A Hybrid Model for Day-Ahead Price Forecasting

Lei Wu, *Member, IEEE*, and Mohammad Shahidehpour, *Fellow, IEEE*

***Abstract*—This paper presents a hybrid time-series and adaptive wavelet neural network (AWNN) model for the day-ahead electricity market clearing price forecast. Instead of using price series, one-period continuously compounded return series is used to achieve more attractive statistical properties. The autoregressive moving average with exogenous variables (ARMAX) model is used to catch the linear relationship between price return series and explanatory variable load series, the generalized autoregressive conditional heteroscedastic (GARCH) model is used to unveil the heteroscedastic character of residuals, and AWNN is used to present the nonlinear, nonstationary impact of load series on electricity prices. The Monte Carlo method is adopted to generate more evenly distributed random numbers used for time series and AWNN models to accelerate the convergence. Several criteria such as average mean absolute percentage error (AMAPE) and the variance of forecast errors are used to assess the model and measure the forecasting accuracy. Illustrative price forecasting examples of the PJM market are presented to show the efficiency of the proposed method.**

***Index Terms*—AMAPE, ARMAX, AWNN, day-ahead price forecast, GARCH, Monte Carlo, time series method, variance of forecast errors.**

NOMENCLATURE

Variables:

a_t	Noise process.
a_{ij}, b_{ij}	Translation and dilation parameters of wavelets.
c	Constant item in the GARCH model.
d	Differential order.
e	Bias of the output node in AWNN.
h_t	Conditional variance of residual series at time t.
i, j	Index.
k	AWNN training iteration index.
m	Number of input layer of AWNN.
MSE_T^k	Mean squared error (MSE) on testing set at iteration k for AWNN.
MSE_V^k	MSE on validation set at iteration k for AWNN.
$\overline{MSE}_V^k$	Average MSE on validation set until iteration k for AWNN.
n	Number of hidden nodes of AWNN.
O_t	Forecast electricity price at time t.
p, b, q	Orders of time series models.
W_i, W_j, W_{ij}	Weights of AWNN.
$\overline{R(d)}$	Average value of $R_t(d)$ series.
t, τ	Index of time series.
α_i, β_i	GARCH coefficients.
$\phi(B)$	Coefficient function for price series.
$\varphi(B)$	Coefficient function for load series.
$\theta(B)$	Coefficient function for noise process.
ρ_k	Sample autocorrelation value of lag k.
ε_t	Residual at time t.
γ^k	Generalization factor of iteration k for AWNN.
$\overline{\gamma}^k$	Average generalization factor until iteration k for AWNN.
σ_E^2	Variance of forecast errors.
σ_V^k	Standard deviation (STD) of MSE on validation set until iteration k for AWNN.
σ_γ^k	STD of generalization factor until iteration k for AWNN.

Constants:

B	Backshift operator, means $B^p \cdot P_t = P(t-p)$.
$F^{-1}(\cdot)$	Inverse probability distribution function.
L_t	Load at time t.
P_t	Electricity price at time t.
R_t	One-period log return at time t.
$R_t(\tau)$	Log return of τ-period at time t.
T	Study period.

Manuscript received December 19, 2008; revised June 21, 2009. First published February 08, 2010; current version published July 21, 2010. Paper no. TPWRS-01020-2008.

The authors are with the Electrical and Computer Engineering Department, Illinois Institute of Technology, Chicago, IL 60616 USA (e-mail: lwu10@iit.edu; ms@iit.edu).

Digital Object Identifier 10.1109/TPWRS.2009.2039948

I. INTRODUCTION

WITH the introduction of restructuring to the electric power industry, the price of electricity is becoming the focus of all activities in electricity markets. Price forecasting techniques are used for bidding purposes and hedging against volatilities. With a good next-day market clearing price forecast, a market participant would be able to delineate better financial decisions. That is, a power producer can develop appropriate strategies to maximize its payoff and a consumer can minimize its utilization cost [1].

Electricity has distinct characteristics as compared to other commodities; it cannot be stored economically and transmission congestion may prevent a free exchange of power among control areas. Thus, electricity price series can exhibit a major volatility and the application of forecasting methods prevailed in other commodities can pose large errors in electricity price forecasting. For different applications, price forecasting can be categorized into very short-term (several minutes to few hours), short-term (few days), midterm (few months) and long-term (few years). In this paper, we focus on the day-ahead market clearing price forecasting in electricity markets. A reasonable forecasting algorithm could capture important properties of electricity prices such as spikes, mean reversion, seasonality, and fat tails.

There are several possible methods for the day-ahead electricity price forecasting. The first approach is based on fundamental models, which is to simulate the exact physical model of power system components and apply algorithms which consider physical characteristics of power networks. Such an approach would express electricity prices based on marginal generation costs with the consideration of transmission congestion, losses, and other ancillary service requests in power markets [1], [2]. Fundamental models account for the impact of physical capacity of power plants and transmission lines, and demand characteristics and fluctuations. The most difficult issue here is that a fundamental model may require significant amounts of real-time data on power systems which can result in a complex computation process.

The second approach is based on mathematical finance models which were originally developed and widely used for stock and interest rate markets. Skantze *et al.* [3] developed a dynamic model to describe price series based on interactions of supply bids and load demands, which are simulated as mean reverting processes and exponential functions, respectively. The market clearing price was calculated as the cross section of aggregated supply and demand bid curves. A diffusion model for electricity prices, based on stochastic models for supply and demand curves, was proposed by Barlow [4] to exhibit price spikes. By fitting to historical data with the maximum likelihood estimation, Barlow concluded that the proposed diffusion model could provide a better fit than models with jumps. The finance models reflect the electricity price volatility caused by supply and demand, thus are suitable for option valuation and risk assessment purposes. The drawback is that it is very difficult to incorporate physical characteristics of power systems, such as transmission congestions, network losses, and contingencies, into mathematical finance models which may cause a discrepancy between the finance model solution and the real-time status of power systems.

The third approach is based on game theory models, which are particularly focused on the impact of bidders' strategic behavior on electricity prices. Bolle [5] used supply and demand functions as instruments to determine market equilibria. It showed that under a certain condition, an equilibrium exists for every finite spread of autonomous demand, including small, nonstrategically acting consumers. Lower bounds of market prices were also computed. Ruibal *et al.* [32] used three oligopoly models, Bertrand, Cournot, and supply function equilibrium, to obtain closed form expressions for expected values and variances of hourly electricity prices as well as average prices, with the consideration of hourly demand volatility and random outages of generating units. It concluded that the introduction of competition may decrease expected prices but variances may actually increase. Li *et al.* [6] used the incomplete information game theory to analyze competitions among transmission-constrained GENCOs. Based on the solution of incomplete information game, the Bayesian Nash equilibrium which represents locational marginal prices (LMPs) was calculated using the ISO's security-constrained economic dispatch (i.e., a market clearing model). The problem was an iterative bilevel process, with the upper process representing individual GENCOs' bidding strategies to maximize payoffs and the lower process solving the ISO's market clearing problem for calculating LMPs. One presumption in the game theory is that all players are rational, which is not always the case in electricity markets. Other drawbacks are represented by time-consuming processes for calculating Nash equilibria.

The fourth method is based on regression models, which include time-series, and artificial intelligence models such as artificial neural network (ANN) and fuzzy logic. Regression models relate electricity price fluctuations to historical prices and other explanatory factors such as temperature, time of day, load demand, etc. The autoregressive integrated moving average (ARIMA) time series was applied to load forecasting [7], [8]. Contreras *et al.* [9] proposed an ARIMA model for the day-ahead electricity price forecasting, with examples from the Spanish electricity market and the PJM electricity market. Backshift factors were introduced into the ARIMA model for reflecting the seasonality of prices. In addition, it reported that the inclusion of explanatory variables, such as demand and available daily production of hydro units, may improve price forecasts with large spikes. Conejo *et al.* [10] proposed a hybrid wavelet transform and ARIMA model to forecast day-ahead electricity prices. Wavelet transform was used to decompose historically ill-behaved price series into a set of well-behaved constitutive series. Then, ARIMA was adopted for each constitutive series and the inverse wavelet transform was used to reconstruct a better forecast result. Gao *et al.* [11] used a three-layer back propagation network to forecast day-ahead market clearing prices (MCPs) and market clearing quantities (MCQs). Historical MCPs and MCQs, system loads, fuel prices, power exchanges, weather, and season indices that influenced market prices were used for network training, validating, and forecasting. Holiday data were pretreated as weighted average of following normal days. Price spikes were truncated for a better accuracy of normal prices. Mandal *et al.* [31] explored an ANN model based on similar days method to forecast day-ahead electricity prices in the PJM market. The impact factors were historical loads and price series of the days similar to those of the forecast day. Pindoriya *et al.* [33]

proposed an adaptive wavelet-ANN by using the Mexican hat wavelet as the activation function for hidden-layer neurons of feed-forward ANN. The forecast results showed good accuracy compared to ARIMA, multi-layer perceptrons, radial basis functions, and fuzzy neural networks. Swanson *et al.* [29] compared various adaptive and nonadaptive, linear and potentially nonlinear models, and concluded that the hybrid models, by grouping the multivariate adaptive linear and nonlinear models, dominate other models and provide least confused forecasts. The regression method is simple and computationally efficient. It requires detailed and correct historical data and proper model (orders of time series model, details of layer structure of artificial intelligence model) for adaptation and forecasting.

