (2), the number of flops is $2^*(2n+1)$, where n is the order of the filters. In computing $y^*(k)$, from equation (5), the number of flops is $(2L+1)$ and L is the order of the Adaline. In adjusting weight vectors from equations (6) and (14), the number of flops is $(3+2L)$. Notice that in Figure 8, there are four filters (one for temperature-sensitive load), five Adalines (one for temperature-sensitive load), and three summation nodes including 5 floating point addition operations. The total number of flops is equal to or less than $12N+20M+25$, where N is the maximum order of the filters and M is the maximum order of the Adalines. This number of flops gives the amount of computation in producing a forecast and is an exact measurement of the efficiency of the algorithm.

7. CONCLUSION

Adaline requires sequential input and output signals. This pre-condition indicates that it is particularly suitable for applications in signal processing and control areas. In this area, it performs well and is highly efficient.

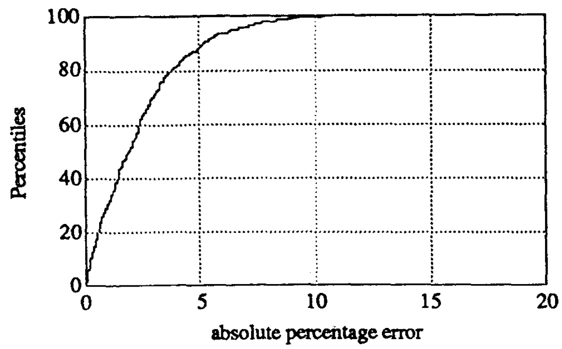

Figure 9. Forecasting Performance

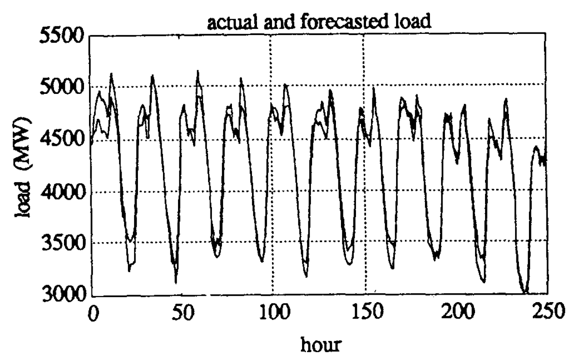

Figure 10. Actual Load and Forecasted Load

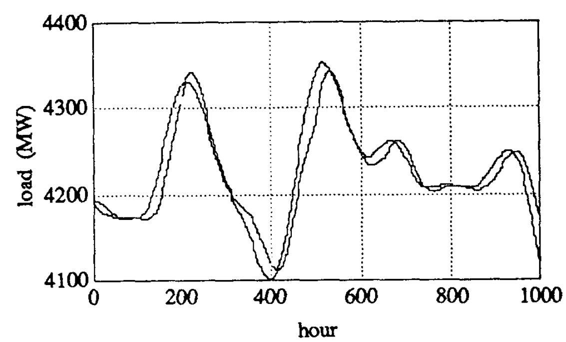

Figure 11. Actual and Forecasted Base Load Components

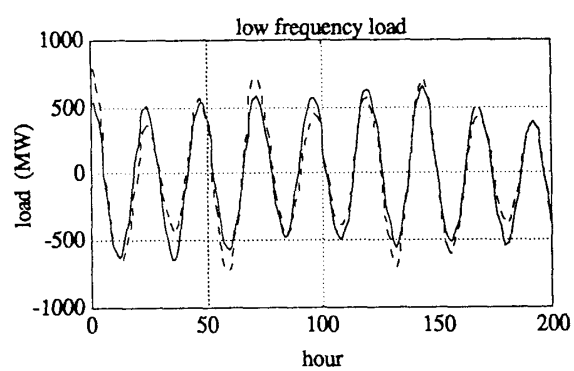

Figure 12. Actual and Forecasted Low-Frequency Load Components

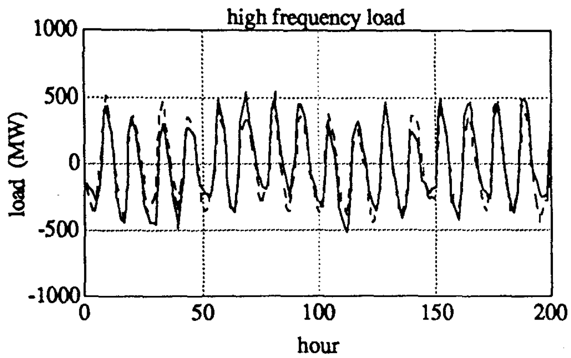

Figure 13. Actual and Forecasted High-Frequency Load Components

This paper presents a new neural network model, sometimes called Widrow model, for short-term load forecasting. It is different from previously proposed feedforward neural network approaches [10-15] in the following ways.

