Electricity Price Forecasting in Ontario Electricity Market Using Wavelet Transform in Artificial Neural Network Based Model

Sanjeev Kumar Aggarwal, Lalit Mohan Saini*, and Ashwani Kumar

Abstract: Electricity price forecasting has become an integral part of power system operation and control. In this paper, a wavelet transform (WT) based neural network (NN) model to forecast price profile in a deregulated electricity market has been presented. The historical price data has been decomposed into wavelet domain constitutive sub series using WT and then combined with the other time domain variables to form the set of input variables for the proposed forecasting model. The behavior of the wavelet domain constitutive series has been studied based on statistical analysis. It has been observed that forecasting accuracy can be improved by the use of WT in a forecasting model. Multi-scale analysis from one to seven levels of decomposition has been performed and the empirical evidence suggests that accuracy improvement is highest at third level of decomposition. Forecasting performance of the proposed model has been compared with (i) a heuristic technique, (ii) a simulation model used by Ontario's Independent Electricity System Operator (IESO), (iii) a Multiple Linear Regression (MLR) model, (iv) NN model, (v) Auto Regressive Integrated Moving Average (ARIMA) model, (vi) Dynamic Regression (DR) model, and (vii) Transfer Function (TF) model. Forecasting results show that the performance of the proposed WT based NN model is satisfactory and it can be used by the participants to respond properly as it predicts price before closing of window for submission of initial bids.

Keywords: Multiresolution analysis, neural network, normal distribution curve, price forecasting, wavelet transform.

NOMENCLATURE

The list of acronyms used throughout the paper is provided below:

A_1,.. A_7	Approximation time series of level 1, 2, 3, 4, 5, 6, and 7 respectively
ARIMA	Autoregressive Integrated Moving Average
CWT	Continuous wavelet transform
D_1,.. D_7	Detailed time series of level 1, 2, 3, 4, 5, 6, and 7 respectively
DR	Dynamic regression
DSPS	Dispatch scheduling and pricing software
DWT	Discrete wavelet transform
FFNN	Feed forward neural network
FT	Fourier transform
GARCH	Generalized autoregressive conditional heteroskedastic
HOEP	Hourly Ontario Energy Price
IESO	Independent electricity system operator
MAPE	Mean absolute percentage error
MCP	Market clearing price
MLR	Multiple linear regression
MRA	Multiresolution analysis
NN	Neural Network
OD	Ontario demand
OEM	Ontario Electricity Market
PDP	Pre-dispatch price forecast by IESO
QMF	Quadrature mirror filters
RMSE	Root mean square error
SSR	System status report
SSRFD	Forecast demand made by IESO in SSR
STFT	Short time Fourier transform
TF	Transfer function
TMD	Total market demand in Ontario
WT	Wavelet Transform

1. INTRODUCTION

Electricity price forecasting is essential for all the

Manuscript received February 29, 2008; accepted May 28, 2008. Recommended by Guest Editor Seung Ki Sul. The authors are thankful to Robert Doyle and Charanjeet Minhas from Customer Communications Department of Ontario IESO for providing the past SSR data from their data bank and for other useful suggestions.

Sanjeev Kumar Aggarwal, Lalit Mohan Saini, and Ashwani Kumar are with the National Institute of Technology, Kurukshetra 136119, India (e-mails: vasusanjeev@yahoo.co.in, lmsaini@rediffmail.com, ashwa_ks@yahoo.co.in).
* Corresponding author.

participants in deregulated electricity markets as a risk management technique [1]. Using this, generation companies can maximize their profits by bidding effectively, whereas; bulk electricity customers can optimize their load schedules. Accurate price forecasting also helps in pricing a range of derivative securities for hedging. Thus forecasting electricity prices is important yet complex task because price series is a non-stationary and highly volatile series with non-constant mean, variance and significant outliers [2]. The main methodologies followed for price forecasting are: (i) Production-cost models [3] and, (ii) Statistical models [4-10]. Production-cost models are complicated to implement and require detailed system operation data like participants' competitive bidding behavior, generation units' data, the transmission network data, hydrological conditions, fuel prices and demand forecasts. The market operator usually adopts them as they provide detailed insights into system prices. Statistical methods require lesser amount of data as compared to production cost models, and may be adopted by generation companies in preparing their bidding strategies and by large consumers for demand side management. Two main techniques of the statistical modeling are: (i) Nonlinear models, and (ii) Linear models. Artificial intelligence based methods employing NN, which can model complex nonlinear relationship between input variables and output variable, have been proposed by different researchers [4-6]. Whereas; the available linear models are MLR [7], AR [8], ARIMA [2,9], multivariate time series models are TF and DR [2,9] and GARCH [10]. Hybrid nonlinear neuro-fuzzy model [6] and game theoretic model [11] have also been reported.

With the help of WT, a signal can be decomposed into a parsimoniously countable set of basis functions at different time locations and resolution levels, known as MRA analysis. These decomposed series can be used to unfold inner characteristics of the signal and hence for more precise forecasting. This issue has been addressed in this work from price forecasting perspective and applied to OEM [12] as a test case system. The rest of the paper is organized as follows: In Section 2, utilization of WT in a price-forecasting problem and contribution of the present work is explained and in Section 3, a brief introduction of WT is presented. Section 4 covers the introduction to OEM and selection process of input variables for price forecasting models. In Section 5, detailed analysis of effect of WT on price signal and its statistical properties is investigated. Various price-forecasting models employed in this work are described in Section 6. Section 7 consists of the results analysis and comparison of their forecasting accuracy. Section 8 is the conclusion.

2. PROBLEM FORMULATION

A price signal exhibits much richer structure than load series and signal-processing techniques like FT, WT are good candidates for bringing out hidden patterns in price series [13]. In order to tackle the problem of non-stationary price series, wavelets have been utilized because they can produce a good local representation of the signal in both time and frequency domains. WT is used for multi-scale analysis of the signal and decomposes the time series signal into one low-frequency sub-series (approximation part) and some high-frequency sub-series (detailed part) in the wavelet domain. These constitutive series have better statistical properties than original price series and hence better forecasting accuracy can be achieved by their appropriate utilization. In WT based models, first of all WT is applied to the price series, prediction is made in the wavelet domain using a predictive model like regression model [2,14], time series model [15,16], or NN [17,18], and then inverse WT is applied to obtain the actual predicted value in time domain. During the process, there may be some loss of information. Moreover, due to characteristics of the some high frequency detailed series, these series cannot be predicted accurately [17]. In order to handle these issues, in this work, inputs to the forecasting model are a combination of original time domain and wavelet domain variables. The main difference is that in other models [2,14-18], prediction is made in the wavelet domain, whereas; in this work prediction is directly made in the time domain using both time domain and wavelet domain input variables. Since wavelet domain variables have been used in addition to the time domain variables, so there is no possibility of loss of information. Other input variables like oil price, capacity shortfall have also been included to improve the forecasting accuracy.