Price spikes were addressed regularly in electricity price forecasting. Previously proposed models preprocess, ignore, or truncate price spikes in order to get better forecasts for hourly electricity prices. Zhao *et al.* [30] developed a data mining based approach to forecast price spikes. Feature selection techniques were described to identify attributes relevant to the occurrence of spikes, including system demand, system supply, seasonality, scheduled interchanges, and dispatchable loads. Two algorithms, i.e., support vector machine and probability classifier, were applied as spike occurrence predictors. Guan *et al.* [12] proposed that price spikes might be the result of power suppliers' strategic gaming behavior. A prisoner's dilemma matrix game was formulated, and the notion of opportunistic tacit collusion was introduced to explain strategic bidding behaviors in which suppliers withhold a generation capacity from the market to drive prices up.

This paper focuses on day-ahead electricity market clearing price forecasts with large volatilities. Return series is used instead of original price series to capture statistical properties. The ARMAX model is used to seize the linear relationship between price return series and load series, the GARCH model is used to present heteroscedastic characteristics of residuals, and AWNN is used to show nonlinear and nonstationary impacts of system loads on electricity prices. Our proposed hybrid model provides a 24-h MCP forecast of the next day based on historical data and forecast explanatory factors. To illustrate our model, price forecasts in the PJM electricity market [13] are calculated and discussed.

The rest of the paper is organized as follows. Section II describes the proposed model and its solution methodology. Section III presents illustrative examples to show the proposed model applied to the PJM electricity market. Conclusions are provided in Section IV.

II. Forecasting Methodologies

In competitive power markets, electricity prices appear to be highly volatile with nonstationary means and variances. The prices correspond to seasonality, weekdays, weekends, and holidays. Furthermore, power system loads as an explanatory factor would impact electricity prices. Here, we propose a hybrid model that uses ARMAX time series to forecast linear relationships between electricity price return series and system load series, followed by GARCH to simulate nonconstant variances of residuals, and applies AWNN to forecast nonlinear and nonstationary impacts of system loads on electricity prices as shown in Fig. 1. In Fig. 1, the ARMAX model is first applied for forecast using historical prices and other explanatory data. Then, forecast results from the ARMAX model are used as input to the GARCH model. The GARCH model output includes nonconstant residuals. The ARMAX and GARCH results are combined and used as input to AWNN. The AWNN output is the final price forecast.

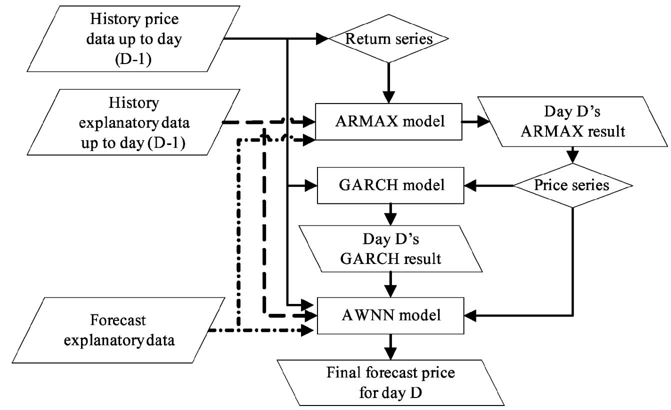

Fig. 1. Flowchart of the proposed forecasting procedure.

A. Price Data Preprocess

A key feature that distinguishes the electricity price series from other time series lies on the assumption that the electricity price is very volatile. Return series has more attractive statistical properties and thus is easier to handle than price series [14]. There are several common definitions of asset return [15]. In this paper, we use the continuously compounded return, also called log return. The one-period continuously compounded return series is defined as

$$R_t = \ln(P_t) - \ln(P_{t-1}) = \ln\left(\frac{P_t}{P_{t-1}}\right). \tag{1}$$

The continuously compounded return enjoys several advantages over the price series. First, the multiperiod return is simply the sum of one-period continuously compounded returns in (2):

$$\begin{aligned} R_t(\tau) &= \ln(P_t) - \ln(P_{t-\tau}) = \sum_{j=0}^{\tau-1} [\ln(P_{t-j}) - \ln(P_{t-j-1})] \\ &= R_t + R_{t-1} + \cdots R_{t-(\tau-1)}. \end{aligned} \tag{2}$$

Second, statistical properties of the continuously compounded return are more tractable. Fig. 2 shows the hourly electricity prices from January 1, 2006 to May 31, 2006 of the PJM market [13]. Fig. 3 shows the return series during the same period. Comparing the two figures, we learn that the mean value of return series is close to zero and the variability of return series over time looks more homogeneous than that of price series. Thus, statistical characteristics of return series may be better comprehended than those of price series. Therefore, we study the return series of assets in the stationary ARMAX model instead of price series. In the following, we refer to return series as the one-period continuously compounded return as in (1).

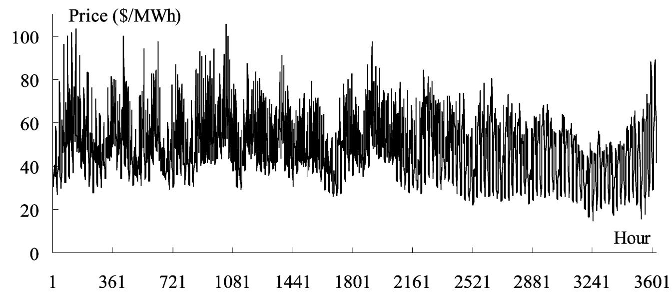

Fig. 2. Price series of PJM market from January 1, 2006 to May 31, 2006.

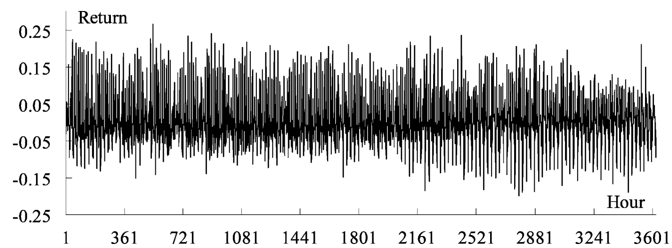

Fig. 3. Return series of PJM market from January 1, 2006 to May 31, 2006.

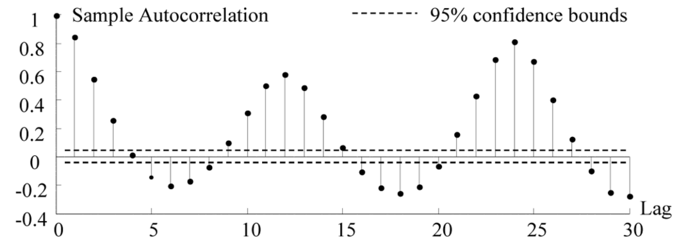

Fig. 4. ACF of price series of PJM market from January 1, 2006 to May 31, 2006.

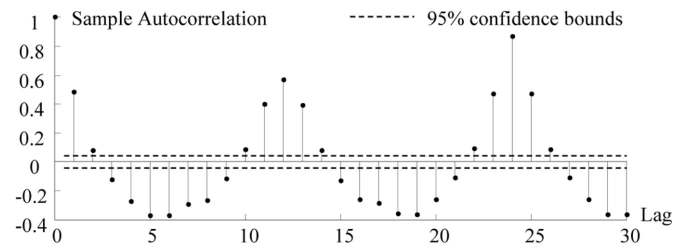

Fig. 5. ACF of return series of PJM market from January 1, 2006 to May 31, 2006.