In feed-forward neural networks, the hidden neurons have a non-linear transfer function. The hidden neurons of the adaptive neural network have a linear transfer function that is a linear combination of the input sampling values. However, both of them provide non-linear mapping from input to output. The non-linearity in Adalines is because their weights depend on both the input and output. This feature determines that both networks can be used for non-linear system problems.

Another significant difference between the two networks is their learning rules. Feed-forward networks are trained by the back-propagation learning rule, which is designed to minimize the error function of selected training cases. The adaptive network is trained by Widrow learning rule, LMS algorithm, which is designed to minimize the expected error value. This learning rule is suitable for a sequential process such as load forecasting. It requires less computation than that of the back-propagation rule.

In this paper, we propose a model that combines the familiar models of Box and Jenkins with the new technology of neural networks. The basic structure of the load model is an autoregressive-moving average model. Prior researchers who have applied this type of model have recognized the limitations based on the need to regularly update the model parameters [9]. In our work, Adalines are incorporated into neural networks to provide continuous updating procedures for the model parameters. These one-week ahead forecasting results of less than 3.4 percent are quite favorable when compared to the 24-hour ahead forecasts using a Box and Jenkins model and nonlinear temperature model of Hagan with 3.73 percent mean absolute percentage error [8]. This comparison takes on greater significance since the proposed model is not reliant on temperature forecasts. Furthermore, limited simulation results of 24-hour ahead forecasts for weekdays Tuesday through Friday show an average absolute percentage error of less than 2.1 percent, again favorably comparable to other published results.

In summary, the proposed modelling approach incorporates a time series model into the neural network structure with Adalines. The results of this synthesis of a recognizable model with an artificial intelligent algorithm suggests a potentially worthwhile direction for further investigation.

REFERENCES

[1] S. Rahman and R. Bhatnagar, "An Expert System Based Algorithm for Short-Term Load Forecast," *IEEE Transactions on Power Systems*, vol. 3, No. 2, pp. 392-399, 1988.

[2] H. Van Meeteren and P. Van Son, "Short-Term Load Prediction with a Combination of Different Models," IEEE Power Industry Computer Applications Conference, pp. 192-197, 1979.

[3] J. Park, Y. Park, and K. Lee, "Composite Modeling for Adaptive Short-Term Load Forecasting," Presented at the IEEE Power Systems Engineering Summer Meeting, 1990.

[4] N.F. Hubele and C.-S. Cheng, "Identification of Seasonal Short-Term Forecasting Models Using Statistical Decision Functions," *IEEE Transactions on Power Systems*, Vol. PWRS-5, No. 1, pp.40-45, 1990.

[5] A.D. Papalexopoulos, and T.C. Hesterberg, "A Regression-Based Approach to Short-Term System load Forecasting," *IEEE Transactions on Power Systems*, Vol. PWRS-5, No. 4, pp. 1535-1547, 1990.

[6] R. Campo, and P. Ruiz, "Adaptive Weather-Sensitive Short Term Load Forecast," *IEEE Transactions on Power Systems*, Vol. PWRS-2, No. 3, pp.592-600, 1987.

[7] S.V. Wemuri, E.F. Hill and R. Balasubramanian, "Load Forecasting Using Stochastic Models," IEEE Power Industry Computer Applications Conference, pp. 31-37, 1973.

[8] M.T. Hagan and S.M. Behr, "The Time Series Approach to Short Term Load Forecasting," *IEEE Transactions on Power Systems*, Vol. PWRS-2, No. 3, pp. 785-791, 1990.

[9] M. Hagan and R. Klein, "On-Line Maximum Likelihood Estimation for Load Forecasting," *IEEE Transactions on Systems, Man and Cybernetics*, Vol. SMC-8, No. 9, pp. 711-715, 1978.