Since, relation between price and its influencing variables is non-linear, therefore; NNs are well suited for this problem because of their ability to model the complex and non-linear relationship involved in price forecasting. The main focus and contribution of this paper is to improve the forecasting accuracy of NN model using WT directly in time-domain and assess the effect of different decomposition levels of price on forecasting accuracy. Effect of decomposition levels from one to seven has been evaluated. It has been empirically proved that by combining both wavelet domain and time domain variables in a single framework forecasting accuracy of a nonlinear model like NN can be improved and third level of decomposition has been found to be best for appropriate forecasting. Forecasting performance has been compared with a heuristic technique, a simulation model used by Ontario's IESO, MLR model, NN model. Neuro-fuzzy [6], ARIMA, TF and

DR [9] models have also been included to present the detailed comparison of linear and non-linear models. Thus, in addition to presenting a new method of price forecasting, detailed analysis of the various price-forecasting methodologies has also been presented. The performance of the proposed model has been found to be satisfactory and it has real world application since it utilizes publicly available information, which can be easily collected by the participants before they need to submit bids in the market and also gives sufficient time to the participants to respond.

3. WAVELET TRANSFORM (WT)

The conventional FT gives the spectral contents of a signal $f(t)$, but it gives no information regarding wherein time those spectral components appear This is suitable only for dealing with signals that do not evolve with time, i.e. stationary signals. The STFT provides the time information by computing different FTs for consecutive time intervals, and putting them together. Consecutive time intervals of the signal are obtained by truncating the signal using a sliding windowing function $W(t)$. STFT gives a fixed resolution at all times. Once the window is chosen, the resolution is set for both time and frequency. Wide analysis window gives poor time resolution and good frequency resolution and vice-versa. WT overcomes the preset resolution of the STFT by using a variable length window. Analysis windows of different lengths are used for different frequencies [19-21]. The function used to window the signal is called the wavelet $\psi(t)$. The CWT of $f(t)$ with respect to a wavelet $\psi(t)$ is defined as:

$$W(a,b)=\frac{1}{\sqrt{|a|}}\int f(t)\cdot\psi^{*}\left(\frac{t-b}{a}\right)dt, \qquad (1)$$

where a and b are real numbers known as scale or dilation variable and time shift or translation variable respectively and * denotes complex conjugation. The normalizing factor $1/\sqrt{|a|}$ ensures that the energy stays the same for all a and b. The CWT offers time and frequency selectivity. The segment of $f(t)$ that influences the value of $W(a,b)$ for any (a,b) is that stretch of $f(t)$ that coincides with the interval over which $\psi_{a,b}(t)$ has the bulk of its energy. This windowing effect results in the time selectivity of the CWT. The frequency selectivity of the CWT can be understood on the basis of its interpretation as a collection of linear, time-invariant filters with impulse responses that are dilations of the mother wavelet reflected about the time axis. For a large value of a, the associated filter has a frequency response centered at a low frequency value and the smaller the value of a, the more the band pass shifts to a higher frequency. At small values of a, the CWT possesses good time resolution and poor frequency resolution. The opposite is true for large values of a. Therefore, CWT is suitable for non-stationary signals, in which rapidly varying high-frequency components are superimposed on slowly varying low-frequency components such as seismic signals, load and price signals. Inverse CWT is also possible and for a real valued $\psi(t)$ a function $f(t)$ can be recovered from its CWT $W(a,b)$ as:

$$f(t)=\frac{1}{C}\int_{a=0}^{\infty}\int_{b=-\infty}^{\infty}\frac{1}{|a|^{2}}W(a,b).\psi_{a,b}(t)\,da\,db. \qquad (2)$$

The region of support $W(a,b)$ is the set of ordered pairs (a,b) for which $W(a,b)\neq 0$. The CWT provides a redundant representation of the signal in the sense that the entire support of $W(a,b)$ need not be used to recover $f(t)$.Therefore, a two-dimensional sequence $D(m,n)$ is defined which is known as DWT of $f(t)$. In this, the dilation parameter $a=2^{m}$ where m is an integer and at any dilation 2^{m} the translation parameter $b=2^{m}n$ where n is an integer. This corresponds to sampling the coordinates (a,b) on a grid known as dyadic grid and the process is called dyadic sampling. The DWT is still the transform of continuous signal and the discretization is only in the variables a and b.

An efficient algorithm to implement the scheme using filters was developed by Mallat [21] and is available in [22]. This algorithm has two stages: decomposition and reconstruction. In first stage, the original signal is passed through two complementary filters and emerges as two signals: approximation (general trend component) and detail (high frequency component). Each of these signals has the same number of data points, and then these are down sampled by two, to get DWT coefficients. This decomposition can be iterated and successive approximations can be decomposed to many lower resolution components. In second stage, these components can be assembled back into the original signal. Thus, wavelet decomposition involves filtering and down sampling, and the wavelet reconstruction involves up sampling and filtering. The low and high

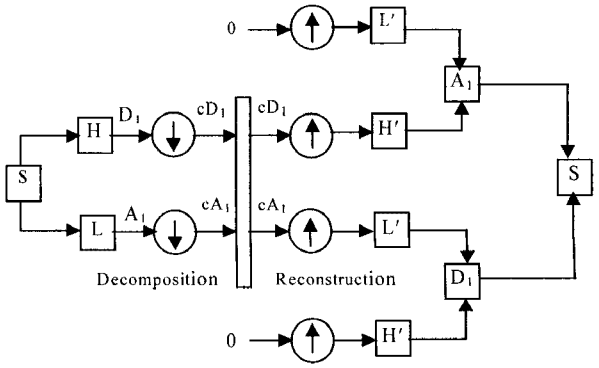


Fig. 1. Single stage decomposition and reconstruction.

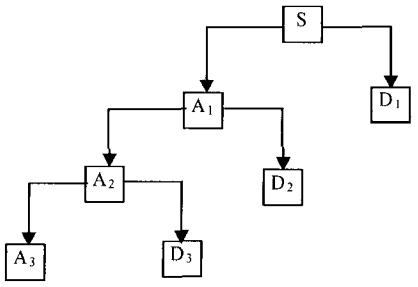


Fig. 2. A three level decomposition (S=A3+D3+D2+D1).

pass decomposition filters (L and H) and their associated reconstruction filters (L′ and H′) form QMF. A single stage decomposition and reconstruction has been shown in Fig 1 and corresponding three-stage decomposition has been shown in Fig. 2.