B. ARMAX Time Series Model

ARMAX is a class of stationary stochastic model. The series in Fig. 2 shows nonequilibrium at a constant mean level, which means the original price series is nonstationary. Using suitable differences, a homogeneous nonstationary time series is transferred to stationary mixed autoregressive-moving average process shown in (3):

$$\phi(B) \cdot (1-B)^d \cdot \ln(P_t) = \varphi(B) \cdot L_t + \theta(B) \cdot a_t \tag{3}$$

where a_t is a Gaussian $N\left(0, \sigma_a^2\right)$ white noise process, $\phi(B) = 1 - \phi_1 \cdot B - \cdots - \phi_p \cdot B^p$, $\varphi(B) = 1 - \varphi_1 \cdot B - \cdots - \varphi_b \cdot B^b$, and $\theta(B) = 1 - \theta_1 \cdot B - \cdots - \theta_q \cdot B^q$.

A homogeneous nonstationary time series is transferred to a stationary time series by taking a proper degree of differencing. From (3), $\ln(P_t)$ is a homogeneous nonstationary time series. By applying a suitable difference of order d, i.e., $R_t(d) = (1-B)^d \cdot \ln(P_t)$, the new return series $R_t(d)$ is transferred to a stationary stochastic process and fitted in a standard ARMAX (p, q, b) model as

$$\phi(B) \cdot R_t(d) = \varphi(B) \cdot L_t + \theta(B) \cdot a_t. \tag{4}$$

The general ARMAX scheme is described as follows:

1) Model Identification: Use the autocorrelation function (ACF) and partial autocorrelation function (PACF) to identify the order of ARMAX (p, q, b) and the suitable difference order d. The autocorrelation of lag k is given as

$$\rho_k = \frac{\sum_{t=1}^{T-k} \left[R_t(d) - \overline{R(d)}\right] \cdot \left[R_{t+k}(d) - \overline{R(d)}\right]}{\sum_{t=1}^{T} \left[R_t(d) - \overline{R(d)}\right]^2}. \tag{5}$$

The stationary property implies that zeros of $\phi(B)$ lie outside the unit circle, which means that the ACF will die out quickly and rather linearly. Figs. 4 and 5 show the ACF of price series and return series from January 1, 2006 to May 31, 2006 in the PJM market, respectively. Dashed lines indicate the approximate upper and lower confidence bounds. That is, if the ACF at a certain lag is smaller than the 95% confidence bound, we assume there is no significant autocorrelation at that lag. After the first order difference (which is carried out in return series), the ACF dies out quickly after lag 2 with a damping sine-cosine wave; so the price series may follow the process with $d = 1$, which is the one-period return series as defined in (1). Thus inputs and outputs of ARMAX model are all price return series. Furthermore, (p, q, b) orders are determined by lags where ACF and PACF die out.

There are two alternatives to the ARMAX model to forecast a day-ahead electricity price return series. One is to build a single ARMAX model based on the entire series, and then consider a 24-step forecast for the next day. The other is to consider different hourly models. That is, dividing the entire series into 24 sets corresponding to different hours of a day, building ARMAX models based on different hourly series, and forecasting a one step ahead for each ARMAX model. The forecast results of the 24 ARMAX models constitute the day ahead forecast. Figs. 6–8 show the ACF of return series for the first hour and two peak hours, i.e., hours 10 and 20, of each day in the PJM market from January 1, 2006 to May 31, 2006. Comparing Figs. 6–8 with 5, we learn that, although hourly ACFs are not exactly the same, the return series ACF of the same-hour dies out more quickly than the return series of the entire day. Hence by considering the hourly-based series, the ARMAX model can better fit a series with proper AR and MA orders. Here we consider different hourly ARMAX models instead of a single ARMAX model for the entire series.

2) Parameter Estimation: After identifying a tentative model, the next step is to estimate the model parameters. The ordinary least squares (OLS) estimation is efficiently used to make inferences about parameters conditional on the adequacy

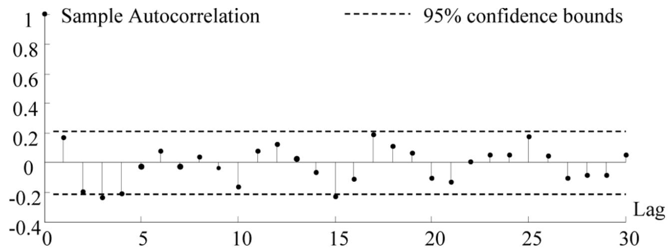

Fig. 6. ACF of return series for the daily hour 01:00 from January 1, 2006 to May 31, 2006.

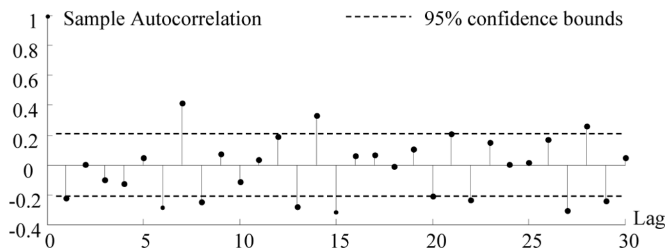

Fig. 7. ACF of return series for the daily hour 10:00 from January 1, 2006 to May 31, 2006.

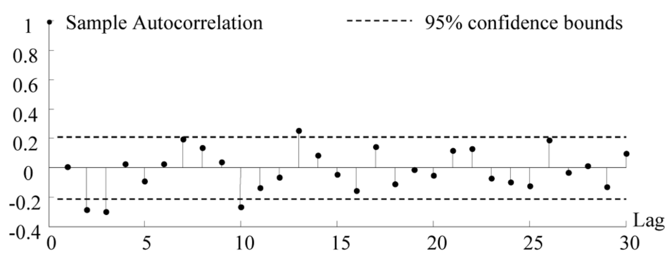

Fig. 8. ACF of return series for the daily hour 20:00 from January 1, 2006 to May 31, 2006.

of the model. Tray *et al.* proposed an iterative regression process to estimate parameters for stationary and nonstationary ARMA models [16].

3) Diagnostic Checking: Once parameters are estimated, we check the model adequacy for representing the series which is intended to reveal model inadequacies and consider improvements. One technique that can be used for diagnostic checking is overfitting, i.e., fitting a more elaborate model to show whether additions are needed. The other is the diagnostic check applied to residuals to show if there are any autocorrelations or partial autocorrelations among residuals. If the current model is inadequate, we return to step one and repeat the iterative procedure.

In conclusion, the time series model building is an iterative procedure. It starts with the model identification and the parameter estimation. After the parameter estimation which is used to analyze the adequacy of the model by diagnostic checking, iterative steps of the model building are repeated until a satisfactory model is obtained. One issue that needs to be considered here is that the fit of a regression model to the new data is nearly always worse than its fit to the original data. In this paper, the task is accomplished by separating the series into two parts, with the first portion of the data used for the model construction and the remaining portion for the evaluation of forecasting

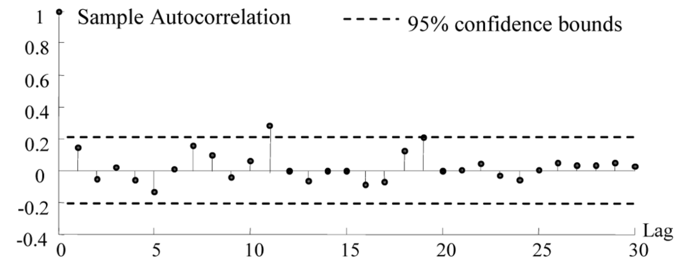

Fig. 9. ACF of residuals after the ARMAX is fit to the PJM market from January 1, 2006 to May 31, 2006.

capability. Copas [17] proposed the preshrunk predictor which anticipates the overfitting and gives predictions with a uniformly lower mean squared error.

C. GARCH—Conditional Heteroscedastic Model

It is generally agreed that both price series and return series present nonconstant deviations over time as demonstrated in Figs. 2 and 3. Thus the OLS estimator of ARMAX model coefficients is no longer asymptotically unbiased and consistent, when error terms are autocorrelated. Thus, the residual analysis is an important step in the regression analysis. GARCH models are widespread tools to deal with series conditional standard deviations [18]. A $\mathrm{GARCH}\,(p, q)$ is modeled as (6) where ν_t is a Gaussian $N\,(0, 1)$ white noise process and $h_t = Var\,(\varepsilon_t | \varepsilon_{t-1})$ represents the conditional variance of time t based on time $(t-1)$:

$$\begin{aligned} h_t &= c + \sum_{i=1}^{p} \alpha_i \cdot h_{t-i} + \sum_{i=1}^{q} \beta_i \cdot \varepsilon_{t-i}^2 \\ \varepsilon_t^2 &= \nu_t^2 \cdot h_t. \end{aligned} \tag{6}$$

The proposed GARCH model described in (6) can forecast beyond one day out. The input series, ε_t, considers the residual of ARMAX model as the actual price minus the forecast from the ARMAX process when the actual price is known, or forecast residual values from the GARCH model when foresting beyond one day out and the actual price is unknown. The prediction quality deteriorates as the number of predicted hours increases. After combining the results of ARMAX and GARCH models, forecast results incorporate the possibility of nonconstant error variance. The application of GARCH model is an iterative procedure which is similar to the ARMAX model. It includes iteratively order determination, parameter estimation, and model diagnostic checking [19].

