SHORT-TERM LOAD FORECASTING

This article is intended to provide an overview of the state-of-the-art methodologies for electric power system short-term load forecasting (STLF). Three categories of STLF methods are described in the second section with technical details for readers who are interested in advanced research in STLF. Enhanced modeling techniques based on statistical methods are presented with simulation results in the third section to demonstrate how to employ the various existing modeling techniques to build a realistic application. This article concludes with the prospective of the development of STLF methods in the future. An introduction to the STLF problem follows.

Power systems are large-scale, nonlinear, and geographically distributed systems. The objective of power systems operators is to provide high-quality services at a reasonable price to the individual end users. This objective can be achieved only through efficient planning, scheduling, and minute-to-minute operation. The availability of sufficient information about a power system is critical for its efficient operation. The future system demand is a crucial piece of information. Much effort has been made in developing sophisticated methodologies to solve this problem.

Power system planning deals with determining the optimal mix and capacity of generation, capacity, and voltage level of transmission and distribution system additions, and the type of facilities required in transmission expansion plans, in the long or medium term. Scheduling focuses on when to start up or shut down generating units, how to coordinate different energy sources (e.g., hydro and thermal energy), and which equipment to maintain at what time, so that consumers are served reliably and economically. Careful scheduling is a prerequisite for high-quality service to consumers. It is the minute-to-minute operation, however, that makes reliable and economic services possible. Without knowledge of future load demand, none of these three facets can be achieved. Load forecasting, therefore, plays a very important role in the planning, scheduling, and operating of power systems.

The close and constant tracking of system load by the system generation is a basic requirement in the operation of power systems. For economically efficient operation and for effective control, this must be accomplished over a broad spectrum of time intervals. STLF is generally defined as forecasting the system load from one hour ahead to one week (168 h) ahead. The principal objective of the STLF function is to provide accurate load predictions for basic generation scheduling tasks, such as economic dispatch, unit commitment, and interchange evaluation, in addition to the important task of system security assessment.

The timeliness and accuracy of STLF have significant effects on power system operation and production costs (1). Utilities have to make up adequate levels of spinning reserve and standby reserve in order to have sufficient generation to satisfy the demand and to have the desired measure of system security and reliability. Underprediction of load results in a failure to provide the necessary reserve which, in turn, translates into higher costs due to the need to use expensive peaking units. Overprediction of load, on the other hand, involves the start-up of too many units resulting in an unnecessary increase in reserve and, hence, operating costs. Thus, by reducing the forecasting error, reserve levels may be reduced while satisfying the demand and maintaining the system security.

The behavior of the system load is directly governed by the customers' activity pattern. These include the seasonal, weekly, and daily periodicities, legal and religious holiday effects, which form the inherent properties of the load of a system. Meanwhile, some external factors, such as weather, special events, and so forth, greatly affect the individual consumption pattern and, in turn, the load behavior of the system. Thus, the system load behavior is not only governed by a self regressive rule but is explicitly affected by a number of external factors. This behavior is system and environment dependent, a fact that makes the load forecasting problem quite complicated. The search for a robust STLF algorithm which is suitable for different systems and different cases has been a challenging task for the power systems community.

The total load demand of a power system consists of energy consumption of individual consumers. An individual consumer may be a residence, a factory, or a shopping mall. The activities of an individual consumer have to some extent certain regularity. For example, a residential user goes to work at around 8 AM and comes home at about 5 PM Monday through Friday, a factory operates 12 h daily seven days a week, a mall is open from 10 AM to 9 PM five days a week except weekends, and so on. This regularity causes the hourly load demands to follow some patterns from one day to the next. Figure 1 shows the plot of hourly loads (in megawatts) on a power system for two consecutive days in January of a given year. It can be seen that there is a daily pattern for hourly loads of the two days, though differences in magnitude exist. This can be generalized and is called daily periodicity. Similarly, there are weekly, seasonal, yearly, and other periodicities.

In principle, one could determine the load patterns of a system if each of the individual consumption patterns were known. However, demand or usage pattern of an individual load customer is quite random and highly unpredictable. In addition, there are other factors affecting hourly loads, such as temperature, humidity, wind speed, holidays, economic fac-

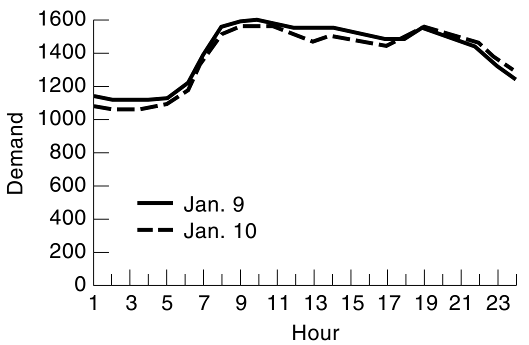

Figure 1. Hourly load demands for two consecutive days.

tors, and other random factors. These factors must be taken into account for accurate load forecasting.

Different load forecasting methods were developed for various purposes. Based on the purpose of application, they can be categorized into three categories:

1. Very short-term load forecasting (minutes or hours ahead)
2. Short-term load forecasting (days to several weeks ahead)
3. Medium- and long-term load forecasting (months to years)

Very short-term load forecasting is mainly intended to provide load demand information in the very near future so that automatic generation control functions are appropriately adjusted to follow load fluctuations economically. Short-term load forecasting is necessary for the operational planning of a power system. Its main objective is to predict the hourly load demands, one day or one week ahead. Medium- or long-term load forecasting is essential for scheduling equipment maintenance, and planning generation, transmission, and distribution of a power system.

The three categories are important. However, this article will concentrate on the category of short-term load forecasting; that is, daily to weekly load forecasting.

The primary application of the STLF function is to drive the scheduling functions that determine the most economic commitment of generation resources that comply with reliability requirements, operational constraints, and environmental constraints. For power systems with hydro energy being the major generation resources, load forecasts are required to determine the optimal releases from the reservoirs and generation levels in the power house. For systems of dominant thermal generations, the load forecasts are used for generating unit commitment to determine the minimal cost hourly strategies for the start-up and shutdown of units to meet the predicted loads. For hybrid hydro and thermal systems, the load forecasts are used to coordinate hydro and thermal generation resources so that production costs are minimized. These scheduling applications require hourly system load forecasts for the next day or the next week for the determination of the least cost operating plans. A closely associated scheduling task is the scheduling and contracting of interchanges between electric utilities. For this application, the STLF is useful for calculation of the economic levels of interchange.

The second application of STLF is for predictive assessment of the power system security. System load forecasts are an essential data requirement for the off-line transmission system analysis for the detection of future conditions under which the power system may be vulnerable. This information permits the operating personnel to prepare the corrective actions so that the system can be operated securely.

The third application of STLF is to provide system operating personnel with timely information, that is, the most recent load forecast. They need this information to operate the system economically and reliably.

The technical literature displays a wide range of methodologies and models for STLF. In general, there are three categories for the existing approaches: (1) statistical methods, (2) expert systems, and (3) artificial neural networks.

Statistical approaches generally employ two kinds of models: static models (1–5) and dynamic models (6–9). Static models assume that the load is a linear combination of some functional elements which describe the variation of weather variables or the basic characteristics of load behavior. The model parameters are estimated by using multiple linear regression or exponential smoothing techniques. These models are structurally simple and require relatively low computational effort. Dynamic models are based on a time series description, or its equivalent state space description (10), of the load behavior. These models treat the load pattern as a time series signal with known seasonal, weekly, and daily periodicities. This gives a rough prediction of the load at the given season, day of the week, and time of the day. The difference between the predicted and the actual load is considered as a stochastic process. The analysis of this random process leads to a more accurate prediction. The techniques used for the analysis of this random process include Kalman filtering (6), the auto-regressive moving average (ARMA) models (9), Box–Jenkins method (11), and the general exponential smoothing (12). ARMA models are the most commonly used dynamic models which model stationary processes with finite variances. Nonstationary processes can be modeled by differencing the original process. The differencing operation produces an auto-regressive integrated moving average (ARIMA) model (11). Some dynamic models explicitly include weather information as input variables (8). Others rely on a more heuristic approach where the load process is preliminarily corrected for weather influences (13).

