FULL PAPER
International Journal of Recent Trends in Engineering, Vol 2, No. 5, November 2009

Electricity price forecasting using Wavelet domain and Time domain features in a Regression based technique

S. K. Aggarwal[1], L. M. saini[2], and A. Kumar[3]
[1, 2, 3] National Institute of Technology /Electrical Engineering, Kurukshetra, India
Email: vasusanjeev@yahoo.co.in
Email: lmsaini @rediffmail.com, ashwa_ks@yahoo.co.in

***Abstract*— A combined wavelet transform (WT) and multiple linear regression (MLR) based technique to forecast price profile in a single settlement real time electricity market has been presented. The historical price and load data has been decomposed into better-behaved wavelet domain constitutive subseries using WT and then combined with other time domain variables to form the set of input variables for the MLR based forecasting model, which models each day using 24 separate equations for 24 hours of the day. Forecasting performance of the proposed model for a period of three years has been compared with the three other models.**

***Index Terms*— Multiple linear regression, Normal distribution, Price forecasting, Wavelet transform.**

I. INTRODUCTION

Forecasting electricity prices is a complex task due to non-storable nature of electrical energy and stiff condition of maintaining demand and supply in real time [1-2]. There are two main methodologies followed for price forecasting: (i) Production cost models [3] and, (ii) Statistical models [4-17]. Production cost models require detailed system operation data and are complicated to implement. Statistical methods require lesser amount of data than the production cost models, and may be adopted by generation companies in preparing their bidding strategies and by large consumers for demand side management. There are two main approaches of the statistical modeling: (i) Non-linear models, and (ii) Linear models. Non-linear models like artificial intelligence (AI) based methods employing neural networks (NNs) have been proposed by different researchers [4-8]. Second type of models are linear models like univariate autoregressive (AR) [9], autoregressive moving average (ARMA) [10], autoregressive integrated moving average (ARIMA) [11], multivariate time series models like transfer function and dynamic regression [11] and Generalized Autoregressive Conditional Heteroskedastic (GARCH) model [12]. Models have been reported, where price series was preprocessed through some signal processing technique like wavelet transfrom (WT) [13-16] and then prediction has been made through any of previously established methods like regression model [13], time series model [14], or NN model [15-16].

A combined WT and MLR model has been proposed in this paper to forecast the price profile in Ontario electricity market (OEM) with forecasting horizon of 24 hours. WT is used to split up the time series into one low-frequency subseries (approximation part) and some high-frequency subseries (detailed part) in the wavelet domain. In models [13-16], after appropriate decomposition, the prediction was made in wavelet domain and then inverse WT was applied to obtain the actual value of the predicted variable. While, in this work, inputs to the forecasting model are based on information from both the original time domain price series and from wavelet domain subseries. Other input variables like oil price have also been included in the forecasting model to improve the forecasting accuracy. In this model, reconstruction is not needed and predicted outputs from the different subseries have been merely added to get the final output. The main contribution of this paper is to propose a model, which combines time domain features and wavelet domain features of price sub-series in a single framework. The MLR has been chosen as the forecasting technique because of its simplicity; ability to incorporate as many as possible explicative variables and it takes less computation time. Forecasting performance of the proposed model for a period of three years has been compared with (i) a heuristic technique, (ii) a simulation model used by Ontario's IESO and, (iii) a MLR model. Forecasting results show that proposed model is more accurate than the other models.

II. ONTARIO ELECTRICITY MARKET

The wholesale price for electricity in Ontario is based on hourly bids and offers into the market from the participants [17]. The market clearing price (MCP) is set for each five-minute interval. In addition to this, each hour, hourly Ontario energy price (HOEP) is determined by taking the average of twelve MCPs during an hour. HOEP usually varies from 20 to 80 $/MWh under normal circumstances and is used as the wholesale price for electricity in OEM and has been predicted in this paper. The 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 for pre-dispatch instructions purpose, 12:00 AM - IESO publishes first pre-dispatch report for day D for each of 24 hours of dispatch day D. (ii) *Dispatch day (D)*: A registered market participant may revise bid data two hours prior to the beginning of each dispatch interval. In addition to this, the IESO starts posting the system status report (SSR) daily on its website [17] at least 24 hours in advance for day D and keeps on updating it. The SSR contains hourly forecast information of twelve variables on supply side; ten on demand side and nine other variables.