Fig. 9 shows the ACF of residual series after the ARMAX model is fit to the PJM market from January 1, 2006 to May 31, 2006. This ACF figure suggests that the residual series fails to have a significant serial correlation. Fig. 10 shows the ACF of squared residuals. The autocorrelations at several lags are larger than bounds, which suggest that the residual series may have a conditional heteroscedasticity.

D. Adaptive Wavelet Neural Network (AWNN)

Considering nonlinear and nonstationary impacts of system loads on price does help improve predictions based on time se-

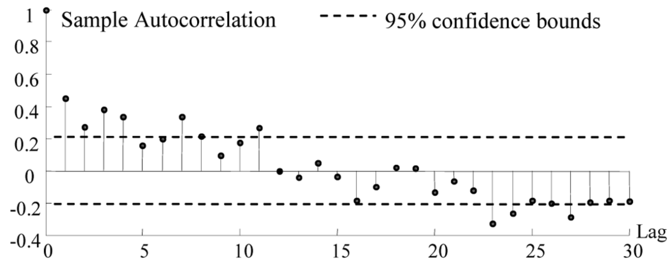

Fig. 10. ACF of squared residuals after the ARMAX is fit to the PJM market from January 1, 2006 to May 31, 2006.

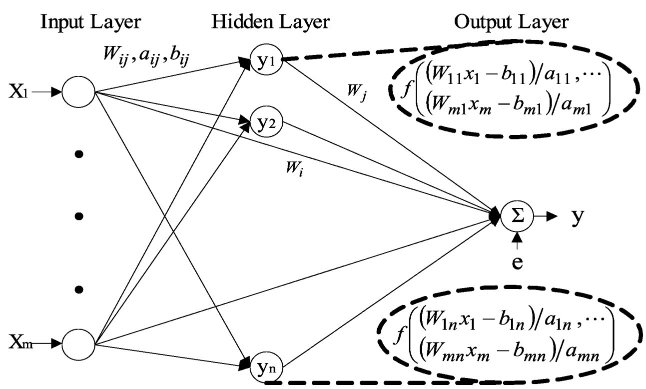

Fig. 11. AWNN model.

ries techniques, especially in the case of sky-high price spikes. AWNN represents a more modern estimation technique for a nonlinear relationship between electricity prices and system loads. AWNN introduces wavelets as activation functions of hidden neurons in traditional feed-forward neural networks with a linear output neuron. The unknown parameters of network include weights and dilation/translation factors which can be learned by gradient-type algorithms. The success of this configuration dwells on the fact that the wavelet network has universal and L^2 approximation properties and is a consistent function estimator [21], [22]. The AWNN structure is shown in Fig. 11.

The output of AWNN is computed as (7) where the multi-dimensional wavelet function $f(\,)$ is calculated by the tensor product of one-dimensional wavelets or some radial functions, and b_{ij}, a_{ij} are translation and dilation parameters:

$$y=\sum_{j=1}^{n} W_j\cdot f\left(\frac{(W_{1j}\cdot x_1-b_{1j})}{a_{1j}},\ldots,\frac{(W_{mj}\cdot x_m-b_{mj})}{a_{mj}}\right)+\sum_{i=1}^{m} W_i\cdot x_i+e. \quad (7)$$

The morlet wavelet shown in (8) is used here as the mother wavelet, and other wavelet functions are dilations and translations derived from this prototype mother wavelet. We compared two most popular mother wavelets given in the literature with numerical tests: Morlet and Mexican Hat. We found that Morlet always gives a better forecast solution. Since dilation parameters $a_{1j},\ldots,a_{mj}$ in (7) are already included for each input, the usage of 1.75 in (8) is an arbitrary coefficient which will not impact the forecasting accuracy. Multidimensional wavelet function is calculated by the tensor product in (9) and by the radial function in (10):

$$f(x)=\cos(1.75x)\cdot\exp\left(-x^2/2\right) \quad (8)$$

$$f\left((W_{1j}\cdot x_1-b_{1j})/a_{1j},\ldots,(W_{mj}\cdot x_m-b_{mj})/a_{mj}\right)=\prod_{i=1}^{m}\cos\left[1.75\cdot(W_{ij}\cdot x_i-b_{ij})/a_{ij}\right]\cdot\exp\left\{-\left[(W_{ij}\cdot x_i-b_{ij})/a_{ij}\right]^2/2\right\} \quad (9)$$

$$f\left((W_{1j}\cdot x_1-b_{1j})/a_{1j},\ldots,(W_{mj}\cdot x_m-b_{mj})/a_{mj}\right)=\cos\left(1.75\sqrt{\sum_{i=1}^{m}\left[(W_{ij}\cdot x_i-b_{ij})/a_{ij}\right]^2}\right)\cdot\exp\left(-\sum_{i=1}^{m}\left[(W_{ij}\cdot x_i-b_{ij})/a_{ij}\right]^2/2\right). \quad (10)$$

In the training process, the network learns by adjusting weights as well as translation and dilation parameters by using the product of gradient and learning rate. Two parameters, i.e., learning rate and momentum, are adjusted to accelerate the learning process [19]. The learning rate controls the size of each step for minimizing the objective function. The calculation of momentum term is to average the changes and determine the proportion of past changes that should be used for new values. Thereby, the term ensures that the search path on the average is in the downhill direction.

One of the critical issues in network training is overfitting which means that the network memorizes the training patterns and consequently loses the ability to generalize. That is, it fits the training set but cannot predict the fit for new data sets. Overfitting is especially a serious problem for price forecasting since there are many price spikes which are blended into historical prices.

In the training process, there is a point at which the training error continues to decrease while the generalization error starts to increase. The training process should stop at this point to avoid overfitting. Accordingly, in order to detect overfitting, the original data set is divided into three disjoint sets, i.e., training set, validation set, and generalization set. The training set is used to train the network model, and the validation set is used to estimate the generalization error while the generalization set is for forecasting. In this paper, two indicators are considered together to detect the point of overfitting. One is to check whether the MSE in each training iteration exceeds the average MSE plus its standard deviation, as expressed in (11):

$$MSE=\frac{1}{T}\sum_{t=1}^{T}(O_t-P_t)^2$$
$$MSE_V^k>\overline{MSE}_V^k+\sigma_V^k. \quad (11)$$

MSE implies an underlying quadratic loss function. It is the assumed loss function for which the conditional mean is the best, i.e., it is the minimum MSE forecast. The other is to measure the generalization factor, given in (12). Overfitting is detected at the training iteration k where $\gamma^k>\chi_\gamma^k$ with χ_γ^k given in (13).

This test always ensures that $\gamma^k \leq 1.0$ before overfitting occurs to make sure that the validation error is smaller than the testing error [22]:

$$\gamma^k = \frac{MSE_V^k}{MSE_T^k} \tag{12}$$

$$\chi_\gamma^k = \min\left\{\chi_\gamma^{k-1}, \overline{\gamma}^k + \sigma_\gamma^k, 1.0\right\}. \tag{13}$$

ARMAX considers different models for different hours. However, it is not the case for WANN. Two AWNN models are used for weekday and weekend forecasts. Both AWNNs for PJM price forecast studies contain 16 hidden nodes and one output. There are two AWNN input structures, one for predicting weekday prices and the other for weekend prices. For weekday forecasts, nine input data are used at hour t on day D, which are loads at hours $(t-1)$, t and $(t+1)$ on day D, historical prices at hour t on days (D-1) and (D-2), and forecast prices at $(t-3)$, $(t-2)$, $(t-1)$ and t based on the forecast results from ARMAX&GARCH models. For weekend forecasts, 14 input data are used at hour t on day D, which are loads at hours $(t-1)$, t and $(t+1)$ on day D, historical prices at hour t, $(t+1)$ and $(t+2)$ on day (D-1), historical prices at hour $(t-3)$, $(t-2)$, $(t-1)$ and t on day (D-2), and forecast prices at $(t-3)$, $(t-2)$, $(t-1)$ and t based on the forecast results from ARMAX and GARCH models. The selection of input features is based on correlation analysis, and the number of hidden nodes, 16, is based on the forecasting experience in the PJM market, which are all system specific.