4. ONTARIO ELECTRICITY MARKET AND PRICE INFLUENCING VARIABLES

4.1. Electricity market introduction

OEM follows a real time single settlement design structure [12]. The MCP for electricity in OEM is based on bids and offers into the market from the participants and is set for each five-minute interval. In addition to the five-minute prices, each hour HOEP is determined by taking the average of twelve MCPs during an hour. Each day consists of 24 trading intervals of one-hour duration. Time-table for market operations is as follows: (i) *Pre-dispatch day (D-1)*: 6:00 AM – Window opens for submission of bids for dispatch day D, 11:00 AM – Initial bids must be submitted for dispatch day D. (ii) *Dispatch day (D)*: A registered market participant may revise bid data two hours prior to the beginning of each dispatch hour. This HOEP has been predicted in this study and the proposed model is able to provide the forecasted HOEP information before submission of initial bids on day D-1. Year wise statistical properties of HOEP are given in Table 1. Following the definition of historical volatility [9], the daily logarithmic return y_t for all market prices can be calculated as:

$$y_t = \ln(p_t) - \ln(p_{t-24}), \quad (3)$$

where p_t is the price information at time t, and p_{t-24} is the price information 24 hours before time t. Historical price volatility (σ) is defined as the standard deviation of y_t over a specified period of time. It can be observed from Table 1 that usual level of volatility is quite high in OEM and therefore, HOEP prediction is difficult.

4.2. Input variable selection

Price is dependent on many independent variables. Following hourly variables have been considered in this work: (i) past HOEP, (ii) IESO's forecasted energy demand, (iii) forecasted energy excess, (iv) forecasted capacity excess (v) crude oil prices, and (vi) temperature. Past HOEP data have been taken from [12], and variables (ii) to (iv) have been taken from SSR published by IESO. Crude oil price and temperature data have been collected from [23] and [24] respectively. Correlation analysis [25] has been used to select explanatory variables and correlation coefficients of the variables with HOEP have been presented in Table 2. It is evident that past HOEP, forecasted energy demand, energy excess, capacity excess and crude oil prices show good linear correlation with the price, whereas; temperature does not exhibit significant correlation with the price.

Table 1. Statistical properties of HOEP.

Year	2002	2003	2004	2005
Mean	52.0	54.0	49.9	68.5
Minimum	0	11.54	5.25	8.6
Maximum	1028.4	548.5	340.4	639.9
Standard deviation	46.08	35.9	21.9	40.7
Historical Volatility (σ)	0.41	0.49	0.36	0.40

Table 2. Correlation analysis of price drivers (July 2004 - June 2005).

S. No.	Category of variable	Explanatory Variable	Correlation coefficient with HOEP (D, t)
1.	Price variables	HOEP (D-n, t), n = 2,3,7,14	0.45, 0.43, 0.37, 0.4
2.		Crude oil price (D-2, t)	0.24, 0.25
3.	Weather variable	Temperature (D-2, t)	-0.03
4.	SSR data	Forecasted energy demand (D, t)	0.6
5.		Forecast energy excess (D, t)	-0.57
6.		Forecast capacity excess (D, t)	-0.63

D: day for which forecast is being made,
D-2: 2 days before the D-day, **t**: hour of the day

5. EFFECT OF WT ON HOEP

The statistical description of the price signal using the wavelet decomposition as a multi-scale analysis has been presented in this section. The main aspect of signal is what information is contained in the signal, and what pieces of information are useful. In order to do that wavelet decomposition was applied to the HOEP series of a particular week (March 1-7, 2004). Price signal is highly volatile and corrupted by the occasional spikes and follows a weekly-daily cycle with each sample of one-hour interval. Before performing wavelet decomposition, two issues need to be resolved: selection of mother wavelet and definition of the number of levels of decomposition. There are many wavelets that can be used in practice [22]. To choose the most appropriate wavelet, the attributes of the mother wavelet and the characteristics of the signal must be taken into account. Daubechies wavelets are the most appropriate for treating a non-stationary series [26]. For these families of wavelets, the smoothness increases as the order of the functions do; nevertheless, the support intervals also increase, which may cause the prediction to deteriorate. Therefore, low order wavelet functions are generally advisable. In this work, Daubechies wavelets of orders 1-4 have been considered and it was observed that decomposed series by wavelet of order 2 are more similar to original signal than by other orders and hence revealed the hidden patterns in HOEP in a more meaningful way.

It is advisable to select a suitable number of decomposition levels based on the nature of the signal. In this paper, based on the features of the price curve, one to seven levels of decomposition have been considered. Fig. 3 shows price signal's approximation, A_1 to A_7, and detailed parts, D_1 to D_7. Approximation curves correspond to low frequency bands and represent the trend of the price signal. On the other hand, detailed curves correspond to high frequency bands and contain the local short-period discrepancies in the price signal due to bidding strategies adopted by the participants. The level, whose approximation series is having the characteristics of normal distribution curve and yet closest to the shape of the original HOEP series, represents the filtered version of the original signal in a better way than the others. Table 3 presents skewness and kurtosis characteristics [25] of these series. Skewness is a measure of the symmetry of the data around the data mean and is zero for an ideal normal curve. Kurtosis is a measure of how outlier prone a distribution is. Kurtosis of normal distribution curve is three. From Fig. 3, it can be observed that, A_1, A_2 and A_3 series are similar in shape to the original signal than the other approximation levels. But statistical properties of A_1 are almost