III. Methodology and Model for Price Forecasting

The modeling task has been divided into four parts: data set selection, input variable selection, data preprocessing, and forecasting models.

A. Data Set

Following hourly variables have been considered in this work: (i) HOEP data, (ii) 1 hour, 2 hour and 3 hour ahead pre-dispatch prices (PDP), (iii) total market demand (TMD), (iv) Ontario demand (OD), (v) imports, (vi) exports, (vii) ten variables from SSR report for D-day (from SSR available before 11 A.M. on (D-1) day), (viii) crude oil prices, (ix) nine weather variables. Data for the period March 1, 2003 to June 30, 2006 has been taken for the study. Variables (i) through (vii) have been taken from [17]. Crude oil price data and weather data has been collected from [18] and [19] respectively.

B. Input Variable Selection

In this paper, final explanatory variables have been selected based upon linear correlation analysis, since MLR is a linear prediction technique. It has been observed that past HOEP, PDP forecasted by IESO, TMD, and OD exhibit good correlation with the HOEP. Since PDPs are highly collinear with HOEP; therefore, their net effect on final output is insignificant. Similarly OD and TMD are collinear, with OD presenting a better option than TMD; since, TMD is sum of OD, imports and exports. Imports and exports show no significant correlation with the price, so OD has only been selected in the final model. Crude oil price has also shown some linear correlation with the price.

Weather variables do not exhibit significant correlation with the price. Out of ten variables from SSR, only four variables have shown significant correlation with the price namely primary demand east, primary demand west, energy excess and capacity excess. Primary demand of east plus primary demand of west is the demand forecast made by IESO for D-day in SSR report. SSR planned outages east and west have not shown any meaningful correlation with the price. Final set of selected explanatory variables has been shown in Table I.

TABLE I. Final Set of Explanatory Variables

Variable	Time lag
HOEP	D-2
Crude oil price	D-2
SSR Forecast demand	D
SSR Energy excess	D
SSR Capacity excess	D

C. Data Preprocessing

The WT converts a price series into a set (typically six) of constitutive series [20-22]. Many wavelet families exist, where Daubechies family of wavelets, which are compactly supported orthonormal wavelets, is the most popular one and has been used in this work. The division of a weekly price series curve in approximation A_3, A_2, A_1 and the corresponding levels of detail D_3, D_2, and D1 has been shown in fig. 1. It can be noticed that the approximations A_3, A_2, and A_1 exhibit essential daily trend in the price signal with A_3 a very good representation of the original price series, which is highly volatile and corrupted by the occasional spikes. The details D_3 and D_2 contain useful higher frequency information and D_1 contains the information regarding the time localization of the spike.

The skewness and kurtosis characteristics of these series have been presented in Table II. Skewness is a measure of the symmetry of the data around the data mean. The skewness of the ideal normal curve is zero. Kurtosis is a measure of how outlier prone a distribution is. The kurtosis of normal distribution curve is three [23-24]. Especially A_3, A_2, and D_3, D_2 presents very close approximation to the normal curve. Curve A_1 and D_1 are skewed and outlier prone and are not very much effective in a linear predictive model like MLR.

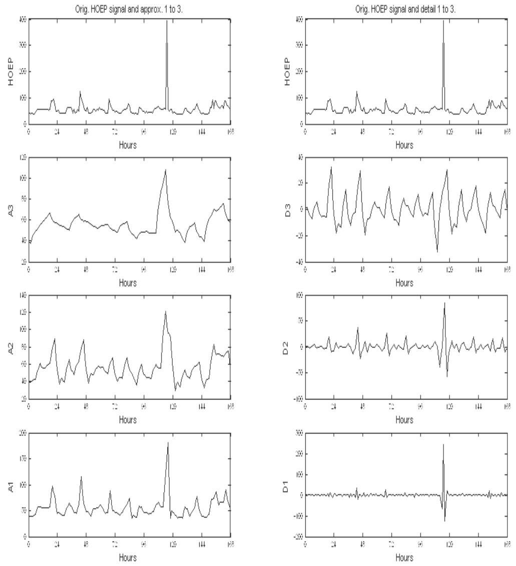

Figure 1. Price and its constitutive series using MRA (March 1, 2005 to March 7, 2005)