The utilization and analysis of large data sets often represent a complex forecasting task. The process of feature selection is regarded as a reduction in dimensionality by choosing a subset of variables as features which are more relevant than others. It is likely that statistically less important components arise from noise which are not relevant to the intrinsic nature of the data. The dimensionality can be reduced by using statistical-based and artificial intelligence algorithms as principal component analysis, factor analysis, feature clustering, etc. [26], [27]. ARMAX considers hourly model structures on each day. However, the only two AWNN models are for weekdays and weekends, respectively.

E. Assessment of Forecasting

Several measurements are used to examine the accuracy of forecast results. The mean absolute percentage error (MAPE) index in (14) is considered here to evaluate the performance of forecast results. MAPE represents the absolute average prediction error between predictions and actual targets:

$$\text{MAPE} = \left(\frac{1}{T} \cdot \sum_{t=1}^{T} \frac{|O_t - P_t|}{P_t}\right) \cdot 100\%. \tag{14}$$

If the actual value is small, (14) will contribute large terms to MAPE even if the difference between actual and forecast values is small. In addition, if the forecast value is small and actual value is large, the absolute percentage error will be close to 100% [28]. In order to avoid the adverse effect of very small prices, the AMAPE defined in (15) is adopted and compared with those in the literature. For (14) and (15), period T could be hourly, daily, or weekly:

$$\text{AMAPE} = \left(\frac{1}{T} \cdot \sum_{t=1}^{T} \left[\frac{|O_t - P_t|}{\left(\frac{1}{T} \cdot \sum_{t=1}^{T} P_t\right)}\right]\right) \cdot 100\%. \tag{15}$$

Before using a single model or a combination of models to predict the future, we assume implicitly that there is a true model or a combination of models for a given series. However, the assumption is rarely accurate. Even if it is true, there is no guarantee that it will be selected as the best fit to the data. Thus the impact of model uncertainty on forecasts needs to be measured [23]. In this paper, we use the variance of forecast errors to measure this uncertainty. The smaller the variance, the less uncertain is the model or more accurate is the forecast results. The variance of error in a time span T is defined as

$$\sigma_E^2 = \frac{1}{T} \cdot \sum_{t=1}^{T} \left(\left[\frac{|O_t - P_t|}{\left(\frac{1}{T} \cdot \sum_{t=1}^{T} P_t\right)}\right] - \text{AMAPE}\right)^2. \tag{16}$$

F. Monte Carlo Random Number Generator

Series a_t in the ARMAX model, series ν_t in the GARCH model, and noise e in AWNN are regarded as Gaussian processes. Also, gradient algorithms used for the training of AWNN may be sensitive to initial conditions (i.e., weights, translation, and dilation factors). Evenly distributed random numbers would improve the universal search and the convergence process for the training of ARMAX, GARCH, and AWNN models. In this subsection, we describe the application of Monte Carlo method for generating more evenly distributed random numbers. We adopt antithetic variates, a variance reduction technique, to accelerate the convergence of results. A low-discrepancy Monte Carlo simulation method (lattice) is adopted with the possibility of accelerating the convergence rate. Random numbers in the lattice method are more evenly distributed. An n-point lattice rule of rank-r in dimension d is defined as

$$\left\{\sum_{i=1}^{r} \frac{k_i}{n_i} v_i \bmod 1,\ k_i = 0, 1, \ldots, n_i - 1\ i = 1, \ldots r\right\} \tag{17}$$

where $v_1, \ldots, v_r$ are linearly independent d-vector of integers. If we draw N independent samples according to the Monte Carlo simulation method, the iteration ends when the relative standard deviation is less than a predefined value (e.g., 95% relative standard deviation is given as $1.96 \cdot \sigma/\sqrt{N}$ where σ is the standard deviation and N is number of simulations). Usually N exceeds several thousands for a relatively small standard deviation. If the low-discrepancy method is used, the convergence will be accelerated from $O\left(1/\sqrt{N}\right)$ to nearly $O\left(1/N\right)$ and we can use a relatively smaller number of samples to reach the same convergence [24], [25]. The method of antithetic variates is a commonly used control variance method. It attempts to reduce the variance by introducing the negative dependence between pairs of replications. It is based on the observation that

$Y_i = F^{-1}(U_i)$ and $\tilde{Y}_i = F^{-1}(1 - U_i)$ have a distribution F but are antithetic to each other. The basic idea is that the variance after including the pair $\left(Y_i, \tilde{Y}_i\right)$ is smaller than just using the same size of random numbers generated by the ordinary Monte Carlo method [24], [25]. The emphasis on using the low-discrepancy method and the control variance method, lattice and antithetic variates, will generate more evenly distributed random numbers and accelerate the convergence.

III. Case Studies

The proposed model is applied to predict electricity prices for the PJM market. It is trained and tested using the data set from January 1, 2005 to December 31, 2006. Price data and other supporting information are given in [13]. In all case studies, for the day D's price forecast, the training set includes actual price and load data up to the day before day D. Also, the demand up to one hour ahead the price forecast hour are assumed known. As reported in the literature [34], [35], short-term load forecasting has reached a comfortable state of performance with 1%–2% error. Thus load forecast errors will have very limited impacts on price forecast errors reported in the paper. The cases are listed as follows:

- Case 1) Five single days, selected for each month from January to May 2006 including weekdays and weekends, are forecast and compared with those in [31].
- Case 2) Two weeks, February 1–7, 2006 representing a low load demand week and February 22–28, 2006 representing a high load demand week, are studied and compared with results reported in [31]. Furthermore, in order to investigate the trend in the prediction quality degradation as we increase the number of predicted hours, the forecast is studied based on day-ahead and represented for the whole month of February 2006. The same hybrid model is used for the entire month.
- Case 3) Four one-week periods, the first seven days in each month of February, May, August, and November representing different seasons in 2006, are studied.

The cases are discussed as follows.

1) Case 1: Five single days

For a fair comparison, same five test days used in [31] are considered, which includes January 20, February 10, March 5, April 7, and May 13 in 2006. Five individual models are developed and forecast prices for those five days, respectively. The forecasts are presented in Table I and compared with those in [31]. The AMAPE given in (15) that is also used in [31] is considered for the results in Table I. In Table I and thereafter, the improvement is calculated as: (the difference between the forecast results reported in [31] and the forecast results reported in this paper) divided by (the forecast results reported in [31]). A lower forecast error indicates that results are more accurate. Table I shows the daily AMAPE for selected days. Comparing the daily AMAPE of the proposed forecasts with those in [31], we learn that the consideration of price series deviations and nonlinear characteristics would help improve the forecast accuracy. Table II shows the variance of forecast errors as a measure of the model uncertainty. The smaller the variance, the less uncertain is the model, thus the more accurate are the forecasts. Notable improvements in AMAPEs and error

TABLE I
Comparison of Daily AMAPEs for Case 1

Day in 2006	Proposed Method	Results in [31]	Improvement
Jan. 20th	3.71%	6.93%	46.46%
Feb. 10th	2.85%	7.96%	64.20%
Mar. 5th	5.48%	7.88%	30.46%
Apr. 7th	4.17%	9.02%	53.77%
May 13th	4.06%	6.91%	41.24%

TABLE II
Comparison of Daily Error Variances for Case 1

Day in 2006	Proposed Method	Results in [31]	Improvement
Jan. 20th	0.0010	0.0034	70.59%
Feb. 10th	0.0015	0.0050	70.00%
Mar. 5th	0.0033	0.0061	45.90%
Apr. 7th	0.0013	0.0038	65.79%
May 13th	0.0015	0.0049	69.39%

TABLE III
Comparison for April 7 for the Effect of GARCH ($/MWh)

Hour	Actual Price	Proposed method	Without GARCH	Hour	Actual Price	Proposed method	Without GARCH
1	34.33	35.76	35.79	13	64.92	61.91	61.90
2	31.85	35.38	35.92	14	62.50	57.80	58.80
3	31.09	34.49	36.36	15	59.00	51.29	53.32
4	30.32	32.24	32.23	16	53.05	51.10	50.82
5	31.59	31.28	30.67	17	50.13	47.27	47.27
6	38.02	42.41	43.30	18	45.95	45.95	45.85
7	60.59	66.46	65.99	19	49.77	50.38	50.92
8	67.06	68.60	68.88	20	62.25	62.25	62.86
9	68.91	68.82	70.09	21	69.08	69.88	70.57
10	68.72	67.50	67.40	22	63.33	60.96	60.76
11	66.15	66.97	66.97	23	46.45	46.05	46.70
12	70.37	71.90	66.20	24	42.44	40.06	39.87

variances of the proposed method as compared to [31] are shown in the last column of Tables I and II. Table III compares the forecast results on April 7 using the proposed method and the method excluding GARCH (that is, only ARMAX and AWNN are used and GARCH is bypassed). Better results are obtained at most hours with GARCH, and the best one occurs at hour 12 with the actual price of 70.37$/MWh, which is 71.90$/MWh as compared to 66.20$/MWh without GARCH. The forecast results with GARCH are slightly worse at hours 4,7, 14,15, 23, and the worst one occurs at hour 15, with the actual price of 59.00 $/MWh, which is 51.29$/MWh as compared to 53.32$/MWh without GARCH. The final daily AMAPE of the proposed model is 4.17% as compared with the one without GARCH 4.72%, which indicates that by including GARCH, the final AMAPE for the whole day is improved by about 13.19% (i.e., $(4.72\% - 4.17\%)/4.17\%$).