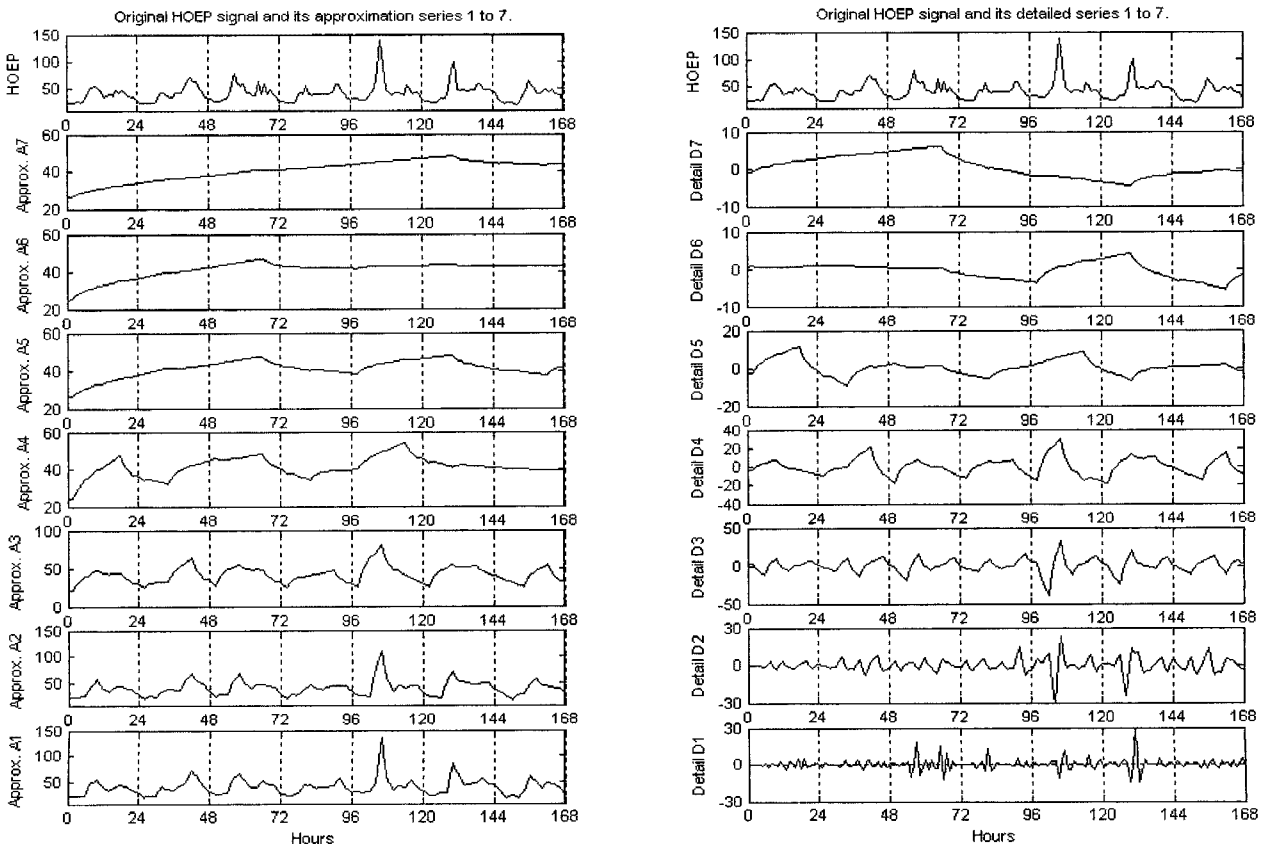


Fig. 3. Price and its constitutive series using WT (March 1-7, 2005).

Table 3. Statistical properties of constitutive series of HOEP.

Time series	Skewness	Kurtosis	Time series	Skewness	Kurtosis
HOEP series	1.49	10.14	HOEP series	1.49	10.14
Approx. A7	0.94	4.47	Detail D7	-0.23	3.54
Approx. A6	0.61	4.46	Detail D6	-0.83	6.64
Approx. A5	0.59	3.93	Detail D5	0.03	6.63
Approx. A4	0.74	3.96	Detail D4	0.33	5.75
Approx. A3	0.97	5.33	Detail D3	-0.27	4.42
Approx. A2	0.83	4.35	Detail D2	1.93	38.22
Approx. A1	1.15	6.31	Detail D1	-0.86	36.55

similar to HOEP series and therefore cannot be utilized for forecasting in a better way. Better statistical properties at levels A_4-A_7 are obtained at the cost of loss of some meaningful information. Thus third and second levels of decomposition describe the regular behavior of the price series in a more thorough and meaningful way and lead to improvement in statistical properties without any loss of information.

The corresponding details have been examined now. From Fig. 3, it can be seen that range of detailed parts is on lower side as compare to range of approximation parts. Further range of details D_4 to D_7 is lower than the range of D_1 to D_3, thus detailed parts D_4 to D_7 are mainly superficial random noise, as is evident from their Kurtosis characteristic as well. The details D_1 and D_2 contain useful high frequency information regarding abrupt changes in the signal. They exhibit signal irregularities and have similar range and mean values. Range of detail D_3 is higher than the details D_1-D_2 and it contains peaks corresponding to time localization of the peak price in the price series. Thus although, second decomposition level appears to be better than third decomposition level in terms of statistical characteristics of the approximation series, but due to the ability of D_3 to detect localized swings in price accurately, third level of decomposition is better positioned for price prediction and indeed empirical evidence regarding this has been provided in Section 7.1 after carrying out sufficient experiments.

6. MODELS AND METHODOLOGY FOR HOEP FORECASTING

This section describes the methods and models used and compared for price forecasting in this work. These are as follows:

6.1. Heuristic method (PM1)

For price forecasting, heuristic method assumes a strong and linear relationship between price and load, whose trends and levels repeat daily, weekly and seasonally. The expected price predicted by this method can be defined as:

$$P_{d,t} = P_{d-comp,t} \times \frac{L_{d,t}}{L_{d-comm,t}}, \quad (4)$$

where

$P_{d,t}$: the expected price for day d at hour t

$P_{d\text{-}comp,,t}$: the price at hour t of the comparable day of forecast day d

$L_{d,t}$: the forecast load for day d at hour t

$L_{d-comp,t}$: the load at hour t of the comparable day of forecast day d.

Comparable day has been assumed to be corresponding day of the previous week i.e., 7 days before the D-day. This has been taken to capture the weekly periodicity. Forecasted demand for D-day was taken as $L_{d,t}$ in (4) to predict the HOEP.

6.2. IESO model (PM2)

This is the model used by IESO and is simulation-based forecast information for the market participants. The DSPS determines the schedules and prices for energy in Ontario. The DSPS consists of several system and data analysis blocks, with a dc-based security-constrained optimal power flow block at its heart [27]. The pre-dispatch run of DSPS is used to provide the market participants with the projected schedules and prices for advisory purposes in advance. The information used by DSPS is dispatch data submitted by registered market participants, IESO's own load forecast, transmission system and ancillary services information. PDPs are available 3 hour, 2 hour and 1 hour before the actual dispatch corresponding to each trading interval. One hour before PDP information from IESO website was taken to make a comparative study with the other models.

6.3. MLR model (PM3)

A regression model is a mathematical equation that describes the relationship between two or more variables.The dependent variable (price) may be written as a linear function of a number of independent variables that are known [25,28]. A multiple linear system model that has n independent inputs (X_1, X_2, X_3, ... X_n) and one output Y at any time t can be described by the following equation:

$$Y = \beta_0 + \beta_1.X_1 + \beta_2.X_2 + \ldots + \beta_k.X_k + u, \quad (5)$$

where β_0, β_1, β_2, ... β_k are unknown regression

Table 4. Input variable set.

S. No.	Variable name	Time lag
P1	HOEP	D-2
P2	HOEP	D-3
P3	Crude oil price	D-2
P4	Forecasted demand	D
P5	Forecasted Energy	D
P6	Forecasted Capacity	D

coefficients, u is the independent and normal distributed random error term having mean zero and constant variance for different observations. The estimation of the regression coefficients is found using the least square estimation technique.

Variable segmentation has been applied to the input data set by separating the whole series into 24 time series, each one corresponding to an hour of the day. Then separate regression coefficients are calculated for each of the 24 time series. These coefficients for the predicted day are calculated using the data of the past ninety-one days. All input variables given in Table 4 were considered for prediction and the model has been implemented using Matlab7 [25].

6.4. NN model (PM4)

The steps for price forecasting procedure are:

Step 1: All variables given in Table 4 have been selected as input variables and complete data set has been divided into 24 separate hourly series. All inputs X_i and output Y_i are scaled to be in the range [-1, 1].