TABLE II. Statistical Properties of Constitutive Series of Price

Time series	Skewness	Kurtosis
HOEP series	3.19	33.2
Approx. A_3	1.17	5.2
Approx. A_2	1.08	4.23
Approx. A_1	1.27	5.06
Detail D_3	0.54	9.1
Detail D_2	0.96	18.45
Detail D_1	6.71	185.9

IV. Forecasting Models

Four different models have been analyzed and compared in this section.

A. 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:

$$P_{d,t} = P_{d-comp,t} \times \frac{L_{d,t}}{L_{d-comp,t}} \quad (1)$$

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

$P_{d-comp,t}$ is the price at hour t of the comparable day of forecast day d.

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

$L_{d-comp,t}$ is 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 seasonality. SSR forecasted demand (SSRFD) for D-day has been taken as the forecast load in (1) to predict the HOEP.

B. IESO model (PM2)

This is the model used by IESO and is simulation-based forecast information for market participants. PDPs are available 3 hour, 2 hour and 1 hour before the actual dispatch. One hour before PDP information from the IESO website [17] has been taken to make a comparative study with the other models.

C. MLR model (PM3)

This model has been implemented using Matlab7.0 [23-24]. Variable segmentation has been applied to the complete data set by dividing it into 24 time series, each one corresponding to an hour of the day. Separate regression coefficients have been calculated for each of the 24 time series. These coefficients for the predicted day have been calculated using the data of the past fifteen weeks. Length of data period from seventeen weeks to seven weeks was tried and optimum results were obtained using data of fifteen weeks. Initially, all input variables given in Table I were considered; but, it was found that by inclusion of SSRFD in the input variables, accuracy of the model gets deteriorated. Thus, variables other than SSRFD in Table I provide the best possible combination.

D. Wavelet-MLR model (PM4)

Basic methodology of this model is similar to PM3 and has bee implemented using Matlab7.0 [22-23]. All the variables from Table I have been used to predict the regression coefficients. The available data to forecast price on day D (Y_D) can be denoted by $X_{m,d}$; m are the number of variables and d = D-105, D-104, . . . , D-1, D. The detailed steps are as follows:

Step 1) Decompose the available historical HOEP series and SSRFD series in a set of four constitutive series (A_3, D_3, D_2, and D_1) using WT. The Daubechies wavelet of order 2 (db2) has been used for decomposition. Since, information contained within D_1 is outlier prone; therefore, series corresponding to D_1 has been discarded and only series A_3, D_3 and D_2 have been used. With this process, there is no loss of information; because, time domain HOEP series is still used as input variable. Wavelets other than db2 were also tried; however, best results have been obtained using db2 only. The original data set is converted into the three data sets.

$$\{X_{m,d}, Y_d;\ m = 1,2,\ldots,8; d = D-105, D-104, \ldots D-1, D\} \Rightarrow$$
$$[\{XA3_{m1,d}, XD3_{m1,d}, XD2_{m1,d}, YA3_d, YD3_d, YD2_d\};\ m1 = 1,2,\ldots 9;\ d = D-105, D-104, \ldots D].$$

when, d = D, then $X_{m,d}$ contains the input variables used for HOEP prediction for D day. Y_D is the output variable vector. *XA3*, *XD3*, *XD2* each consist of six variables and a constant as given in the Table III. *YA3*, *YD3*, and *YD2* are the corresponding output vectors.

Step 2) Convert each of constitutive series $XA3_{m1,d}$, $XD3_{m1,d}$, and $XD2_{m1,d}$ and corresponding output vectors into 24 number of separate hourly series.