Expert systems-based approaches have emerged as a result of advances in the field of artificial intelligence in the last two decades. An expert system is a computer program (though not algorithmic) which has the ability to act as an expert. This program can reason, explain, and have its knowledge base expanded as new information becomes available to it. In the case of the STLF problem, the forecasting system emulates the knowledge, experience, and analogical thinking of experienced system operators. The objective is to identify variables and rules that are used by system operators in estimating or forecasting the system load and the criteria for employing different rules in different situations. Examples demonstrating the application of this approach to STLF problems can be found in various references (14–16). In conjunction with expert systems, fuzzy set theory was proposed as a tool to handle the uncertainties in load models, weather variables, and operators' heuristic rules (17). Expert system models seem to be robust and adaptable to changing conditions than other methods.

In recent years, artificial neural network (ANN) techniques have been applied to perform STLF (18–23). The ANNs can extract the implicit nonlinear relationship between past load or weather variables and forecasted load. They do not rely on explicit function representation of input variables and the load to be forecast. It also has the capability to adapt to a changing forecasting environment through the concept of self-learning. ANN application in STLF has been an ongoing research area and promising results have been achieved.

STLF METHODS

Statistical Methods

There are many different variations of statistical methods for STLF. They can be divided into the following four general methodologies:

1. Multiple linear regression
2. Stochastic time series
3. General exponential smoothing
4. State space and Kalman filter

Multiple Linear Regression. In the multiple linear regression method, the load is represented in terms of explanatory variables and weather and non-weather variables that influence the electrical load. The multiple linear regression model STLF can be described in the following form:

$$y(t) = a_0 + a_1 x_1(t) + \cdots + a_n x_n + \epsilon(t) \tag{1}$$

where $y(t)$ is the electrical load demand to forecast, $x_1(t), \ldots, x_n(t)$ are explanatory variables correlated with $y(t)$, $\epsilon(t)$ is a random variable with zero mean and constant variance, and $a_0, a_1, \ldots, a_n$ are regression coefficients.

The explanatory variables of this model are identified on the basis of correlation analysis on each of these independent variables with the load variable to forecast. The estimation of the regression coefficients is usually computed with the least-square estimate technique. Statistical tests, such as the F-statistic test, are applied to determine the significance of these regression coefficients. The t-ratios resulting from these tests determine the significance of each of the coefficients, and correspondingly the significance of the associated variables with these coefficients.

Stochastic Time Series. Stochastic time series based STLF methods appear to be the most popular. This method has been widely utilized and is still being used for STLF in electric power utilities. The theory of stochastic time series is discussed in many text books, and there are many publications on stochastic time series based STLF methods. With this method, the load series, $y(t)$, is modeled as the output from a linear filter that has a random series input, $\epsilon(t)$, usually called a white noise. Depending on the characteristic of the linear filter, different models can be classified as follows (11).

Auto-Regressive Process. In the auto-regressive (AR) process, the current value of the time series, $y(t)$, is expressed linearly in terms of its previous values $(y(t-1), y(t-2), \ldots)$ and a random noise $\epsilon(t)$. The order of this process depends on the oldest previous value at which $y(t)$ is regressed on. For an AR process of order p, this model can be written as:

$$y(t) = \phi_1 y(t-1) + \phi_2 y(t-2) + \cdots + \phi_p y(t-p) + \epsilon(t) \tag{2}$$

With the introduction of the back-shift operator B that defines $y(t-1) = By(t)$, and consequently $y(t-m) = B^m y(t)$, Eq. (2) can be re-written in the following form:

$$\phi(B)y(t) = \epsilon(t) \tag{3}$$

where $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots \phi_p B^p$.

Moving Average Process. In the moving average (MA) process, the current value of the time series $y(t)$ is expressed linearly in terms of current and previous values of a white noise series $\epsilon(t), \epsilon(t-1), \ldots$. This noise series is constructed from the forecast errors or residuals when load observations become available. The order of this process depends on the oldest noise value at which $y(t)$ is regressed on. For a moving average model of order q, this model can be written as:

$$y(t) = \epsilon(t) - \theta_1 \epsilon(t-1) - \theta_t \epsilon(t-2) - \cdots - \theta_q \epsilon(t-q) \tag{4}$$

A similar application of the back-shift operator on the white noise series would allow Eq. (4) to be re-written as:

$$y(t) = \theta(B)\epsilon(t) \tag{5}$$

where $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \theta_p B^q$.

Auto-Regressive Moving-Average Process. In the auto-regressive moving-average process, the current value of the time series $y(t)$ is expressed linearly in terms of its values at previous periods $(y(t-1), y(t-2), \ldots)$ and in terms of current and previous values of a white noise series $(\epsilon(t), \epsilon(t-1), \ldots)$. The order of the ARMA process is determined by both the oldest previous value of the series and the oldest white noise value at which $y(t)$ is regressed on. For an ARMA process of order p and q, the model is written as:

$$\begin{aligned} y(t) &= \phi_1 y(t-1) + \cdots + \phi_p y(t-p) + \epsilon(t) \\ &\quad + \theta_1 \epsilon(t-1) - \theta_q \epsilon(t-q) \end{aligned} \tag{6}$$

By using the back-shift operator previously defined, Eq. (6) can be re-written as follows:

$$\phi(B)y(t) = \theta(B)\epsilon(t) \tag{7}$$

Auto-Regressive Integrated Moving Average Process. The time series that can be defined as an AR, MA, or as an ARMA process, is called a *stationary process*. This means that the mean of the series and the covariances among its observations do not change with time. If the process is nonstationary, transformation of the series to a stationary process needs to be performed to obtain a stationary process. This can be achieved by differencing operations on the nonstationary process. By introducing the operator ∇, a differenced time series of order 1 can be written as $\nabla y(t) = (1 - B)y(t)$ using the definition of the back-shift operator B. Therefore, an order d differenced time series is written as $\nabla^d y(t) = (1 - B)^d y(t)$. The differenced stationary series can be modeled as an AR, MA, or an ARMA process to yield an ARI, IMA, or ARIMA time series model. For a series that needs to be differenced d times and has orders p and q for the AR and the MA components, the model can be expressed as follows:

$$\phi(B)\nabla^d y(t) = \theta(B)\epsilon(t) \tag{8}$$

Seasonal Processes. As was pointed out in the first section, there are daily, weekly, yearly, or other periodicities in load demand time series. As a result, a different class of models that can describe this property is designated as seasonal processes. Seasonal time series could be modeled as an AR, MA, ARMA, or ARIMA seasonal models similar to the nonseasonal time series. The general multiplicative mode $(p, d, q) \times (P, D, Q)_S$ for a time series model can be written in the following form:

$$\phi(B)\Phi(B^S)\nabla^d\nabla_S^D y(t) = \theta(B)\Theta(B^S)\epsilon(t) \tag{9}$$

where ∇^d, $\phi(B)$, and $\theta(B)$ were defined previously. Similar definitions for ∇_S^D, $\Phi(B^S)$, and $\Theta(B^S)$ are given:

$$\nabla_S^D = (1 - B^S)^D \tag{10}$$

$$\Phi(B^S) = 1 - \Phi_1 B^S - \Phi_2 B^{2S} - \cdots - \Phi_P B^{PS} \tag{11}$$

$$\Theta(B^S) = 1 - \Theta_1 B^S - \Theta_2 B^{2S} - \cdots - \Theta_Q B^{QS} \tag{12}$$

The model described in Eq. (9) can be extended to the case where two seasonalities are accounted. The order of the model is $(p, d, q) \times (P, D, Q)_S \times (P', D', Q')_{S'}$, and is expressed as follows:

$$\phi(B)\Phi(B^S)\Phi'(B^{S'})\nabla^d\nabla_S^D\nabla_{S'}^{D'} y(t) = \theta(B)\Theta(B^S)\Theta(B^{S'})\epsilon(t) \tag{13}$$

where definitions for $\nabla_{S'}^{D'}$, $\Phi'(B^{S'})$, and $\Theta'(B^{S'})$ are similar to Eqs. (10), (11), and (12).