[10] D. Park, M. El-Sharkawi, R. Marks, A. Atlas and M. Damborg, "Electric Load Forecasting Using an Artificial Neural Network" *IEEE Transactions on Power Systems*, Vol. PWRS-6, No. 2, pp. 442-449, 1991.

[11] M.A. El-Sharkawi, S. Oh, R.J. Marks, M.J. Damdorg and C.H. Brace, "Short Term Electric Load Forecasting Using an Adaptively Trained Layered Perceptron," Proceedings of the First International Forum on Applications of Neural Networks to Power Systems, pp. 3-6, 1991.

[12] J.T. Conner, L.E. Atlas, and D. Martin, "Recurrent Neural Networks and Load Forecasting," Proceedings of the First International Forum on Applications of Neural Networks to Power Systems, pp. 22-25, 1991.

[13] K.Y. Lee, Y.T. Cha and C.C. Ku, "A Study of Neural Networks for Short-Term Load Forecasting," Proceedings of the First International Forum on Applications of Neural Networks to Power Systems, pp. 26-30, 1991.

[14] T.M. Peng, N.F. Hubele and G.G. Karady, "Advancement in the Application of Neural Networks for Short-Term Load Forecasting," Presented at the IEEE Power Engineering Society Summer Meeting, San Diego, July 28-Aug. 1, 1991.

[15] S. Chen, D. Yu and A. Moghaddamjo, "Weather Sensitive Short-Term Load Forecasting Using Nonfully Connected Artificial Neural Network," Presented at the IEEE Power Engineering Society Summer Meeting, San Diego, July 28-Aug. 1, 1991.

[16] B. Widrow and M.E. Hoff, "Adaptive Switching Circuits," 1960 *WESCON Convention Record Part IV*, 96-104, 1960.

[17] B. Widrow, "Generalization and Information Storage in Network of Adaline 'Neuron'," Self-Organizing Systems, M. Yovitz, G. Jacobi, and G. Goldstein, Eds. Washington, DC: Spartan Books, pp. 435-461, 1962.

[18] B. Widrow and M.A. Lehr, "30 Years of Adaptive Neural Networks: Perceptron, Madaline, and Backropagation," *Proceedings of the IEEE*, vol. 78, No. 9, September 1990, pp. 1415-1442.

[19] B. Widrow and S.D. Stearns, *Adaptive Signal Processing*, Prentice-Hill, Inc. N.J., 1985

[20] D.J. Defatta, J.G. Lucas and W.S. Hodgkiss, *Digital Signal Processing: A System Design Approach*, John Wiley & Sons, Inc., 1988.

[21] A.S. Lapedes and R. Farber, "Nonlinear Signal Processing Using Neural Networks: Prediction and System Modeling," Technical Report, Los Alamos National Laboratory, Los Alamos, New Mexico, 1987.

[22] G. Hripcsak, "Problem Solving Using Neural Networks," M.D. Computing, Vol. 5, No. 3, 25-37, 1988.

Tie Mao Peng was born in Heibei, China. He received his B.S. in Electrical Engineering from the North China Institute of Electric Power (China) in 1982, and the M.S. and Ph.D degree in Electric Engineering from Arizona State University in 1988 and 1991 respectively. Since

May 1991, He has been working as a consultant in Pacific Gas & Electric Company's Electric Supply Unit Computer Service Department.

Norma Hubele (M '88) was born in Lawrence, Massachusetts. She received her B.S. in Mathematics from the University of Massachusetts at Amherst in 1975, her M.S. in Operations Research and Statistics, and her Ph.D. in Computer and Systems Engineering from Rensselear Polytechnic Institute, Troy, New York, in 1977 and 1984.

She currently holds a faculty position in the Industrial and Management Systems Engineering Department of Arizona State University and is Associate Director of the Statistics and Engineering Applications for Quality Laboratory with the computer Integrated Manufacturing Systems Research Center. She is also a member of the center for Energy Systems Research. Her technical Interests include statistical analysis of large databases, statistical forecasting, and control. She has worked extensively in electrical load forecasting for both industry and research.

Dr. Hubele is a member of IEEE Power Engineering, the Institute of Industrial Engineers, American Statistical Association, and American Society for Quality Control.