Figs. 12 and 13 show the day-ahead price forecasts for chosen days of April 7 and May 13, 2006. It is reported in [31] that the prediction is particularly inaccurate for the morning and evening peaks of April 7, 2006. Fig. 12 shows that the prediction is improved by applying the proposed method for the morning and evening peaks, and the daily AMAPE is improved from 9.02% to 4.17%, with a 53.77% improvement. In Fig. 13, the prediction is improved for the daily and evening peaks especially at hours 9:00 and 21:00. The forecast inaccuracy is improved here especially for very large peaks, which are essential to GENCOs' bidding strategies.

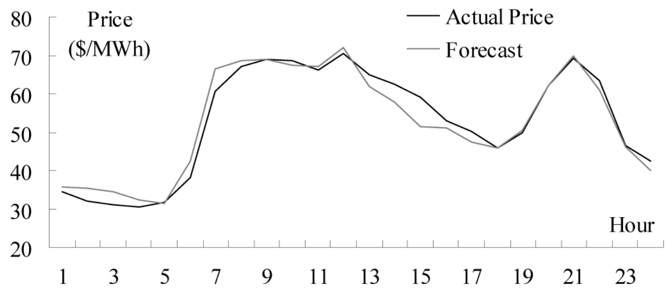

Fig. 12. Actual and forecast prices for April 7, 2006.

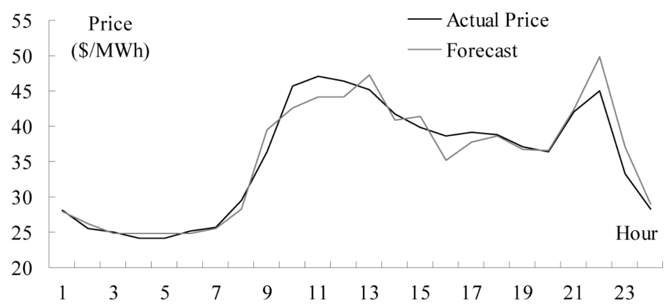

Fig. 13. Actual and forecast prices for May 13, 2006.

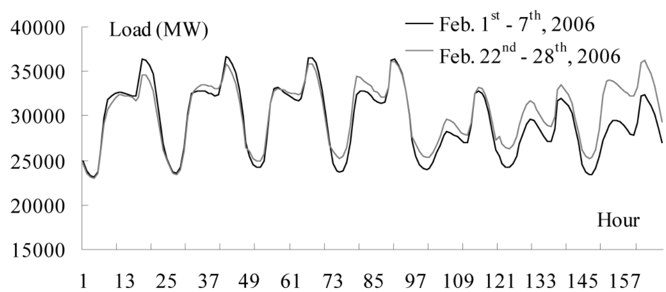

Fig. 14. Actual load profile for February 1–7 and February 22–28, 2006.

2) Case 2: Two weeks with low and high load demands

Two weeks, February 1–7, 2006 representing a low demand week and February 22–28, 2006 representing a high demand week, are studied. Load profiles for these two weeks are shown in Fig. 14.

Tables IV and V compare the performance of the proposed method with that of [31]. For these two weeks, both the AMAPE and error variance are notably improved. Especially in the week of February 22–28, 2006, the AMAPE is decreased from 8.88% to 5.01%, which is improved by 43.58%. With the proposed forecasting method, the maximal daily AMAPE of 8.25% is even smaller than the weekly AMAPE of 8.88% as reported in [31]. These forecasts are based on day-ahead and represented for one week. That is, the actual prices of the PJM market on February 1, 2006 are used as inputs to forecast the price on February 2, 2006 and so on; however, only one hybrid model is used for the entire week.

The weekly forecasts for February 1–7, 2006 are presented in Fig. 15 which shows price forecasts obtained from the proposed model follow the trend of actual market prices. When price spikes appear during peak hours in the last three days of this week, the proposed model outperforms that in [31] by

TABLE IV
COMPARISON OF WEEKLY AMAPEs FOR CASE 2

Week (2006)		Proposed Method	Results in [31]	Improvement
Feb. 1st 7th	Max. Daily AMAPE	8.21%	11.32%	27.47%
	Min. Daily AMAPE	2.98%	5.94%	49.83%
	Weekly AMAPE	5.27%	7.66%	31.20%
Feb. 22nd-28th	Max. Daily AMAPE	8.25%	12.37%	33.31%
	Min. Daily AMAPE	2.11%	5.66%	62.72%
	Weekly AMAPE	5.01%	8.88%	43.58%

TABLE V
COMPARISON OF WEEKLY ERROR VARIANCES FOR CASE 2

Week (2006)	Proposed Method	Results in [31]	Improvement
Feb. 1st 7th	0.0037	0.0066	43.94%
Feb. 22nd-28th	0.0025	0.0047	46.81%

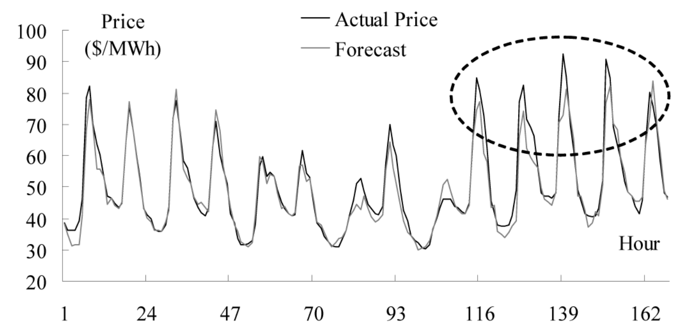

Fig. 15. Actual and forecast prices for February 1–7, 2006.

capturing actual prices more closely as highlighted in Fig. 15. The weekly AMAPE proposed by the hybrid model is 5.27% as compared to 7.66% reported in [31], which shows a 31.20% improvement.

Similarly, weekly forecasts of February 22–28, 2006 are presented in Fig. 16. According to Fig. 14, load patterns for the first five days are similar in these two weeks. Large loads occur in later days of the week, which increase electric prices. In comparison with [31], the proposed weekly AMAPE is decreased from 8.88% to 5.01%, with a 43.58% improvement. Also, the proposed accuracy of forecasts is improved especially for the last day of the week as highlighted. The quality of prediction may deteriorate as the number of predicted hours increases since the same set of hybrid models will be used to for the entire week instead of developing a set of new models for each day.

Table VI gives the forecast error variance, corresponding to the same hours on different days of the week of February 1–7, 2006, to measure the model uncertainty. The smaller the variance, the less uncertain is the model and the more accurate are the forecasts. Variances for off-peak hours are much less than those of peak hours. The largest variances, which are less than 0.008, occur during hours 5–7. The second largest variances of 0.005 occur during hours 19–21 which are the evening. Variances less than 0.002 occur at off-peak hours.

The prediction quality deteriorates gradually as the number of predicted hours increases. In order to show the degradation of the forecast based on the increasing number of predicted hours, forecasts based on the single day-ahead model for the whole month of February 2006 are presented in Fig. 17 and

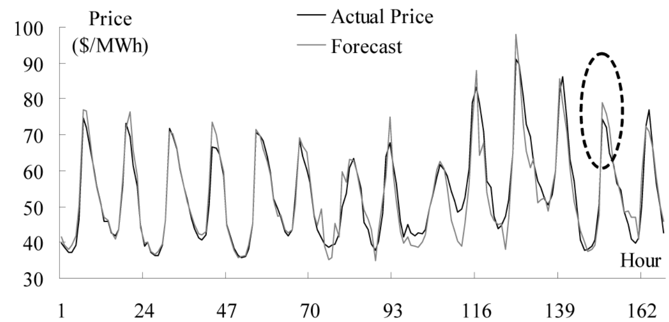

Fig. 16. Actual and forecast prices for February 22–28, 2006.