An example demonstrating the seasonal time series models is the model for an hourly load time series. If the daily periodicity is determined to be included in the model, Eq. (9) can be applied with $D = 1$ and $S = 24$. If data analysis shows that both daily and weekly periodicities should be reflected in the time series model, Eq. (13) can be used to build an ARIMA model with $D = 1$, $D' = 1$, $S = 24$, and $S' = 168$.

Transfer Function Method. The previous models permit $y(t)$ to be expressed in terms of its history and a white noise. If other variables affect the value of $y(t)$, the effect of these variables should be accounted for in the model. This can be achieved by using the transfer function model. For the case of one independent variable $x(t)$, such as the temperature, the transfer function model can be written in the following form:

$$y(t) = \frac{\omega(B)}{\sigma(B)} x(t - b) + \epsilon(t) \tag{14}$$

where

$$\omega(B) = \omega_0 - \omega_1 B - \omega_2 B^2 - \cdots - \omega_r B^r \tag{15}$$

$$\sigma(B) = 1 - \sigma_1 B - \sigma_2 B^2 - \cdots - \sigma_s B^s \tag{16}$$

b is response lag time of the independent variable $x(t)$ with respect to $y(t)$, and $\epsilon(t)$ is a colored (nonwhite) noise series. The series $\epsilon(t)$ can be modeled in terms of its past values and a white noise by using any of the techniques discussed previously.

The identification of the time series models can be obtained by analyzing the raw historical load data. This analysis may include the use of the range-mean, autocorrelation function, and partial autocorrelation function plots. This usually leads to initial guesses of the required data transformation and degree of differencing to obtain a stationary process. The degrees of the AR and the MA polynomials are initially determined by means of autocorrelation function analysis and the partial autocorrelation function analysis. For the case of the transfer function time series model, the cross-correlation function plot between the historical load time series, $y(t)$, and the independent variable, $x(t)$, is also needed in order to evaluate the response lag time, b, along with the orders of r and s of the polynomials $\omega(B)$ and $\sigma(B)$.

The estimation of parameters of the identified load forecasting model is usually achieved through the use of an efficient estimation method. For a pure AR model, the application of the Yule–Walker equation solution gives the estimates of the parameters of this process (11). Other methods such as the maximum likelihood techniques are capable of being applied to other models. In the third section, an adaptive weighted recursive least-square estimate method is described. Along with the estimation of the load forecasting model, estimation of the standard deviation and correlation of the parameters of the model with the variances and covariances of the residuals are established for analysis.

The load forecast model obtained can be assumed to be correct only if the model passes the diagnostic checking test. This test can be performed simply by checking whether the residual series is a white noise. If not, the inadequacy of the model needs to be corrected in view of the autocorrelation function and partial autocorrelation function.

General Exponential Smoothing. With the general exponential smoothing (GES) method, the load at time t, $y(t)$, can be modeled by using a fitting function as follows:

$$y(t) = c(t)^T f(t) + \epsilon(t) \tag{17}$$

where $f(t)$ is the fitting function vector for the process, $c(t)$ is the coefficient vector, $\epsilon(t)$ is a white noise, and T is vector transpose operator.

The estimates of the coefficients can be found by using weighted mean square error for the recent N sampled intervals. This is achieved by minimizing the following function:

$$\sum_{j=0}^{N-1} w^j \left[y(N - j) - f^T(-j)c \right]^2 \qquad 0 < w < 1 \tag{18}$$

This minimization gives the estimate of the coefficients in the following form:

$$\hat{c}(N) = F^{-1}(N)h(N) \tag{19}$$

where

$$F(N) = \sum_{j=0}^{N-1} w^j f(-j) f^T(-j) \tag{20}$$

$$h(N) = \sum_{j=0}^{N-1} w^j j(-j) y(N-j) \tag{21}$$

The forecast of the series at lead time l is found as:

$$\hat{y}(N+l) = f^T(l)\hat{c}(N) \tag{22}$$

The coefficient estimates and the forecasts can be updated, as new observations are available, respectively as follows:

$$\hat{c}(N+1) = L^T\hat{c}(N) + F^{-1} f(0)\left[y(N+1) - \hat{y}(N)\right] \tag{23}$$

$$\hat{y}(N+1+l) = f^T(l)\hat{c}(N+1) \tag{24}$$

where $F = \lim_{N\to\infty} F(N)$. The L matrix is called the *transition matrix* and is constructed on the basis that the model will have a fitting function satisfying the following relationship:

$$f(t) = Lf(t-1) \tag{25}$$

State Space and Kalman Filter. State space and Kalman filter is a general forecasting approach. It can include the previously described methods and more, such as time-varying coefficient models. In this method, the load is modeled as a state variable by using state space formulation. The state space formulation is designated by two sets of equations.

System state equations:

$$y(k+1) = \Phi(k)y(k) + W(k) \tag{26}$$

Measurement equations:

$$z(k) = H(k)y(k) + V(k) \tag{27}$$

where $y(k)$ is an $(n \times 1)$ process state vector at time t_k, $\Phi(k)$ is an $(n \times n)$ state transition matrix relating $y(k)$ to $y(k+1)$ when no forcing function exists, $W(k)$ is an $(n \times 1)$ white noise with a known white covariance $Q(k)$, $z(k)$ is an $(m \times 1)$ load measurement vector at time t_k, $H(k)$ is an $(m \times n)$ matrix relating $y(k)$ to $z(k)$ without noise, and $V(k)$ is an $(m \times 1)$ load measurement error that is a white noise with a know covariance $R(k)$.

The covariance matrices for vectors $W(k)$ and $V(k)$ have the following properties:

$$E\left[W(k)W(i)^T\right] = \begin{cases} Q(k) & i = k \\ 0 & i \neq k \end{cases} \tag{28}$$

$$E\left[V(k)V(i)^T\right] = \begin{cases} R(k) & i = k \\ 0 & i \neq k \end{cases} \tag{29}$$

The process noise, $W(k)$, and the measurement noise, $V(k)$, are also assumed uncorrelated. This assumption is described in the following:

$$E\left[W(k)V(k)\right]^T = 0, \qquad \text{for all } k \text{ and } i \tag{30}$$

At any time t_k, there will be an estimate for the process based on knowledge of the process up to t_{k-1}. This estimate is called the apriori estimate and is expressed as $y(k/k-1)$. The associated error between the actual and the previous estimates of the process is given as follows:

$$e(k/k-1) = y(k) - y(k/k-1) \tag{31}$$

This error vector has an error covariance matrix expressed by Eq. (32):

$$E\left[e(k/k-1)e(k/k-1)^T\right] = P(k/k-1) \tag{32}$$

The posteriori estimate is obtained as a linear combination from the apriori estimate and the measurement noise as in Eq. (33).

$$y(k/k) = y(k/k-1) + K(k)\left[y(k) - H(k)y(k/k-1)\right] \tag{33}$$

where $y(k/k)$ is the updated estimate and $K(k)$ is the blending factor.