George Karady (SM'70, F'78) was born in Budapest, Hungary. He received his BSEE and Doctor of Engineering degree in electrical engineering from Technical University of Budapest in 1952 and 1960, respectively. Dr. Karady was appointed in Salt River Project Chair Professor at Arizona State University in 1986 where he is responsible for electrical power education and performs research in power electronics high-voltage techniques and electric power. Previously, he was with EBASCO Services where he served as Chief Consulting Electrical Engineer, Manager of Electrical Systems, and Chief Engineer of Computer Technology. He was Electrical Task supervisor for the Tokomak Fusion Test Reactor Project in Princeton, New Jersey. From 1969 to 1977 he worked for the Hydro Quebec Institute of Research as a Program Manager and in 1976 was elected Research Fellow. Between 1952 and 1969 he worked for the Technology University of Budapest where he progressed from Post Doctoral Student to Deputy Department Head. In addition, Dr. Karady worked as visiting lecturer/professor in England and in Bagdad, Iraq. Between 1980 and 1986 he served as an Adjunct Professor of the Brooklyn Polytechnic Institute. Dr. Karady is a registered professional engineer in New York, New Jersey, and Quebec, Canada. He is the author of more than 80 technical papers.

Dr. Karady is a member of the CIGRE U.S. National Committee, where he served as vice president and presently Secretary-Treasurer. He is the Chairman of the IEEE Subcommittee on Lightning and Insulator and WG on Non-Ceramic Insulators.

He was a member of the Canadian Electrical Engineering Association and the Electrical Engineering Association of Hungary, where he held different positions between 1952 and 1966.

Discussion

Z. Ouyang and **N. I. Deeb** (Sargent & Lundy Engineers, Chicago, IL): This paper presents a new method in short term load forecasting practice. Regarding the methodology and implementation of the proposed method, we would appreciate if the authors can share their ideas with us on the following items:

1. Regarding the order of the Adaline L, did the authors experiment the effect of varying L on the forecasting error, and is there an optimal range of L for a practical problem size?
2. The optimal weight matrix W^* of the Adaline is solved iteratively using Equation (14). For time index k, the weights are updated by the error function shifted 24 hours back $[e(k - 24)]$. Therefore the lead time is 24 hours. Since the high frequency component has 12 hour period, would it be reasonable to use 12 hour lead time for the high frequency component Adaline? It is noted that either 24 hour or 12 hour lead time is based on heuristics.
3. This method seems not to encounter one important aspect of load forecasting, which is special days such as national holidays. Did the authors perform a pre-screening on the load sequence for such days, or any other considerations? In our opinion, holidays present some special low frequency components and may be treated similarly as the temperature factor in the mothed.
4. In Section 3.2, based on the 400 hour load sequence given in Figure 1, the passband frequencies for the three components are provided. Are the same passband widths used for the simulations in Section 6, and could the authors discuss how the passband widths are determined and why the low frequency band is significantly wider than the others?

Finally, we congratulate the authors for presenting another novel paper in the application of neural network to the power system study.

Manuscript received July 27, 1992.

T.M. PENG, N.F. HUBELE, G.G. KARADY: We wish to thank Drs. Ouyang and Deeb for their interest in our paper. We respond to their question sequentially.

1. In general, a larger order of the Adaline can produce a higher forecasting accuracy. This improvement of the accuracy, however, will diminish when the order is "large enough". We used L=20 in the presented work because no further improvement could be made by using the larger order.

2. We used e(k-24) to update the weights because it was assumed that we did not know the errors of e(k), e(k-1), ..., e(k-23). Depending on the application of the algorithm in practice, however, one can use the more recent error to update the network weights, such as that suggested by the discussers.

3. The load forecasting method presented in this paper did not incorporate a holiday day model. The holidays were not included in both the training and testing data. To deal with this aspect, we consider each holiday as an isolated event which can be represented as a set of impulses. This set of impulses will move the load curve downward in the special time of a year. The magnitudes of the impulses can be estimated from the previous holiday and the most recent weekend data. When using this holiday model, an additional item can be added to the network output node. This item has a zero value on regular days and has the above mentioned impulse values on holidays.

4. The selection of the passband widths of the digital filters was based on heuristics and tests. We estimated the digital filter passbands by observing the spectrum of the load signal. Then, by observing the output signal curve of the filter, we adjusted the passband frequency until a desired output signal shape was obtained. The same passband frequencies were used in Section 3.2 and Section 6.1.

We appreciate your inquiries and encourage continued experimentation, discussion and application of this interesting load forecasting methodology.

Manuscript received September 28, 1992.