TABLE VI
VARIANCE OF HOURLY FORECAST ERRORS FOR FEBRUARY 1–7, 2006 IN CASE 2

Hour	Variance	Hour	Variance	Hour	Variance
1	0.00042	9	0.00202	17	0.00024
2	0.00133	10	0.00242	18	0.00023
3	0.00254	11	0.00374	19	0.00566
4	0.00199	12	0.00044	20	0.00175
5	0.00436	13	0.00031	21	0.00492
6	0.00559	14	0.00029	22	0.00084
7	0.00759	15	0.00052	23	0.00062
8	0.00145	16	0.00120	24	0.00067

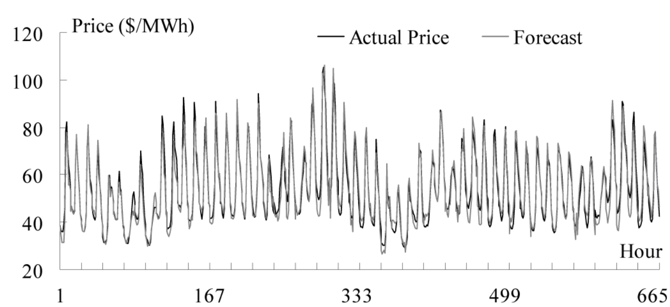

Fig. 17. Actual and forecast prices for February 1–28, 2006.

TABLE VII
COMPARISON OF MAPES FOR THE MONTH OF FEBRUARY 1–28, 2006

Test Week (2006)	Feb. 1^{st}-7^{th}	Feb. 8^{th}-14^{th}	Feb. 15^{th}-21^{st}	Feb. 22^{nd}-28^{th}
Weekly AMAPE	5.27%	6.09%	6.26%	6.70%

Table VII. The results show that the forecast performance deteriorates slowly as the number of predicted hours increases. As shown in Table IV, with a newly developed model for the week of February 22–28, 2006, the weekly AMAPE is 5.01%. In comparison, the same period results based on the monthly forecast would increase to 6.70% as shown in Table VII. This test shows that a single model can be applied to several weeks.

The hybrid model would be updated as required if the current forecasting model is not accurate enough. Accordingly, we train the new model offline using the historical data. The technology of incremental learning would be applied for reducing the frequency of developing the new model. The technology would update the existing model in an incremental fashion to accommodate the new data without compromising the performance [36].

3) Case 3: Four one-week periods for seasons in 2006

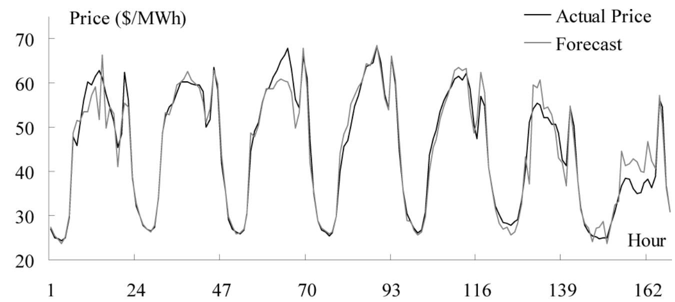

Fig. 18. Actual and forecast prices for May 1–7, 2006.

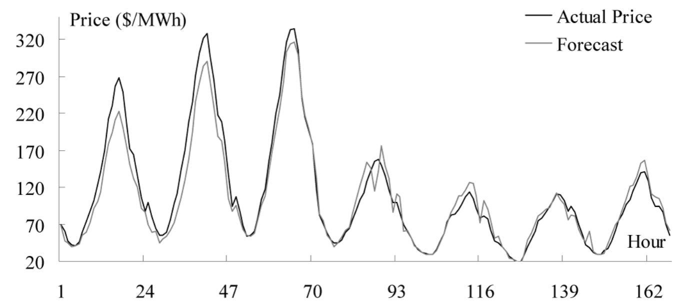

Fig. 19. Actual and forecast prices for August 1–7, 2006.

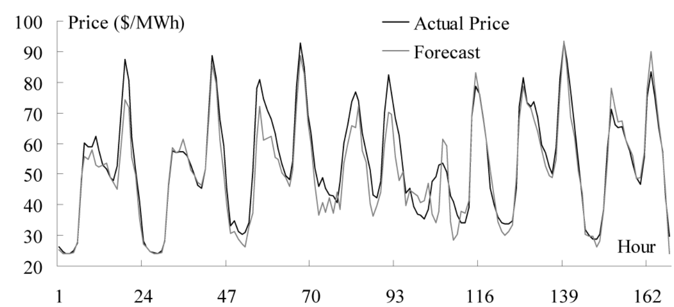

Fig. 20. Actual and forecast prices for November 1–7, 2006.

TABLE VIII
FORECAST RESULTS OF WEEKLY FORECAST PERFORMANCE OF CASE 3

Test Week (2006)	Feb.	May	Aug.	Nov.
Weekly AMAPE	5.27%	4.64%	11.28%	7.70%
Max. Daily AMAPE	8.21%	8.75%	16.04%	11.94%
Min. Daily AMAPE	2.98%	1.79%	5.59%	3.63%
Error Variance	0.0047	0.0026	0.0126	0.0050

In this case, four weeks, the first seven days in each month of February, May, August, and November representing different seasons in year 2006, are studied. Table VIII shows the forecasts. The smallest weekly AMAPE occurs in the week of May, and the largest occurs in the week of August, which is mainly due to large spikes occurring during that week. Figs. 18–20 show price forecasts for weeks of May, August, and November. The different price curves show seasonal variant volatilities for the PJM electricity market. Daily prices in May and November appear to be more periodic than those in August, and the two weekly forecasts follow the actual prices more closely.

In the week of August 2006, spikes as large as over 350$/MWh occur on August 3, 2006. Another two spikes occur on peak hours of August 1 and 2, 2006. Fig. 19 shows that

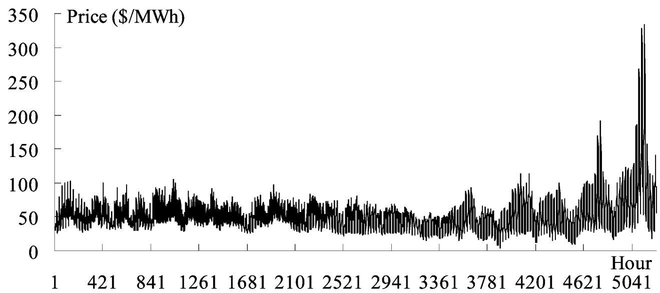

Fig. 21. Actual price from January 1, 2006 to August 7, 2006.

TABLE IX
FORECAST FOR THREE LARGE SPIKES FOR THE WEEK OF AUGUST 1–7, 2006

Hour	Actual price ($/MWh)	Forecast price($/MWh)	Error
17:00 on Aug. 1st, 2006	267.99	223.02	16.78%
17:00 on Aug. 2nd, 2006	328.04	289.83	11.65%
17:00 on Aug. 3rd, 2006	333.91	316.61	5.18%

the proposed price forecasts follow the trend of actual market price spikes when spikes are directly related to the load values, although the spike quantity is not predicted very accurately. Fig. 21 shows the actual price series from January 1, 2006 to August 7, 2006. The spikes, as large as 267.99$/MWh, occurring on August 1, 2006 are the first large spikes in 2006. Thus, spike forecasts for August 1, 2006 would hardly capture actual spikes, which is mainly due to the lack of historical spike information when the model is built. Although forecast spike magnitudes are smaller than those of actual prices, the proposed algorithm does predict the spike vertex to some level of accuracy. Reference [31] did not report any results on such large spikes in the PJM market. In fact, the literature on price spike forecasts is very limited. Zhao *et al.* [30] forecasted price spikes of the Queensland Electricity market in June 2004 with forecast errors of 20%. Table IX gives the hourly forecasts at peak hours on the first three days. The largest hourly spike forecast error, which is 16.78%, occurs at 17:00 on August 1, 2006, and the smallest, as low as 5.18%, occurs at 17:00 hours on August 3, 2006.

Compared with [30], the accuracy of the proposed price spike forecasts is sufficiently good considering the extreme volatility of spike signals. Furthermore, forecasts capture actual prices very well when prices suddenly drop by about 50% on August 4, 2006 from over 350$/MWh to as low as 157$/MWh. Other explanatory factors that may affect electricity prices such as weather, available ancillary services, power exchanges, and the availability of generators and transmission lines are not included in the proposed day-ahead price forecasting model. Such factors are less important for price forecasting in most situations, and their inclusion may cause overfitting or even deteriorate the accuracy of the forecasting method. With modern SCADA/EMS systems, one can monitor/evaluate the power system status and perform very short-time price forecasts with the proposed model, which improves the forecasting of spikes.