The error associated with the actual and the posteriori estimate of the process is:

$$e(k/k) = y(k) - y(k/k) \tag{34}$$

The covariance matrix of this error vector is expressed by:

$$E\left[e(k/k)e(k/k)^T\right] = P(k/k) \tag{35}$$

The blending factor $K(k)$ is computed such that $y(k/k)$ is optimal in some sense such as the minimum mean squares error criterion. This factor is known as Kalman gain and the procedure for implementing Kalman filter for load prediction is as follows (24):

1. Find the process apriori estimate $y(k/k-1)$ and the error covariance matrix associated with it, $P(k/k-1)$.
2. Compute the Kalman gain

$$K(k) = P(k/k-1)H(k)^T\left[H(k)P(k/(k-1))H(k)^T - R(k)^{-1}\right] \tag{36}$$

3. Compute the updated estimate error covariance matrix

$$P(k/k) = \left[1 - K(k)H(k)\right]P(k/(k-1)) \tag{37}$$

4. Calculate the apriori estimate $y(k+1/k)$ and the error covariance matrix $P(k+1/k)$ associated with it

$$y((k+1)/k) = \Phi(k)y(k/k) \tag{38}$$

$$p((k+1)/k) = \Phi(k)P(k/k)Q(k)^T + Q(k) \tag{39}$$

5. Go to Step 2 moving to the next time step.

The state space method is very attractive for real-time applications as a result of the recursive nature of the Kalman filter. The optimal forecast is based on the assumed model. The model has to be known prior to using the Kalman filter. Identifying the model is the main task of this method, in addition to estimation of the noise covariances $Q(k)$ and $R(k)$.

Expert System Based Method

An expert system based load forecasting model can be built by using the knowledge about the load forecast domain from a human expert in the field. The knowledge engineer extracts this knowledge from load forecast domain expert by what is called the acquisition module component of the expert system. This knowledge is represented as facts and IF-THEN production rules. This representation is built in what is called the knowledge base component of the expert system. The search for solution or reasoning about the conclusion drawn by the expert system is performed by the inference engine of the expert system. An expert system has to have the capability to trace its reasoning if asked by the user. This facility can be built through an explanatory interface module.

The main efforts of building an expert system based model are acquiring expert knowledge in the STLF domain and organizing the acquired knowledge. Specifically for STLF, variables that affect load values need to be identified first. The relationship between the identified variables and the load need also to be constructed. Correlation analysis is useful for variable identification. Relating the variables and the electric load is essentially to build the IF-THEN production rules of the expert system. It should be recognized that the relationship between the variables and the electric is not static, but is dynamic. This means that the same magnitude of temperature changes affect load values to different degrees for different day types and seasons. Numerical values are expected outputs of an expert system based STLF. Therefore, some production rules may be qualitative and others are quantitative.

As was mentioned previously, the expert system based model is used to emulate the knowledge, experience, and analogical thinking of an experienced human expert. An advantage of a human expert is his/her capability to learn from new cases and to upgrade his/her knowledge and reasoning. Therefore, self-learning capability is a desired feature of an experts system.

Multilayer Feedforward Artificial Neural Network Based Methods

A multilayer feed forward artificial neural network (ANN) as a computing system consists of nodes or neurons connected by links. The nodes are divided into several layers: (1) the input layer, (2) the output layer, and (3) some hidden layers in between. A three-layer neural network is shown in Fig. 2. The nodes in the input layer take the input signal, and the nodes in the output layer provide the desired output signal. The required number of hidden layers and the number of nodes for each layer are problem dependent. Usually, one or two hidden layers are sufficient for the STLF problem. The number of nodes in the input layer depends on the number of the identified primary variables that affect load demand significantly. The number of nodes in the output layer depends on how the ANN is designed to generate outputs. For instance, if an ANN is designed for each hour of a day and it generates the hourly load forecasting, then one output node is needed. But, if the ANN is designed to provide hourly load forecasts for each of the 24 h of a day with one put-through of the input signal, 24 output neurons are needed. Other variants of the ANN architecture are also possible.

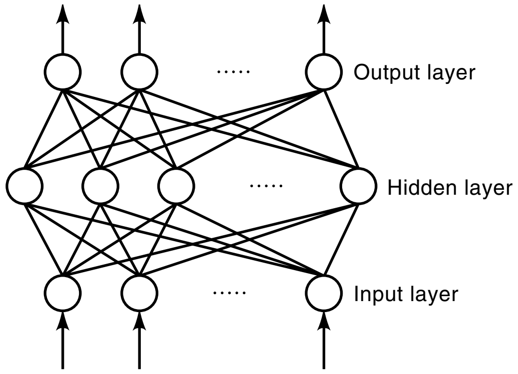

Figure 2. A multilayer feedforward artificial neural network.

In the feedforward ANN, signals can only be propagated from the input layer to the hidden layers and from the hidden layers to the output nodes. Signal propagation between nodes within the same layer or from the input layer directly to the output layer is not permitted.

For each neuron in the input layer, the neuron output is the same as the neuron input. For each neuron in the hidden layer or the output layer, the net input is given by:

$$\text{net}_j = \sum_i w_{ij} o_i \tag{40}$$

where i is a neuron in the preceding layer, o_i is the output of node i, and w_{ji} is the connection weight from neuron i to neuron j. The neuron output is given by Eq. (41):

$$o_j = \frac{1}{1 + e^{-(\text{net}_j + \theta_j)}} = f_j(\text{net}_j, \theta_j) \tag{41}$$

where net_j is the input signal to node j and θ_j is a bias. In the training process, θ_j can be regarded as a connection weight between node j and a fictitious node whose output always remains at unity.

The feedforward network is trained with an error back-propagation learning algorithm via selected sets of input–output training patterns. The algorithm begins with assigning a set of random numbers to the connection weights. When a pattern p with target output vector $t_p = [t_{p1}, t_{p2}, \ldots, t_{pM}]^T$ is presented, the connection weights can be updated as follows:

$$\Delta\omega_{ji}(p) = \eta \delta_{pj} o_{pi} + \alpha \Delta\omega_{ji}(p-1) \tag{42}$$

where, for the output nodes:

$$\delta_{pj} = (t_{pj} - o_{pj}) o_{pj} (1 - o_{pj}) \tag{43}$$

and for other nodes:

$$\delta_{pj} = \left(\sum_k \delta_{pk} \omega_{kj} \right) o_{pj} (1 - o_{pj}) \tag{44}$$

These equations are derived based on the criterion of minimizing the following error function:

$$E_p = \frac{1}{2} \sum_{j=1}^{M} (t_{pj} - o_{pj})^2 \tag{45}$$

where M is the number of output nodes. The method of steepest descent is used.

It is noted that the learning rate η and the momentum constant α in Eq. (42) affect the learning speed of an ANN significantly. Fine-tunings of them are usually needed when ANN training is performed, as their optimal values may be problem dependent.

An ANN can make forecasts only on the basis of the way it is trained. In power systems, numerous previous load and temperature data can show divergent load patterns and dynamic ranges. The selection of training cases from the data available can significantly affect the forecasting accuracy. Methods to select appropriate training cases need to be devised. For instance, the immediate previous two weeks data may be used for training and the load is forecasted for the present week. In this case, the ANN needs to be retrained each week. The general criterion of selecting training cases is that the ANN trained with cases of high similarity to the case to be forecasted can produce better forecasting results.

HYBRID STATISTICAL METHODS FOR STLF: AN EXAMPLE

Statistical STLF technology has been intensively studied over the past years and is widely used by electric utilities. Physically meaningful models have been developed. They are capable of describing influences due to weather pattern deviations from normal, and random correlation effects using a few explanatory variables. Such models are usually set up via off-line simulations. This reduces their adaptability and robustness.

Improved forecasting performance of the statistical-based STLF techniques can be obtained by:

1. Adaptive identification of explanatory variables
2. Advanced adaptive modeling methods
3. Adaptive parameter updating schemes

Different adaptive model parameter estimation algorithms have been devised for STLF (25,26), but models with fixed structure were used. The models to be described in this section focus on developing robust and adaptive modeling techniques based on statistical methodologies (30).

The algorithm considers the load to be composed of base load, weather-sensitive load, and random load components. Hybrid modeling techniques, which include an adaptive GES model, a nonlinear weather-sensitive load component model, and an adaptive AR model, are adopted to model the three load components. Power spectrum analysis and partial autocorrelation coefficients are applied to identify explanatory variables of the GES and AR models, respectively. The *AIC* criterion (29) is utilized to determine the optimal order of both GES and AR models.