Step 3) Each day has been divided into four segments. For prediction of price in segments 1 and 3, all three constitutive series, *XA3*, *XD3*, and *XD2* are required; whereas, for segments 2 and 4, only *XA3* is needed. This has been done by considering the behavior of past load and price curves. Price curve in segments 2 and 4 shows high volatility as compared to segments 1 and 3.

Step 4) Apply MLR to calculate regression coefficients for D-day using variables as shown in Table III and constitutive series in column 3 of Table IV.

Step 5) Calculate individual price components *YA3*, *YD3*, *YD2* for each hour using separate regression coefficients calculated for each sub-series *XA3*, *XD3*, and *XD2*.

Step 6) Calculate HOEP for each hour using *YA3*, *YD3*, *YD2* as shown in column 4 of Table IV.

TABLE III. INPUT VARIABLE SET FOR PM4

XA3 variables	*XD3* variables	*XD2* variables	Time lag
HOEP	HOEP	HOEP	D-2
A_3 of HOEP	D_3 of HOEP	D_2 of HOEP	D-2
A_3 of SSRFD	D_3 of SSRFD	D_2 of SSRFD	D
Crude oil price	Crude oil price	Crude oil price	D-2
SSR energy excess	SSR energy excess	SSR energy excess	D
SSR capacity excess	SSR capacity excess	SSR capacity excess	D
Constant	Constant	Constant	-

TABLE IV. VARIABLE SEGMENTATION SET FOR PRICE CURVE

Seg-ment	Hour No.	Constitutive series used to calculate regression coefficients	Predicted price component
1	2 to 7	$XA3_{m1,dr}$, $XD3_{m1,dr}$, $XD2_{m1,dr}$	$Y_D = YA3_D + YD3_D + YD2_D$
2	8 to 14	$XA3_{m1,dr}$	$Y_D = YA3_D$
3	15 to 18	$XA3_{m1,dr}$, $XD3_{m1,dr}$, $XD2_{m1,dr}$	$Y_D = YA3_D + YD3_D + YD2_D$
4	19to24, 1	$XA3_{m1,dr}$	$Y_D = YA3_D$
Here **dr** = D-105, D-104, …, D-1.			

V. RESULTS AND DISCUSSION

The forecasting test period is from July 1, 2003 to June 30, 2006.

A. Comparison of forecasting performance

Two types of accuracy criteria have been adopted to assess and compare the performance of the models: (1) Mean absolute percentage error (MAPE), (2) Root mean square error (RMSE). MAPE can be defined as:

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{X_t - X_f}{X_t}\right| \times 100 \quad (2)$$

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.

RMSE can be defined as:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(X_t - X_f\right)^2} \quad (3)$$

Overall yearly MAPE and RMSE comparison has been presented in Table V and monthly MAPE comparison has been given in Table VI. It can be observed that overall MAPE performance of PM4 is better than PM1, PM2 and PM3 by 22.89%, 11.96% and 6.65% respectively. During thirty-four out of thirty-six months, PM4 performs better than PM1. PM4 performs better than PM2 during twenty-six out of thirty-six months. PM4 outperforms PM3 during twenty-five months and performance of both the models is comparable during rest of the months. Yearly RMSE performance of PM4 is better than PM1 and PM2 in all the three years. The RMSE performance of PM4 and PM3 are comparable.

B. *Analysis of performance*

The average annual volatility [11] during calendar years 2003, 2004, and 2005 is 0.49, 0.36, and 0.4 respectively. The performance of PM1 deteriorates sharply during the months of high volatility. During the periods of low volatility, its performance is co mparable to other models. Except a few months, like July 2004, the performance of PM2 also suffers during high volatility period. However, PM4 has performed consistently well during all periods. Its maximum monthly MAPE is 41.53%, during April 2006, which is a month of high volatility having value of 0.57. During this month PM1, PM2 and PM3 have shown MAPE of 55.01%, 45.35% and 49.53% respectively. On the other hand, during October 2004, which is a low volatility month of value 0.2 only, PM4 has shown a MAPE of 12.89%; whereas, MAPEs of PM1, PM2 and PM3 are 15.25%, 15.76% and 13.26% respectively. The performance of PM3 and PM4 is moderate during January 2006, which is a month of moderate volatility of 0.29. Thus, during periods of high volatility PM3 and PM4 perform better than PM1 and PM2, while during periods of low volatility PM1 and PM2 may perform better than PM4 with a very little probability. It can be concluded that PM4 performs consistently under all conditions of varying volatility.