IV. CONCLUSIONS

Several case studies are chosen for the PJM electricity market with various levels of price spikes to test and validate the proposed model. The proposed hybrid time-series and AWNN model, composed of linear and nonlinear relationships of prices and explanatory variables, improves the performance of forecast results. The usage of one-period continuously compounded return series and AWNN has an advantage of modeling nonstationary electricity prices, especially price spikes. The use of AWNN as a consistent function estimator and two overfitting detection indications diminishes the overfitting issue when including history spikes in the training data set for price spikes forecasting. It is observed that the proposed hybrid model outperforms the literature, and considering the extreme volatility of the spike signal, the price spike forecast accuracy level of the proposed model is sufficiently good as compared with those reported in the literature.

REFERENCES

[1] M. Shahidehpour, H. Yamin, and Z. Y. Li, *Market Operations in Electric Power Systems*. New York: Wiley, 2002.
[2] Y. Fu and Z. Li, "Different models and properties on LMP calculations," in *Proc. IEEE Power Eng. Soc. General Meeting*, Montreal, QC, Canada, Jun. 2006.
[3] P. Skantze, M. Ilic, and J. Chapman, "Stochastic modeling of electric power prices in a multi-market environment," in *Proc. IEEE Power Eng. Soc. Winter Meeting*, Singapore, Jan. 2000, pp. 1109–1114.
[4] M. T. Barlow, "A diffusion model of electricity prices," *Math. Fin.*, vol. 12, pp. 287–298, 2002.
[5] F. Bolle, "Competition with supply and demand functions," *Energy Econ.*, vol. 23, pp. 253–271, 2001.
[6] T. Li and M. Shahidehpour, "Strategic bidding of transmission-constrained GENCOs with incomplete information," *IEEE Trans. Power Syst.*, vol. 20, no. 1, pp. 437–447, Feb. 2005.
[7] G. Gross and F. D. Galiana, "Short-term load forecasting," *Proc. IEEE*, vol. 75, no. 12, pp. 1558–1573, Dec. 1987.
[8] M. T. Hagan and S. M. Behr, "The time series approach to short term load forecasting," *IEEE Trans. Power Syst.*, vol. 2, no. 3, pp. 785–791, Aug. 1987.
[9] J. Contreras, R. Espínola, F. J. Nogales, and A. J. Conejo, "ARIMA models to predict next-day electricity prices," *IEEE Trans. Power Syst.*, vol. 18, no. 3, pp. 1014–1020, Aug. 2003.
[10] A. J. Conejo, M. A. Plazas, R. Espínola, and A. B. Molina, "Day-ahead electricity price forecasting using the wavelet transform and ARIMA models," *IEEE Trans. Power Syst.*, vol. 20, no. 2, pp. 1035–1042, May 2005.
[11] F. Gao, X. H. Guan, X. -R. Cao, and A. Papalexopoulos, "Forecasting power market clearing price and quantity using a neural network method," in *Proc. IEEE Power Eng. Soc. Summer Meeting*, Seattle, WA, Jul. 2000, pp. 2183–2188.
[12] X. Guan, Y. C. Ho, and D. L. Pepyne, "Gaming and price spikes in electric power markets," *IEEE Trans. Power Syst.*, vol. 16, no. 3, pp. 402–408, Aug. 2001.
[13] PJM Electricity Market Data. [Online]. Available: http://www.pjm.com/.
[14] J. Y. Campbell, A. W. Lo, and A. C. MacKinlay, *The Econometrics of Financial Markets*. Princeton, NJ: Princeton Univ. Press, 1997.
[15] R. S. Tsay, *Analysis of Financial Time Series*, 2nd ed. New York: Wiley, 2005.
[16] R. S. Tsay and G. C. Tiao, "Consistent estimates of autoregressive parameters and extended sample autocorrelation function for stationary and nonstationary ARMA models," *J. Amer. Statist. Assoc.*, vol. 79, no. 385, pp. 84–96, Mar. 1984.
[17] J. B. Copas, "Regression, prediction and shrinkage," *J. Roy. Statist. Soc.*, ser. B 45, no. 3, pp. 311–354, 1983.
[18] T. Bollerslev, "Generalized autoregressive conditional heteroscedasticity," *J. Econ.*, vol. 31, pp. 307–327, 1986.
[19] A. P. Engelbrecht, *Computational Intelligence: An Introduction*. New York: Wiley, 2002.
[20] Q. Zhang and A. Benveniste, "Wavelet networks," *IEEE Trans. Neural Netw.*, vol. 3, no. 6, pp. 889–898, Nov. 1992.

[21] J. Zhang, G. G. Walter, Y. Miao, and W. N. W. Lee, “Wavelet neural networks for function learning,” *IEEE Trans. Signal Process.*, vol. 43, no. 6, pp. 1485–1497, Jun. 1995.
[22] A. Röbel, The Dynamic Pattern Selection Algorithm: Effective Training and Controlled Generalization of Backpropagation Neural Networks, Tech. Univ. Berlin, Berlin, Germany, 1993, Tech. Rep. 93–23.
[23] C. Chatfield*, Time-Series Forecasting*. London, U.K.: Chapman & Hall/CRC, 2000.
[24] L. Wu, M. Shahidehpour, and T. Li, “Stochastic security-constrained unit commitment,” *IEEE Trans. Power Syst.*, vol. 22, no. 2, pp. 800–811, May 2007.
[25] G. Paul*, Monte Carlo Simulation Method in Financial Engineering*. New York: Springer, 2003.
[26] I. Guyon and A. Elisseeff, “An introduction to variable and feature selection,” *J. Mach. Learn. Res.*, vol. 3, pp. 1157–1182, Mar. 2003.
[27] A. Jain and D. Zongker, “Feature selection: Evaluation, application, and small sample performance,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 2, pp. 153–158, Feb. 1997.
[28] H. Y. Yamin, S. M. Shahidehpour, and Z. Li, “Adaptive short-term electricity price forecasting using artificial neural networks in the restructured power markets,” *Elect. Power Energy Syst.*, vol. 26, pp. 571–581, 2004.
[29] N. R. Swanson and H. White, “A model selection approach to real-time macroeconomic forecasting using linear models and artificial neural networks,” *Rev. Econ. Statist.*, vol. 79, no. 4, pp. 540–550, Nov. 1997.
[30] J. Zhao, Z. Dong, X. Li, and K. Wong, “A framework for electricity price spike analysis with advanced data mining methods,” *IEEE Trans. Power Syst.*, vol. 22, no. 1, pp. 376–385, Feb. 2007.
[31] P. Mandal, T. Senjyu, N. Urasaki, T. Fundabashi, and A. K. Srivastava, “A novel approach to forecast electricity price for PJM using neural network and similar days method,” *IEEE Trans. Power Syst.*, vol. 22, no. 4, pp. 2058–2065, Nov. 2007.
[32] C. M. Ruibal and M. Mazumdar, “Forecasting the mean and the variance of electricity prices in deregulated markets,” *IEEE Trans. Power Syst.*, vol. 23, no. 1, pp. 25–32, Feb. 2008.
[33] N. M. Pindoriya, S. N. Singh, and S. K. Singh, “An adaptive wavelet neural network-based energy price forecasting in electricity markets,” *IEEE Trans. Power Syst.*, vol. 23, no. 3, pp. 1423–1432, Aug. 2008.
[34] T. Senjyu, P. Mandal, K. Uezato, and T. Funabashi, “Next day load curve forecasting using hybrid correction method,” *IEEE Trans. Power Syst.*, vol. 20, no. 1, pp. 102–109, Feb. 2005.
[35] D. W. Bunn, “Forecasting loads and prices in competitive power markets,” *Proc. IEEE*, vol. 88, no. 2, pp. 163–169, Feb. 2000.
[36] R. Polikar, L. Udpa, S. S. Udpa, and V. Honavar, “Learn++: An incremental learning algorithm for supervised neural networks,” *IEEE Trans. Syst., Man, Cybern. C. Appl. Rev.*, vol. 31, no. 4, pp. 497–508, Nov. 2001.

Lei Wu (M’07) received the Ph.D. degree in electrical engineering from Illinois Institute of Technology, Chicago, in 2008.

Presently, he is a visiting Professor at Illinois Institute of Technology. His research interests include power systems restructuring, reliability, and economics.

Mohammad Shahidehpour (F’01) is the Bodine Distinguished Professor in the Electrical and Computer Engineering Department at Illinois Institute of Technology (IIT), Chicago. He is an honorary professor at the North China Electric Power University and the Sharif University in Iran.

Dr. Shahidehpour is the VP for Publications of the IEEE Power & Energy Society. He is the recipient of an honorary doctorate from the Polytechnic University of Bucharest.