Description of the Hybrid Models

Based on the properties of the load process, the following statistical-based hybrid models for the STLF are constructed. The load at hour t can be expressed as follows:

$$y(t) = yb(t) + yt(t) + yr(t) + \epsilon(t) \tag{46}$$

where $y(t)$ is hourly load at hour t, $yb(t)$ is base load component reflecting the daily or weekly regular changing patterns of the load and independent of weather conditions at hour t, $yt(t)$ is weather-sensitive load component at hour t, $yr(t)$ is random changing load component at hour t, and $\epsilon(t)$ is a white noise with zero mean and constant variance.

The base load component is dominated mainly by daily and weekly periodic load changes in normal conditions (except holidays). The GES model is quite suitable for modeling this component (26,27). Previously reported GES models are built through off-line spectral analysis of historical load data and applied on-line without modification. As load patterns change with time, especially during season-changing periods, the primary harmonics, whose composition reflects the changing load profile, also change. An adaptive GES modeling approach is able to adapt to the load pattern changes and is described subsequently.

The random load component, $yr(t)$, represents a stochastic process. Its analysis can always result in higher forecasting accuracy. An adaptive AR modeling method, enhanced by partial autocorrelation analysis, is used. This model is found to be highly effective in modeling the random load component.

Weather conditions have a strong influence on the behavior of electric load demand. They are generally presented in a nonlinear manner with considerable uncertainty (25). A general nonlinear weather-sensitive load model is used here and its description of this model is given.

In the proposed STLF algorithm, the GES model is identified weekly and the AR models are identified daily. With the proposed models, the weighted recursive least-square estimate (WRLSE) algorithm with a variable forgetting factor is employed for model parameter estimation. The utilization of this algorithm improves model tracking capability and numerical stability (26). This is important since in a real-time environment, STLF results are updated using the most recent load and weather information when situations such as those of abnormal or abrupt weather changes, unexpected social events, or even holidays occur. Also, for the purpose of on-line economic dispatch, interchange transaction evaluations, and network security assessment, updating is required for more accurate system analysis. Details of the proposed techniques follow.

Adaptive Modeling Techniques

Enhanced Adaptive GES Modeling

Model Description. The base load component, $yb(t)$, is modeled with the GES model, which can be expressed as:

$$yb(t) = c(t)^T f(t) \tag{47}$$

where $c(t)$ is a column vector of estimates of "n" locally constant coefficients, $f(t)$ is a column vector of "n" linearly independent fitting functions.

Equation (47) can be constructed by a finite Fourier series, as follows:

$$yb(t) = a_0 + \sum_{i=1}^{n} (a_i \sin \omega_i t + b_i \cos \omega_i t) \tag{48}$$

where $\omega_i = 2\pi/168\ k_i$, k_i is an integer (<84, Nyquist limit). Also, $f(t) = [1\ \sin\omega_1 t\ \cos\omega_1 t\ .\ .\ .\ \sin\omega_n t\ \cos\omega_n t]^T$ and $c(t) = [a_0\ a_1\ b_1\ .\ .\ .\ a_n\ b_n]^T$.

A major task in building an appropriate GES model for STLF is to identify the primary harmonics to be included in Eq. (48). An adaptive modeling scheme to perform this task was developed. A presentation of this scheme follows.

Adaptive Identification of Primary Harmonics. A systematic approach to adaptively determine the order of the GES model via power spectrum analysis is devised. Normalized power spectrums are defined and utilized to build the GES models.

The power spectrum of the load series $\{y(t)\}$, is defined by:

$$P(\omega_i) = \frac{1}{\pi}\left[\sigma_Y^2 + 2\sum_{j=1}^{J_l} R_{YY}(j)\cos j\omega_i\right] \tag{49}$$

where $R_{YY}(k)$ is the auto-covariance of $\{y(t)\}$, σ_Y is the variance of $\{y(t)\}$, J_1 is determined such that the $R_{YY}(J_1)$ calculated using the available sample data is statistically reliable. Usually, $J_1 \le N/M$ with M greater than 10 ($M = 13$ is found to be appropriate for the test case to be presented), and N is the number of data used for modeling.

The magnitudes of the coefficients (a_i and b_i) of the GES model are directly proportional to those of power spectrums of the corresponding harmonics (28). As a result, a harmonic with a larger power spectrum magnitude has a more significant effect on the GES model representing load change patterns. Inclusion of such harmonics improves the model accuracy.

The normalized power spectrums, instead of the values calculated using Eq. (49), were utilized because they reveal the relative significance of the harmonics in the GES model. They are defined by:

$$P_n(\omega_i) = \frac{P(\omega_i)}{P_{av}} \tag{50}$$

where

$$P_{av} = \frac{1}{J_2}\sum_{j=1}^{J_2} |P(\omega_j)|$$

is the average power spectrum value. In general, the upper bound for J_2 is the Nyquist limit. Simulation tests are usually needed to determine a proper value: $J_2 = 50$ is used in the following forecasting results.

The proposed adaptive GES modeling method calls for adding harmonics one by one into Eq. (48) in the ascending order of $P_n(\omega_i)$ until the sum of the squares of the model residuals reaches minimum. The model thus obtained can be considered to possess the optimal forecasting performance.

The amount of error reduction evaluated by the sum of the squared residuals may be insignificant and does not justify the additional computational effort due to the increased model order. In general, a model determined by this manner is not parsimonious. Model parsimony can be achieved by using the *AIC* criterion (29). The *AIC* for the GES model with k harmonics is calculated by Eq. (51):

$$AIC(k) = N \cdot \log\left[\tilde{\sigma}_W^2(k)\right] + 2(2k+1) \tag{51}$$

where the model residual is $\{w(t) = y(t) - yb(t)\}$, and $\tilde{\sigma}_W$ is the estimated variance of the residuals of the kth order GES model. According to *AIC* criterion, the optimal model order corresponds to the minimum *AIC* value computed in Eq. (51).

The order of a GES model is usually high. This makes the previously defined modeling procedure computationally inefficient. To speed up the modeling process, the following general heuristic rules are designed: Include in the initial GES model, those harmonics with $P_n(\omega_i) \ge \alpha$ (α, a prespecified threshold value); and rank the remaining harmonics with $P_n(\omega_i) < \alpha$ in ascending order in terms of their $P_n(\omega_i)$ and include them in the initial GES model one by one until Eq. (51) reaches minimum.

A threshold $\alpha = 0.5$ is considered to be appropriate. This choice of α is based on the observation that $P_n(\omega_i) \ge 0.5$ indicates that the relative significance of the harmonic in the GES model is above average compared to other harmonics.

Weather-Sensitive Load Component Model. The identified GES model does not take into account the effect of weather deviations from normal patterns on electric load consumption. Explicit inclusion of weather variables in the forecasting model is necessary to improve model accuracy.

Weather factors that are generally considered to significantly affect load demand are temperature, humidity, wind speed, and cloud cover. The relationship between changes in load and weather is appropriately described by nonlinear polynomials (26,27). The following general model for weather-sensitive load component was applied:

$$yt(t) = c_0 + \sum_{j=l}^{L_1}\sum_{k=l}^{M_j}\sum_{i=0}^{L_2} c_{ijk}T_j(t-i)^k \tag{52}$$

where L_1 is the number of weather variables to be modeled, M_j is a constant reflecting the nonlinearity of the relationship between weather variable and the weather-sensitive load component, c_{ijk} is the coefficient of the kth order of the jth weather variable at the ith lagging hour, $T_j(t)$ is the value of the jth weather variable at hour t, and L_2 is a constant reflecting time delay effect of weather change.

Inclusion of a weather variable in the weather-sensitive load component model depends on the following factors: strength of dependence of the weather-sensitive load component on this variable; and accuracy of the forecasted value of the weather variable. Use of a weather variable that has very high unpredictability may deteriorate rather than improve the load forecasting accuracy.

Based on these considerations, which variables to include in the model may be system dependent and need to be tested on specific system data. In the following example case, temperature is considered to be the most prominent weather variable. Accordingly, a third-order polynomial weather-sensitive model is used in the form of Eq. (52).