Step 2: a three-layered FFNN (Fig. 4) has been selected having six input nodes (equal to number of input variables), four hidden nodes with log-sigmoid transfer function and one output node with linear transfer function, for each series. So for predicting the each price point, 33 parameters need to be estimated. This network was trained with gradient descent with momentum training algorithm. The momentum constant and learning rate have been kept equal to 0.6 and 0.1 respectively.

Step 3: a moving window of past 91 days data has been used for NN training and estimating the parameters for D-day. The maximum epochs were set equal to 30,000 and regularization has been used to avoid over fitting. The performance function used is:

$$M = \gamma M_1 + (1-\gamma) M_2, \tag{6}$$

where γ is the performance ratio and has been set at 0.7. M_1 is mean squared errors for training data and M_2 is the mean squared of network weights and biases. The performance goal was kept at 0.05 [29].

6.5. Wavelet-NN model (PM5)

The basic forecasting steps followed are similar to PM4. The difference is in the input variable set, which

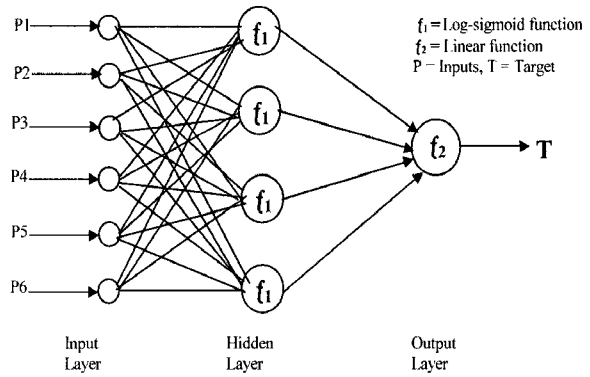


Fig. 4. Architecture of NN Model (PM4).

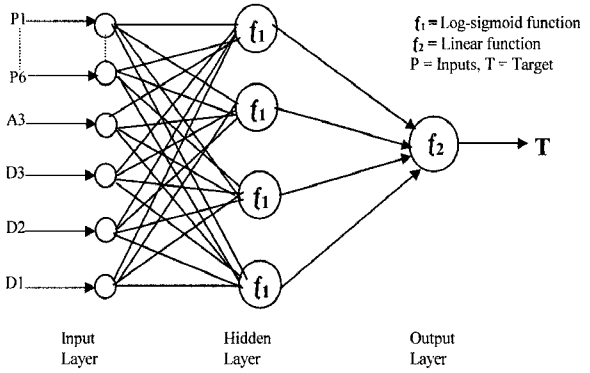


Fig. 5. Architecture of Wavelet NN Model at third decomposition level (PM5-D3).

consists of both time domain and wavelet domain variables (Fig. 5). To obtain wavelet domain variables, multi-scale analysis using wavelet transform has been applied. Daubechies wavelet of order 2 has been used for decomposition of HOEP. This model has been further explained in Section 7.

7. FORECASTING RESULT ANALYSIS AND DISCUSSION

MAPE has been adopted as the accuracy criteria to assess and compare the performance of the models.

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{X_t - X_f}{X_t}\right| \times 100, \tag{7}$$

where X_t is the actual value of the predicted variable and X_f is the forecasted value. N is the number of observations used for analysis. N = 24, for daily MAPE calculations.

Three time periods, each of two weeks duration, have been selected for testing the performance of the models [9]. The first test period (TP1) is from April 26 to May 9, 2004. This is spring's low demand period. The second test period (TP2) is from July 26 to August 8, 2004, which is summer peak-demand period. Winter high-demand period, from December 13-26, 2004, has been selected as third test period (TP3).

7.1. Effect of multi-scale analysis on the performance of Neural Network

A combination of time domain and wavelet domain variables has been used in PM5. The effect of different levels of decompositions levels was verified after conducting several experiments. First of all, HOEP series, consisting of data up to day D-2, was decomposed up to level 1, and then decomposed approximate part (A_1) and detailed part (D_1) were included as additional wavelet domain variables in the six time domain variable set given in Table 4. All parameters of NN viz. learning rate, momentum constant etc. were kept same as that in model PM4. In total seven experiments were conducted, by considering wavelet domain variables from decomposition level 1 to 7. The testing results corresponding to these experiments have been presented in Table 5 and average MAPE during these experiments has been shown in Fig. 6. It can be observed that, better forecasting accuracy has been achieved by including wavelet domain variables. Performance of models of PM5 is better than PM4. As far as effect of different decomposition levels is concerned, MAPE decreases up to level third (Fig. 6) and then it starts increasing. Thus results are best at third level of decomposition with PM5-D3 outperforming PM4 by 3.71%.

Table 5. Forecasting accuracy comparison at different decomposition levels.

Model	Additional HOEP wavelet domain variables in addition to P1-P6	MAPE (TP1)	MAPE (TP2)	MAPE (TP3)	Average MAPE
PM4	-	17.954	18.859	18.359	18.391
PM5-D1	A_1, D_1	17.497	18.464	17.813	17.925
PM5-D2	A_2, D_1, D_2	17.269	18.362	17.773	17.801
PM5-D3	**A_3, D_1, D_2, D_3**	**17.235**	**18.266**	**17.626**	**17.709**
PM5-D4	A_4, D_1, D_2, D_3, D_4	17.318	18.339	17.594	17.75
PM5-D5	A_5, D_1, D_2, D_3, D_4, D_5	17.375	18.458	17.661	17.831
PM5-D6	A_6, D_1, D_2, D_3, D_4, D_5, D_6	17.124	18.583	17.879	17.862
PM5-D7	A_7, D_1, D_2, D_3, D_4, D_5, D_6, D_7	17.129	18.48	18.482	18.03

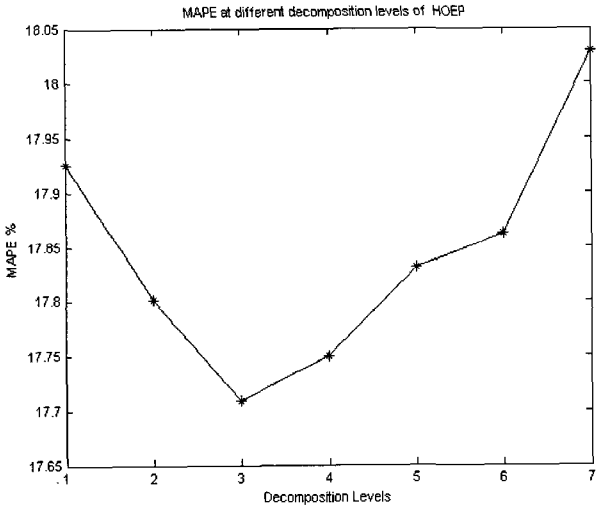


Fig. 6. MAPE at different decomposition levels of HOEP (Model PM5-D1 to PM5-D7).