Histograms of DMAPE have been shown in fig. 2. It can be observed that maximum DMAPE in case of PM1, PM2, PM3 and PM4 is 343%, 149%, 173% and 120% respectively. The number of days when DMAPE is above 70% is 62, 17, 14 and 6 for PM1, PM2, PM3 and PM4 respectively. The number of days when DMAPE is below 20% is 335, 359, 408 and 415 for PM1, PM2, PM3 and PM4 respectively. On both criteria, PM4 has shown the best performance.

VI. CONCLUSION

In this work, a new approach for price forecasting with variable segmented MLR has been proposed in which input variables from both original time domain series and wavelet domain are present. Input variables other than price and load have also been considered in formulating the WT based MLR model (PM4). The proposed model has been compared with a heuristic method (PM1), a model followed by IESO (PM2), and a variable segmented MLR model (PM3) over a period of three years. A great deal of improvement in forecasting performance has been observed during on-peak hours as compared to PM2 and PM3. PM1 and PM2 have been found to be giving reasonable performance during periods of extremely low volatility, but low volatility period is very small in duration. By analyzing the forecasting performance of all the four models, it can be concluded that the proposed wavelet-MLR based method provides forecast with reasonable degree of accuracy and will be especially helpful during on-peak hours and periods of high volatility. PM4 can easily find real-world price forecasting application as it predicts price before submission of initial bids (much before than the IESO model) and utilizes publicly available information only.

TABLE V. YEARLY MAPE AND RMSE COMPARISON

	MAPE				RMSE			
Year	PM1	PM2	PM3	PM4	PM1	PM2	PM3	PM4
2003-04	32.68	31.27	27.22	26.80	21.11	18.40	16.00	16.66
2004-05	26.85	21.42	20.81	19.15	19.36	14.68	14.26	14.34
2005-06	36.04	31.04	30.94	27.76	31.82	26.00	24.64	24.21
Average	31.86	27.91	26.32	24.57	24.10	19.69	18.30	18.40

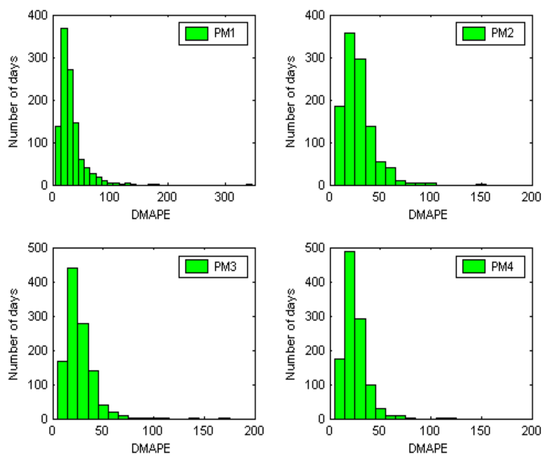

Figure 2. Histograms of DMAPE

ACKNOWLEDGMENT

The authors are thankful to Robert Doyle and Charanjeet Minhas from Customer Communications Department of Ontario Independent Electricity System Operator (IESO) for providing the past SSR data.