A hybrid model formed by combining the base load model component in Eq. (48) and the weather-sensitive load component model is constructed to estimate model parameters using the WRLSE algorithm.

Enhanced Adaptive AR Modeling. Once the task of building the base load and the weather-sensitive load components' models is completed, the random load component, $yr(t)$, can be calculated by:

$$yr(t) = y(t) - \tilde{a}_0 - \sum_{i=1}^{n}(\tilde{a}_i \sin\omega_i t + \tilde{b}_i \cos\omega_i t) - \sum_{j=l}^{L_l}\sum_{k=l}^{M_j}\sum_{i=l}^{L_2}\tilde{c}_{ijk}T_j(t-i)^k \quad (53)$$

where tilde parameters refer to the estimated values. The random series, $\{yr(t)\}$, can be considered a stationary time series and can generally be modeled by an ARMA model.

ARMA modeling methods have been used in STLF (1,9,26,27). It has been shown in (9) that AR models have forecasting quality equivalent to ARMA models. Promising forecasting was obtained using AR models (26). In addition, parameter estimation of AR models is computationally much less demanding than that of ARMA models. This is an attractive feature for the consideration of developing AR models with an enhanced modeling algorithm for use in this example. Based on the assumption that the most recent lagging term has a larger partial autocorrelation coefficient than the subsequent term, AR models were identified in (9) by successively adding the immediate lagging terms into the AR model one by one. This process continues until the sum of squares of the model residuals reaches minimum. Fixed AR models were utilized in (26,27). An improved adaptive AR modeling algorithm enhanced with partial autocorrelation analysis is described later. In addition to being adaptive, this algorithm recognizes that the most recent lagging term does not necessarily have a larger partial-autocorrelation coefficient than the subsequent one.

According to the enhanced AR modeling algorithm, the general form of an AR model of order p can be described by:

$$\theta(B)yr(t) = \epsilon(t) \quad (54)$$

where $\theta(B) = 1 - \sum_{i=1}^{p}\theta_{q_i}B^{q_i}$, θ_{q_i} is coefficient of the AR process, B is back-shift operator, q_i refers to the ith lagging term with q hours preceding the current hour in the AR(p) model. It is determined by partial autocorrelation analysis.

An outline of the proposed adaptive AR modeling method follows.

1. Compute the partial autocorrelation coefficients (PACs) of $\{yr(t)\}$ to the order J_3 as follows:

$$\rho_{YR-YR}(j) = \frac{\dfrac{1}{N-j-1}\sum_{t=1}^{N-j}\left[yr(t)-\overline{yr}\right]\left[yr(t+j)-\overline{yr}\right]}{\dfrac{1}{N}\sum_{t=1}^{N}\left[yr(t)-\overline{yr}\right]^2} \quad (55)$$

where

$$\overline{yr} = \frac{1}{N}\sum_{t=1}^{N}\left[yr(t)\right]$$

J_3 is determined in the same manner as J_1. The PACs are ranked in ascending order $\{q_1, q_2, \ldots, q_{J_3}\}$ in terms of their magnitudes.

2. Add the lagging term at the top of the current ranking to the AR model.
3. Estimate parameters of the current AR model.
4. Compute the sum of the squared model residuals.
5. Evaluate the significance of the decrease of the sum of the squared residuals using the *AIC* criterion. The *AIC* for the pth order AR model is calculated by Eq. 56:

$$AIC(p) = N \cdot \log[\tilde{\sigma}_\epsilon^2] + 2p \quad (56)$$

where $\tilde{\sigma}_\epsilon$ is the estimated variance of the residuals of the pth order AR model. If $AIC(p) < AIC(p-1)$, go to step 2; otherwise, the optimal parsimonious AR model is obtained.

WRLSE Algorithm. For a model expressed by:

$$y(t) = H(t)^T X(t) + \epsilon(t) \quad (57)$$

where $x(t)$ is the vector of the known values of the variables, $H(t)$ is the column vector of coefficients to be estimated, and $y(t)$ is the output of the model. The WRLSE algorithm with a variable forgetting factor λ can be described by the following set of equations:

$$R(t) = \frac{R(t-1)}{\lambda(t-1)}\left[I - \frac{x(t)x(t)^T R(t-1)}{\lambda(t-1) + x(t)^T R(t-1)x(t)}\right] \quad (58)$$

$$H(t) = H(t-1) + R(t)x(t)\left[y(t) - H(t-1)^T x(t)\right] \quad (59)$$

$$\hat{\epsilon}(t) = y(t) - H(t)^T x(t) \quad (60)$$

$$Q(t) = x(t)^T R(t)x(t) \quad (61)$$

$$S(t) = \sum_{j=1}^{t}\lambda(j)^{t-j}\hat{\epsilon}(j)^2 \quad (62)$$

$$\lambda(t) = \frac{1}{2}\left\{1 - Q(t) - \frac{\hat{\epsilon}(t)^2}{S(t)} + \sqrt{1 - Q(t) - \frac{\hat{\epsilon}(t)^2}{S(t)} + 4Q(t)}\right\} \quad (63)$$

Test Results

The adaptive modeling techniques were implemented. A cubic temperature-sensitive load model was used in order to evaluate the proposed adaptive modeling techniques against the fixed structure approach. The WRLSE algorithm with the variable forgetting factor was applied to estimate model parameters. In this simulation, four-week historical data (N = 672) were used for building the GES and AR models. Identification of the GES model was performed weekly; whereas AR models were identified daily. The error statistics were calculated with respect to daily peak loads. A threshold error value of one percent was used to activate the updating algorithm.

Historical data from two utilities were used. Utility A has a maximum and a minimum hourly load, respectively, of 32,343 and 11,071 MW. The maximum and minimum hourly average temperatures were 36.9°C (98.5°F) and 8.4°C (16.9°F). Utility B has a maximum and a minimum hourly load of 2,043 and 633 MW, respectively. The maximum and minimum hourly average temperatures were 33.9°C (93°F) and 30.6°C (−23°F). Utility A is a summer peak system with one yearly peak in July. Utility B is a summer and winter peak system with yearly peaks in August and February. The standard deviation of daily average temperature of Utility A is smaller than that of Utility B. Actual recordings of temperature from the two utilities were used in the following simulation studies.

A fixed modeling technique, which is being used in several utility systems (26), is utilized as the base to evaluate the improvement that can be achieved by the enhanced modeling techniques. The fixed models are described as follows:

Harmonics in the GES model:

$$\{1, 2, 3, 6, 7, 8, 13, 14, 15, 20, 21, 22, 28, 35, 49\}$$

Nonlinear temperature sensitive model:

$$yt(t) = \sum_{j=1}^{3} \{c_1 lj T_a(t)^j + c_{2j} T_a(t-1)^j + c_{3j} T_a(t-2)^j\}$$

where c_{1j}, c_{2j}, and c_{3j} are model parameters and $T_a(t)$ is the average equivalent temperature for two adjacent hours.

The AR model:

$$yr(t+1) = (\theta_0 + \theta_1 B + \theta_{23} B^{23}) yr(t)$$

Table 1 shows absolute mean error (AME) and standard deviation error (SDE) statistics of 24 h for Utility A by the proposed method. Highly accurate forecasting was achieved with seasonal average errors less than 1.5%. Compared with the fixed modeling method, the ratios of improvement in accuracy from the proposed method ranges from 10 to 30% for the four time periods on average. This is a result of using the proposed adaptive modeling techniques.