7.2. Effect of variation of NN design on the performance of PM5

While designing an NN-based forecasting system, there are a large number of choices that need to be made. The following issues are important: (i) number of input neurons, (ii) number of hidden layers and neurons in each layer, (iii) number of output neurons. Since a system of 24 NNs in parallel has been used, therefore number of output neurons is one for each NN. The advantage of this method is that the individual networks are relatively small, and so they are not likely to be over fitted [30]. For number of hidden layers, it has been shown in [30], that one layer is sufficient to approximate any continuous function. Therefore one hidden layer has been selected to keep the model simple and smooth. The number of input neurons is equal to the number of input variables, which have been identified using linear correlation analysis, as already discussed in Section 4.2 for time domain variables and subsequently in Section 7.1 for wavelet domain variables. It has been proved empirically that third level of decomposition is the best; therefore number of input variables is equal to 10 (Fig. 5).

Gradient descent with momentum training algorithm updates the weights and biases in the direction of the negative gradient of the performance function. Its learning rate and momentum parameters need to be adjusted to achieve faster and smoother training. Since, in a price-forecasting problem, accuracy is more important than speed. Therefore, these parameters must be adjusted appropriately to get smooth training. The proposed model confirms this requirement has been shown in Fig. 7, where training feature is plotted. This shows that training of the model is smooth and free of bumps and undue oscillations. For designing NN based model, the frequently encountered problem is that of over fitting, which usually means estimating the training data well, but yet producing the poor forecasts on testing data. This problem has been handled in this work, by modifying the performance function by adding an additional penalty term to the mean squared error function. This term penalizes for the complexity of the model. The performance ratio (γ) controls the effect of this penalty term. Another aspect that needs to be

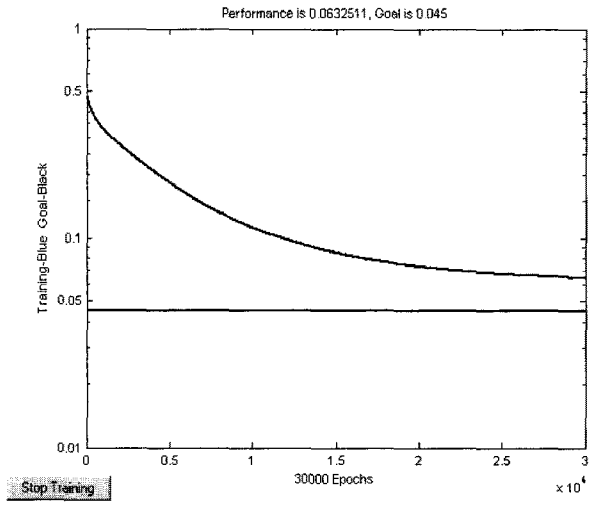


Fig. 7. Training feature for PM5.

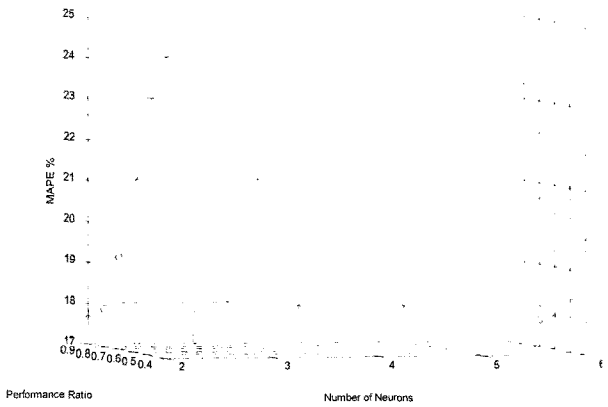


Fig. 8. Effect of NN parameter variation on model performance.

checked to avoid over fitting is the number of hidden neurons. There is no theoretical basis for the selection of hidden neurons [30]. In this work, a number of experiments were carried out to see the effect of variation of γ and number of hidden neurons on model performance. The number of neurons was varied from 2 to 6 and γ was varied from 0.4 to 0.9. The effect of these variations has been plotted in Fig. 8. When the number of neurons is low, this leads to under fitting and vice versa. When γ is increased from 0.4 to 0.9, in each case accuracy increases up to a certain value of γ and then it decreases. The optimum performance of the model has been observed, when the number of neurons is 4 and γ is 0.8. These parameters were then set to make the final comparison with the other models as explained in the Section 7.3.

7.3. Performance comparison with the other models

The performance of the five models explained in this work was compared with the earlier models proposed for OEM. Two other studies carried out for HOEP prediction are [6] and [9]. In [6], one-hour ahead HOEP forecasts have been reported using a nonlinear neuro-fuzzy model with average MAPEs varying from 19.83% to 24% with different configurations of NN and input variables. In [9], HOEP have been predicted, for the same six-weeks test period, using three linear models namely ARIMA, TF and DR. The overall MAPE comparison of proposed models for the same period has been presented in Table 6. The MAPE range of proposed model (PM5) is (15.21 - 18.72) whereas that of ARIMA, TF, and DR models is 13.6 – 21.5, 12.3 – 18.3 and 13.0 – 19.0 respectively as observed from [9]. Thus, the MAPE range of proposed model (PM5) is comparable to the existing models. PM5 has been developed after considering forecasted demand for D-1 day as well in addition to other variables in Table 4 and PM5-D3 in Table 5. The following points can be observed from Table 6.

Accuracy of PM5 is better than the other models in Week 1, Week 5 and Week 6. In five out of six weeks PM5 performs better than PM4. Overall, accuracy of PM5 is better than PM1, PM2, PM3 and PM4, by 29.65%, 13.36%, 3.63% and 4.79% respectively. Although PM1 looks simple, but it has the capacity to outperform other complicated models. In Week 2, PM1 has given the best performance. The performance of PM2 is better than the other models in Week 3.

But these results should be considered in the light of following points:

PM2 predicts HOEP one hour ahead only since one- hour pre-dispatch prices have been taken for comparison. ARIMA, TF and DR models have been individually identified and estimated for each of the six weeks [9] and this makes the models somewhat fragile [2]. On the other hand, in case of PM5 the identical model has been used for the full six weeks test period. Moreover forecasts by ARIMA, TF and DR for D- day are available on D-1 day around 11

Table 6. MAPE comparison for the six weeks test period.

Test period	Week no.	PM1	PM2	PM3	PM4	PM5
April 26 to May 2, 2004	**Week 1**	**21.70**	**23.78**	**16.26**	**16.56**	**15.21**
May 3-9, 2004	Week 2	17.80	25.26	19.23	19.34	18.62
July 26 to August 1, 2004	Week 3	22.92	10.41	17.69	17.45	17.91
August 2-8, 2004	Week 4	37.77	16.22	20.55	20.27	18.72
December 13-19, 2004	**Week 5**	**24.60**	**22.06**	**16.73**	**17.03**	**16.61**
December 20-26, 2004	**Week 6**	**24.55**	**23.51**	**18.54**	**19.69**	**18.02**
	Average	24.89	20.21	18.17	18.39	17.51

P.M. i.e., during last hour of trading on day (D-1) which is a time after closing of window for submission of initial bids for D-day. In case of PM5, the forecasting price is available between 9.30 A.M. to 11 A.M, which is a time before closing of window for submission of initial bids. Since, in Table 4, variables P1 to P3 are available on day D-2 and other variables P4 to P6 have been taken from the second SSR, which is usually available before 9 A.M. on D-1 day for D-day [12]. One of the principal factors affecting the accuracy of the forecasting models is the lead-time and as the lead-time increases, the accuracy of the forecast deteriorates [31]. The lead-time in case of PM5 is greater than the other works. Moreover, due to this time differentiation, PM5 can be easily utilized for bidding preparation by the participants, since it gives forecasts before closing of window for submission of initial bids.