TABLE VI. COMPARATIVE MONTHLY MAPES (JULY, 2003 TO JUNE, 2006)

	2003-04				2004-05				2005-06			
Month	PM1	PM2	PM3	PM4	PM1	PM2	PM3	PM4	PM1	PM2	PM3	PM4
July	35.57	28.88	27.16	30.4	32.97	22.91	25.13	26.07	34.7	33.22	25.79	23.93
August	39.05	27.22	27.85	25.62	28.43	21.04	22.4	20.14	43.12	35.37	29.59	27.3
September	25.45	23.37	26.01	25.31	19.83	14.31	18.53	14.4	41.68	29.44	28.21	25.52
October	31.2	22.22	20.49	23.02	15.25	15.76	13.26	12.89	44.02	38.99	43.25	32.43
November	30.05	21.89	32.92	28.96	26.59	27.7	20.71	21.39	35.57	34.46	35.3	32.76
December	38.85	46.7	28.49	30.09	26.99	23.12	18.43	19.09	53.02	35.34	30.84	29.16
January	31.34	46.26	29.14	30.7	26.52	23.5	18.22	18	23.05	19.54	27.68	25.82
February	35.96	37.54	28.47	26.92	19.64	17.22	13.35	13.82	22.06	20.64	20.63	21.74
March	28.06	27.79	27.82	22.76	23.92	21.52	16.28	16.23	21.67	27.54	28.35	22.13
April	26.64	33.25	23.1	22.93	32.6	24.88	30.98	23.01	55.01	45.35	49.53	41.53
May	37.02	30.97	28.73	28.69	27.41	19.81	23.94	20.36	27.18	32.39	26.19	23.74
June	32.72	28.98	26.58	26.26	41.72	25.12	28.3	24.39	30.59	19.32	25.38	27.04
Average	32.68	31.27	27.22	26.8	26.85	21.42	20.81	19.15	36.04	31.04	30.94	27.76

REFERENCES

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

[2] N. Amjady, M. Hemmati, "Energy price forecasting - Problems and proposals for such predictions", *IEEE Power and Energy Magazine*, March-April 2006, pp. 20-29.

[3] J. Bastian, J. Zhu, V. Banunaryanan, R. Mukherji, "Forecasting energy prices in a competitive market," *IEEE Computer Applications in Power Magazine*, July 1999, pp. 40-45.

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

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

[6] N. Amjady, "Day ahead price forecasting of electricity markets by a new fuzzy neural network", *IEEE Trans. on Power Systems*, vol. 21, No. 2, May 2006, pp. 887-896.

[7] Li Zhang, P. B. Luh, "Neural network-based market clearing price prediction and confidence interval estimation with an improved extended Kalman Filter method", *IEEE Trans. on Power Systems*, vol. 20, no. 1, February 2005, pp. 59- 66.

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

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

[10] M. Zhou, Z. Yan, Y.X. Ni, G. Li, Y. Nie, "Electricity price forecasting with confidence-interval estimation through an extended ARIMA approach", *IEE Proc. of Generation, Transmission and Distribution*, vol. 153, no. 2, March 2006, pp. 187-195.

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

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

[13] Chang-il Kim, In-Keun Yu, Y.H. Song, "Prediction of system marginal price of electricity using wavelet transform analysis”, *Energy Conversion and Management*, vol. 43, 2002, pp. 1839-1851.

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

[15] 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, 2000, pp. 537-549.

[16] D. Benaouda, F. Murtagh, "Hybrid Wavelet Model for Electricity Pool-price Forecasting in a Deregulated Electricity Market", *IEEE International Conference on Engineering of Intelligent Systems*, 22-23 April, 2006, pp. 1-6

[17] Website of IESO of Ontario electricity market. (http://www.ieso.ca)

[18] U.S. Department of Energy, Energy Information Administration website (http://www.eia.doe.gov/)

[19] Weather Underground website (http://www.wunderground.com)

[20] R.M. Rao, A.S. Bopardikar, *Wavelet Transforms – Introduction to theory and applications*, Third Edition, Pearson Education Asia, 2002.

[21] Robi Polikar, *The Engineer’s Ultimate Guide To Wavelet Analysis - The Wavelet Tutorial*, Rowan University, Available: polikar@rowan.edu

[22] MATLAB Wavelet Toolbox, The Mathworks, Available: http://www.mathworks.com/

[23] MATLAB Statistical Toolbox, The Mathworks, Available: http://www.mathworks.com/.

[24] Amir D. Aczel, *Complete Business Statistics*, McGraw Hill, 1999.