Table 1. AME Error Statistics of 24 Hours, Utility A

	Jan.–Mar.		April–June		July–Sept.		Oct.–Dec.	
Hour	AME	SDE	AME	SDE	AME	SDE	AME	SDE
1	0.859	1.068	0.536	0.692	0.526	0.670	0.600	0.751
2	0.805	1.104	0.763	0.966	0.642	0.710	0.819	1.015
3	0.909	1.154	0.860	1.039	0.684	0.861	0.874	1.072
4	1.346	1.562	0.888	1.121	0.588	0.779	1.042	1.290
5	1.053	1.098	0.795	1.065	0.581	0.706	0.864	1.066
6	2.187	2.610	1.531	1.950	1.388	1.532	2.270	2.676
7	1.356	1.692	1.298	1.632	1.117	1.406	1.431	1.819
8	1.725	1.926	1.257	1.559	0.879	1.127	1.759	2.060
9	1.001	1.283	1.099	1.171	0.959	1.232	1.071	1.239
10	1.153	1.398	0.977	1.057	0.956	1.161	0.936	1.227
11	0.923	1.147	0.994	1.119	0.953	1.178	0.776	0.999
12	0.993	1.198	1.094	1.103	0.853	1.060	0.931	1.186
13	0.923	1.199	1.089	1.172	1.036	1.244	1.115	1.403
14	0.955	1.028	1.222	1.312	0.779	0.974	0.903	1.216
15	0.852	1.075	0.940	1.104	1.032	1.239	1.201	1.541
16	0.798	0.994	1.267	1.584	1.063	1.293	1.193	1.484
17	1.398	1.093	1.076	1.293	0.949	1.263	1.227	1.564
18	1.453	1.744	1.330	1.828	1.109	1.366	1.914	1.947
19	1.362	1.577	1.335	1.526	1.120	1.423	1.718	1.989
20	1.467	1.452	1.543	1.949	1.176	1.412	1.434	1.826
21	1.225	1.182	1.385	1.334	1.199	1.489	0.843	1.077
22	0.708	0.904	0.967	1.169	0.746	0.890	0.821	1.019
23	0.826	1.039	0.812	1.034	0.720	0.862	0.588	0.740
24	0.908	1.094	0.651	0.831	0.755	0.967	0.788	1.041
Avg.	**1.133**	**1.468**	**1.071**	**1.391**	**0.909**	**1.179**	**1.130**	**1.532**

Table 2. AME Error Statistics of 7 Day Types, Utility A

	Jan.–Mar.		April–June		July–Sept.		Oct.–Dec.	
Day	AME	SDE	AME	SDE	AME	SDE	AME	SDE
Mon.	1.275	1.629	1.266	1.612	1.042	1.305	1.344	1.753
Tue.	1.095	1.403	1.128	1.369	0.979	1.182	1.173	1.621
Wed.	1.054	1.352	1.009	1.345	0.880	1.144	1.089	1.391
Thu.	1.103	1.363	0.977	1.278	0.983	1.235	1.159	1.522
Fri.	1.220	1.497	0.998	1.263	0.864	1.099	1.112	1.452
Sat.	0.894	1.180	1.009	1.293	0.826	1.044	1.144	1.476
Sun.	1.298	1.748	1.197	1.554	0.940	1.217	1.340	1.730

Table 2 presents the seasonal error statistics in terms of day types. Results show that AME and SDE are less than 1.5% and 2%, respectively, for all seven day types. The adaptive modeling approaches exhibit very stable performance for all day types. This verifies the designed adaptability and robustness.

The error statistics of peak load forecasting are given in Table 3. It is noted that errors for the time period from Oct. to Dec. are larger than those for the other three periods. This is because of holidays such as Thanksgiving Day, Christmas Day, and the New Year Eve. When compared to the fixed modeling approach, the ratios of improvement in accuracy reach up to 34% on average. Considering the importance of peak load forecasting, this improvement is quite significant.

By using the proposed adaptive modeling techniques, the identified models for Utility A in two consecutive weeks of April follow.

The identified harmonics for the two GES models are

$$\{1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 16, 18, 19, 21, 28, 35, 49\}$$

$$\{1, 2, 3, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 16, 18, 19, 21, 24, 26, 28, 35, 49\}$$

The harmonics used in the fixed structure model are {1,2,3,6,7,8,13,14,15,20,21,22,28,35,49}. In addition to most of these harmonics, other harmonics which vary from one week to another are also identified for the GES models by the modeling method described previously. Different thresholds were used to identify the primary harmonics in the GES model and

Table 3. AME Error Statistics of Peak Load Forecast, Utility A

	Jan.–Mar.		April–June		July–Sept.		Oct.–Dec.	
Day	AME	SDE	AME	SDE	AME	SDE	AME	SDE
Mon.	1.148	1.655	1.032	1.340	0.937	1.097	1.990	2.223
Tue.	1.041	1.235	1.635	1.677	0.811	0.863	2.044	1.874
Wed.	1.091	1.553	1.205	1.343	1.298	1.643	1.533	1.824
Thu.	1.413	1.470	0.884	1.191	1.055	1.228	1.579	1.536
Fri.	1.172	1.450	1.096	1.206	1.123	1.333	1.625	1.598
Sat.	1.145	1.610	0.709	0.877	1.088	1.185	1.342	1.771
Sun.	1.730	0.959	0.984	0.969	0.942	1.181	1.250	1.436
Avg.	**1.249**	**N/A**	**1.078**	**N/A**	**1.036**	**N/A**	**1.623**	**N/A**

a threshold of 0.5 was used as a better tradeoff between computational effort and modeling accuracy. In particular, this threshold value generally guarantees that the primary harmonics are all included in the final GES model.

The AR models of 7 successive days in the two weeks:

$$\text{Mon./Tue.:} \quad yr(t) = \sum_{i=1}^{10} \theta_i yr(t-i) + \sum_{i=21}^{23} \theta_i yr(t-i)$$

$$\text{Wed.:} \quad yr(t) = \sum_{i=1}^{9} \theta_i yr(t-i) + \sum_{i=21}^{24} \theta_i yr(t-i)$$

$$\text{Thu.:} \quad yr(t) = \sum_{i=1}^{8} \theta_i yr(t-i) + \sum_{i=21}^{24} \theta_i yr(t-i)$$

$$\text{Fri.:} \quad yr(t) = \sum_{i=1}^{8} \theta_i yr(t-i) + \sum_{i=21}^{23} \theta_i yr(t-i)$$

$$\text{Sat.:} \quad yr(t) = \sum_{i=1}^{5} \theta_i yr(t-i) + \sum_{i=23}^{24} \theta_i yr(t-i)$$

$$\text{Sun.:} \quad yr(t) = \sum_{i=1}^{5} \theta_i yr(t-i) + \sum_{i=22}^{24} \theta_i yr(t-i)$$

Simulations were also performed using one year historical hourly load data of Utility B. Reductions of daily average errors range from 6 to 12%. The reductions of peak load forecasting error range from 2 to 11%. This is achieved through identifying the appropriate harmonics for GES models and lagging terms for AR models, in terms of normalized power spectrum and partial autocorrelation analysis, respectively. The developed AR modeling approach produced AR models that are quite different from one day to the next and can better reflect the load change. This example demonstrates that improved forecasting has resulted from the enhanced modeling techniques.

PROSPECTIVE OF THE DEVELOPMENT OF THE STLF METHODS

Load forecasting is crucial for the expansion planning, reliability maintenance scheduling, and operational planning of power systems. STLF is a particularly important function in power system operations as the principal driving element for daily and weekly operations scheduling. In the last two decades, various modeling and forecasting techniques have been developed and have advanced the state-of-the-art of STLF methods dramatically.

Of the many models published in the literature and applied in power utilities, the time series based (or dynamic) models, like the exemplar STLF modeling method presented in the previous section, are the most popular. Such models are capable of describing time-correlated random phenomena with relatively few explanatory variables and parameters. This type of model is relatively easily developed and updated, with modest computational requirements. However, it should be realized that time series based approaches contain the fundamental assumption that from the behavior of the time series in the past, its behaviors in the future can be determined. This assumption may not always hold, because the system represented by the time series may not be self-explanatory, then this assumption will lead to erroneous predictions. Errors will be greater if some abrupt changes in the system occur, or if a long lead time is chosen. Some adaptive modifications of time series can improve the performance of the dynamic STLF models to some extent, but they will still have the bias of historical data. Static models, such as the multiple linear regression model, depend on identifying the so-called explanatory variables via complicated correlation analysis or even empirical knowledge. The main source of errors for static models is that functional relationships between explanatory and dependent variables are not stationary for load forecasting. They change over time. Similarly, various updating techniques can be used to alleviate this error. While each of the statistical STLF approaches demonstrates considerable success in forecasting accuracy, they are all subject to the risk that the future cannot be fully reflected by history. This is the risk associated with the fundamental assumption of all statistical modeling approaches.