Considering all these points, performance of model PM5 is satisfactory.

7.4. Validation of model PM5

Price forecasting is a complex task as a number of factors interact in an intricate manner and the associated uncertainty is very high. A reasonable forecasting technique can be properly validated if (i) its accuracy is either comparable or better than the well accepted methods; (ii) the comparison is based on the performance on test samples; (iii) the size of the test samples are adequate, which is two years [30,31]. Considering these, the comparison of the accuracy of all the five models was carried out for a period of two years (July 2003 to June 2005) and has been given in Table 7.

In this work, PM1 to PM4 have been selected as benchmark and it can be observed from Table 7 that PM5 performs better than PM1 and PM2 by 24% and 14.2% respectively. Performance of PM5 is better than the linear model PM3 by 2.4% and the ordinary NN model PM4 by 3.04%. Thus, the performance of PM5 is best among all the five models. The detailed comparison with the other forecasting models could not be carried out because no case study for price forecasting in OEM has considered such a long testing period. In this context, this should also be considered as a contribution of the present work. Also, reasonable efficiency gains have been achieved by applying WT to HOEP series, since performance of PM5 is better than PM4. All comparison has been made on the basis of 731 days test sample and is sufficient as it includes all seasonal, cyclical and other kind of variations like inflation, state of economy etc.

The other important questions need to be addressed are: whether the model considers all the important factors affecting the price and has the bad data detection and correction capabilities. The answer to both the questions is yes. Since all the major six variables affecting the HOEP have been included in the model and WT provides the filtration capability. The ability of the model to predict the turning point has been shown in Figs. 9 and 10. Although the model cannot predict the peaks accurately, never the less it predicts the trend and price movement very well.

The proposed method PM5 is easier to implement, utilize publicly available information only and provide forecast results before bidding time on (D-1)

Table 7. MAPE comparison for the two years test period.

Test period	PM1	PM2	PM3	PM4	PM5
1 July 2003 to 30 June, 2004	32.68	31.27	26.73	26.69	25.92
1 July 2004 to 30 June, 2005	26.85	21.42	19.63	19.96	19.33
Average	**29.77**	**26.35**	**23.18**	**23.33**	**22.62**

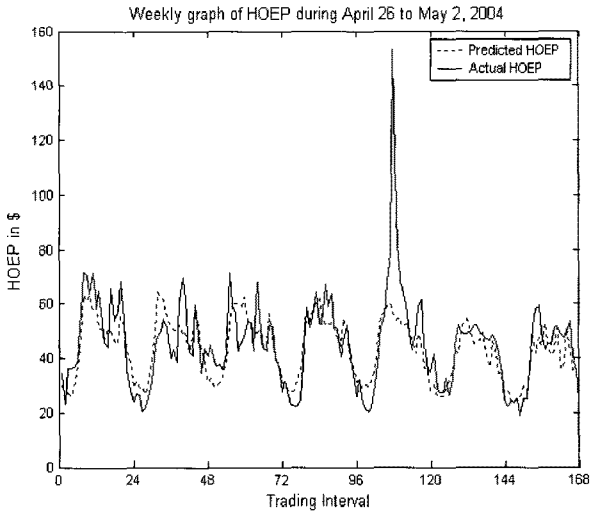


Fig. 9. Weekly predicted HOEP curve during week 1 (April 26 to May 2, 2004).

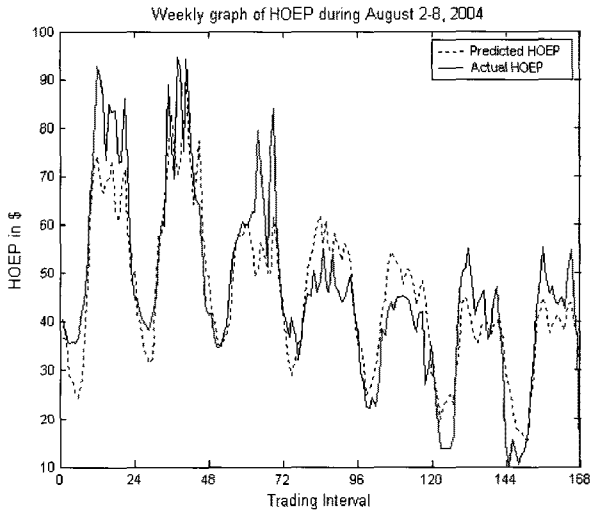


Fig. 10. Weekly predicted HOEP curve during week 4 (August 2-8, 2004).

day; whereas, PM2 predicts one hour before each settlement period on D-day. The lead-time in other works is also smaller than the PM5. Thus PM5 gives sufficient time to the participants to respond and achieves its objective of practical application.

8. CONCLUSION

In this work, a NN based method for price forecasting involving application of WT has been presented. WT has been applied to ill-behaved price series to convert it into its constitutive series and their statistical properties are more like a normal distribution curve than the original series and can be utilized for better prediction. Therefore, input variable set to the NN consists of variables from original time domain as well as from wavelet domain. Input variables other than price and load have also been considered in formulating the model. The proposed model (PM5) gives better accuracy than the benchmark NN model (PM4). It has been shown empirically that the third level of decomposition of price series is optimum for accurate forecasting. Since wavelet domain variables have been used as additional variables along with time domain variables, therefore there is no possibility of loss of information. The proposed model (PM5) has been also been compared with a heuristic method (PM1), a model followed by IESO (PM2), MLR model (PM3) for a period of two years. Results of ARIMA, TF and DR models have also been included for comparison. By comparing the forecasting performance of all the models, it can be concluded that the proposed wavelet-NN based model (PM5) provides better forecast with reasonable degree of accuracy and since its forecasts are available before submission of initial bids, it gives sufficient time to the participants, especially bulk electricity customers, to respond.

REFERENCES

[1] M. Shahidepour, H. Yamin, and Z. Li, *Market Operations in Electric Power Systems Forecasting, Scheduling and Risk Management*, Willey-Interscience, 2002.

[2] A. J. Conejo, J. Contreras, R. Espinola, and M. A. Plazas, "Forecasting electricity prices for a day-ahead pool-based electric energy market," *International Journal of Forecasting*, vol. 21, no. 3, pp. 435-462, 2005.

[3] J. Bastian, J. Zhu, V. Banunaryanan, and R. Mukherji, "Forecasting energy prices in a competitive market," *IEEE Computer Applications in Power Magazine*, vol. 12, no. 3, pp. 40-45, 1999.

[4] H. Y. Yamin, S. M. Shahidehpour, and Z. Li, "Adaptive short-term electricity price forecasting using artificial neural networks in the restructured power markets," *Electrical Power and Energy Systems*, vol. 26, pp. 571-581, 2004.

[5] R. Gareta, L. M. Romeo, and A. Gil, "Forecasting of electricity prices with neural networks," *Energy Conversion and Management*, vol. 47, pp. 1770-1778, 2006.

[6] C. P. Rodriguez and G. J. Anders, "Energy price forecasting in the Ontario competitive power system market," *IEEE Trans. on Power Systems*, vol. 19, no. 3, pp. 366-374, 2004.

[7] Y. H. Song, X. Wang, and J. Z. Liu, *Operation of Market oriented Power Systems*, Springer, 2003.

[8] J. C. Cuaresma, J. Hlouskova, S. Kossmeier, and M. Obersteiner, "Forecasting electricity spotprices using linear univariate time-series models," *Applied Energy*, vol. 77, pp. 87-106, 2004.

[9] H. Zareipour, C. A. Canizares, and K. Bhattacharya, "Application of public-domain market information to forecast Ontario's wholesale electricity prices," *IEEE Trans. on Power Systems*, vol. 21, no. 4, pp. 1707-1717, 2006.

[10] R. C. Garcia, J. Contreras, M. Akkeren, and J. B. C. Garcia, "A GARCH forecasting model to predict day-ahead electricity prices," *IEEE Trans. on Power Systems*, vol. 20, no. 2, pp. 867-874, 2005.

[11] D. W. Bunn, "Forecasting loads and prices in competitive power markets," *Proc. of the IEEE*, vol. 88, no. 2, pp. 163-169, 2000.

[12] Website of Ontario electricity market, http://www.ieso.ca

[13] J. D. Nicolaisen, C. W. Richter, Jr., and G. B. Sheble, "Price signal analysis for competitive electric generation companies," *Proc. of Conf. Electric Utility Deregulation and Restructuring and Power Technologies*, London, U.K., pp. 4-7, April 2000.

[14] C. Kim, I.-K. Yu, and Y. H. Song, "Prediction of system marginal price of electricity using wavelet transform analysis," *Energy Conversion and Management*, vol. 43, pp. 1839-1851, 2002.

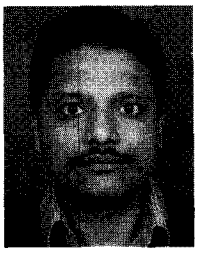
[15] A. J. Conejo, M. A. Plazas, R. Espinola, and A. B. Molina, "Day-ahead electricity price forecasting using the wavelet transform and ARIMA models," *IEEE Trans. on Power Systems*, vol. 20, no. 2, pp. 1035-1042, 2005.

[16] H. Xu and T. Niimura, "Short-term electricity price modeling and forecasting using wavelets and multivariate time series," *Proc. of Power Systems Conference and Exposition, IEEE, PES*, vol. 1, pp. 208-212, 10-13 October 2004.

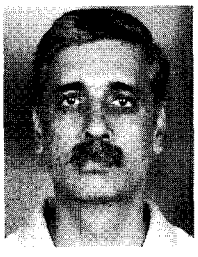
[17] S. J. Yao, Y. H. Song, L. Z. Zhang, and X. Y. Cheng, "Prediction of system marginal price by wavelet transform and neural network," *Electric Machines and Power Systems*, vol. 28, pp. 537-549, 2000.

[18] D. Benaouda and F. Murtagh, "Hybrid wavelet

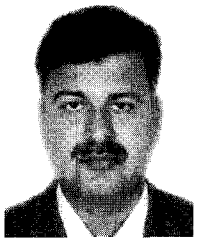
model for electricity pool-price forecasting in a deregulated electricity market," *Proc. of IEEE International Conference on Engineering of Intelligent Systems*, pp. 1-6, 22-23 April 2006.
[19] R. M. Rao and A. S. Bopardikar, *Wavelet Transforms - Introduction to Theory and Applications*, 3rd edition, Pearson Education Asia, 2002.
[20] R. Polikar, *The Engineer's Ultimate Guide To Wavelet Analysis - The Wavelet Tutorial*, Rowan University, Available at: polikar@rowan.edu
[21] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, 1998.
[22] MATLAB Wavelet Toolbox, The Mathworks, Available at: http://www.mathworks.com/
[23] U.S. Department of Energy, Energy Information Administration website http://www.eia.doe.gov/
[24] Weather underground website http://www.wunderground.com
[25] MATLAB Statistical Toolbox, The Mathworks, Available at: http://www.mathworks.com/
[26] A. J. Rocha Reis and A. P. Alves da Silva, "Feature extraction via multi-resolution analysis for short term load forecasting," *IEEE Trans. on Power Systems*, vol. 20, no. 1, pp. 189-198, 2005.
[27] H. Zareipour, C. Canizares, and K. Bhattacharya, "An overview of the operation of Ontario's electricity market," *Proc. IEEE Power Engineering Society Annual General Meeting*, pp. 2463-2470, June 2005.
[28] A. D. Aczel, *Complete Business Statistics*, McGraw Hill, 1999.
[29] MATLAB Neural Network Toolbox, The Mathworks, Available at: http://www.mathworks.com/
[30] H. S. Hippert, C. E. Pedreira, and R. C. Souza, "Neural networks for short-term load forecasting: A review and evaluation," *IEEE Trans. on Power Systems*, vol. 16, no. 1, pp. 44-55, February 2001.
[31] G. Gross and F. D. Galiana, "Short-term load forecasting," *Proc. of the IEEE*, vol. 75, no. 12, pp. 1558-1573, December 1987.



Sanjeev Kumar Aggarwal received the B.E. from M.D. University and M.Tech from NIT, Kurukshetra. He worked for five years for NTPC Ltd. Presently, he is pursuing a Ph.D. in the area of price and load forecasting in deregulated electricity markets from NIT, Kurukshetra. His research interests are load and price forecasting, AC-DC load flow analysis, neural networks and signal processing.



Lalit Mohan Saini received the B.E. from Panjab University and M.Tech from REC, Kurukshetra and Ph.D. from Kurukshetra University. He is an Assistant Professor in EED, NIT, Kurukshetra. His main areas of interest are load and price forecasting, neural networks, power system automation and control.



Ashwani Kumar received the B.Tech. from Pant Nagar University and M.Tech. from Panjab University, Chandigarh. He received the Ph.D. from IIT, Kanpur. He is an Assistant Professor in EED, NIT, Kurukshetra. His main areas of interest are power system optimization, restructuring of electric supply industry, applications of FACTs devices.