In recent years, AI-based approaches have received intensive attention as an alternative method to STLF problems. Expert systems belong to this class of methods. Expert systems can employ human experience to validate intuitive insight about a process with a certain degree of success. This approach has a high reliance on the existence of an expert capable of making accurate forecasts, from whom computer software can be designed to emulate the expert. An inherent difficulty with expert opinions, however, is that they may not always be consistent, or the reliability of such opinions may be in question. Similar to maintaining models for statistical methods, updating of knowledge base and production rules of an expert system is very likely to happen relatively frequently to reflect the latest changes in load data and expert opinions. The main challenges to applications of expert systems to STLF are as follows: (1) The approach is not easily reproducible, that is, an expert system based on selected experts from certain utilities cannot assume an easy portability to other utilities, and (2) the lack of self-learning mechanisms, which are fundamentally different from a human expert whose knowledge continuously evolves.

Another category of AI-based methods are ANNs. Neural networks are a more promising area of artificial intelligence for STLF since they do not rely on human experience, but attempt to draw a link between a set of input data and observed outputs. This approach does not rely on an explicit adoption of a functional relationship between past load or weather variables and forecasted load. It has the capability to adapt to a changing forecasting environment through self-learning. In this sense, the ANN-based methods are very different from the other methods and demonstrate obvious advantages. However, the way in which an ANN is trained influences its performance significantly. If the training data set is not thoughtfully chosen, the resulting network is unlikely to hold up well. Massaging the set of training data probably demands most of the efforts in building a real-world ANN. A systematic method of preparing the set of training data helps create a more robust ANN with less effort.

As was discussed, any individual load forecasting method has its own merits and defects, and none of them has a decisive advantage. Therefore, there is always room for improvement. Hybrid forecasting techniques that blend various types of techniques for STLF, such as combining statistical and expert system based methods incorporating fuzzy logic into the conventional ANNs and others, represent the trend of the future research.

BIBLIOGRAPHY

1. P. C. Gupta and K. Yamada, Adaptive short-term load forecasting of hourly loads using weather information, *IEEE Trans. Power Appar. Syst.,* **91**: 1972, 2085–2094.
2. K. Srinivasan and R. Pronovost, Short-term load forecasting using multiple correlation models, *IEEE Trans. Power Appar. Syst.,* **94**: 1975, 854–858.
3. M. Abou-Hussain et al., An accurate model for short-term load forecasting, *IEEE Trans. Power Appar. Syst.,* **100**: 1981, 4158–4165.
4. S. Or, A weather sensitive model for short-term load forecasting, *Proc. IEEE/CSEE Joint Conf. High Voltage Transmission Syst. China,* Beijing, Oct. 17–22, 1987, 245–251.
5. A. D. Papalexopoulos and T. C. Hesterberg, A regression-based approach to short-term load forecasting, *IEEE Trans. Power Syst.,* **5**: 1990, 1535–1547.
6. M. A. Abu-El-Magd and N. K. Sinha, Two new algorithms for on-line modeling and forecasting of load demand of a multi-node power system, *IEEE Trans. Power Appar. Syst.,* **100**: 1981, 3246–3253.
7. B. Krogh, E. S. Llinas, and D. Lesser, Design and implementation of on-line load forecasting algorithm, *IEEE Trans. Power Appar. Syst.,* **101**: 1982, 3284–3289.
8. A. Keyhani and S. M. Miri, On-line weather-sensitive and industrial group bus load forecasting for micro-processor application, *IEEE Trans. Power Appar. Syst.,* **102**: 1983, 3868–3876.
9. S. Vemuri, W. L. Huang, and D. J. Nelson, On-line algorithm for forecasting hourly loads of an electric utility, *IEEE Trans. Power Appar. Syst.,* **100**: 1981, 3775–3784.
10. L. Ljung and T. Soderstrom, *Theory and Practice of Recursive Identification,* Cambridge, MA: MIT Press, 1983.
11. G. E. P. Box and G. M. Jenkins, *Time Series Analysis—Forecasting and Control,* Oakland, CA: Holden-Day, 1976.
12. W. R. Christiaanse, Short-term load forecasting using general exponential smoothing, *IEEE Trans. Power Appar. Syst.,* **90**: 1971, 900–910.
13. M. Honig and D. Messerschmitt, *Adaptive Filter, Structures, Algorithms, and Applications,* Hingham, MA: Klumer Academic Publishers, 1984.
14. S. Rahman and R. Bhatnagar, An expert system based algorithm for short-term load forecast, *IEEE Trans. Power Syst.,* **3**: 1988, 392–399.
15. K. Jabbour et al., ALFA: Automated load forecasting assistant, *IEEE Trans. Power Syst.,* **3**: 1988, 908–914.
16. S. Rahman and O. Hazim, Generalized knowledge based approach to short-term load forecasting, *IEEE Trans. Power Syst.,* **8**: 1993, 508–514.
17. K. Ho et al., Short-term load forecasting of Taiwan power system using a knowledge-based expert system, *IEEE Trans. Power Syst.,* **5**: 1990, 1214–1221.
18. D. C. Park et al., Electric load forecasting using an artificial neural network, *IEEE Trans. Power Syst.,* **6**: 1991, 442–449.
19. K. Y. Lee, Y. T. Cha, and J. H. Park, Short-term load forecasting using an artificial neural network, *IEEE Trans. Power Syst.,* **8**: 1992, 124–132.
20. K. Ho, Y. Hsu, and C. Yang, Short-term load forecasting using a multi-layer neural network with an adaptive learning, *IEEE Trans. Power Syst.,* **7**: 1992, 141–149.
21. S. Chen, D. Yu, and A. R. Moghhaddamjo, Weather-sensitive short-term load forecasting using non-fully connected artificial neural network, *IEEE Trans. Power Syst.,* **7**: 1992, 1098–1105.
22. T. Peng, N. Hubele, and G. Karady, Advancement in the application of neural networks for short-term load forecasting, *IEEE Trans. Power Syst.,* **7** (1): 1992, 250–257.
23. C. N. Lu, N. T. Wu, and S. Vemuri, Neural network based short-term load forecasting, *IEEE Trans. Power Syst.,* **8**: 1993, 334–342.
24. G. G. Brown, *Introduction to Random Signal Analysis and Kalman Filtering,* New York: Wiley, 1983.
25. Q. C. Lu et al., An adaptive nonlinear predictor with orthogonal escalator structure for short-term load forecasting, *IEEE Trans. Power Syst.,* **4**: 1989, 158–164.
26. J. Y. Fan and J. D. McDonald, A real-time implementation of short-term load forecasting for distribution power systems, *IEEE Trans. Power Syst.,* **9**: 1994, 988–994.
27. M. T. Hagan and S. M. Behr, The time series approach to short-term load forecasting, *IEEE Trans. Power Syst.,* **2**: 1987, 785–791.
28. R. G. Brown, *Smoothing, forecasting and prediction of discrete time series,* Englewood Cliffs, NJ: Prentice-Hall, 1965.
29. H. Akaike, A new look at the statistical model identification, *IEEE Trans. Automat. Contr.,* **19**, 1974, 716–723.
30. A. A. El-Keid and X. Ma, Advancement of statistical based modeling techniques for short-term load forecasting, *Electric Power Systems Research,* **35**, 1995, 51–58.

X. Ma
Cegelec ESCA Corporation

A. A. El-Keib
The University of Alabama

H. Ma
ABB System Control

M. Kaddoura
The University of Alabama

SHORTWAVE BROADCASTING. See High-frequency broadcasting.

SHUNTS, CURRENTS. See Current shunts.