An Optimization Method of Genetic Algorithm for LSSVM in Medium Term Electricity Price Forecasting

Intan Azmira Wan Abdul Razak[1], Izham Zainal Abidin[2], Yap Keem Siah[2], Aidil Azwin Zainul Abidin[2], Titik Khawa Abdul Rahman[3], Nurliyana Baharin[1], Mohd. Hafiz Bin Jali[1]

[1]Faculty of Electrical Engineering, Universiti Teknikal Malaysia Melaka, Malacca, Malaysia
[2]College of Engineering, National Energy University, Selangor, Malaysia,
[3]Faculty of Engineering Girl Campus, King Abdulaziz University, Jeddah, Saudi Arabia.
intan.azmira@utem.edu.my

***Abstract*—Predicting electricity price has now become an important task for planning and maintenance of power system. In medium term forecast, electricity price can be predicted for several weeks ahead up to a year or few months ahead. It is useful for resources reallocation where the market players have to manage the price risk on the expected market scenario. However, researches on medium term price forecast have also exhibited low forecast accuracy. This is due to the limited historical data for training and testing purposes. Therefore, an optimisation technique of Genetic Algorithm (GA) for Least Square Support Vector Machine (LSSVM) was developed in this study to provide an accurate electricity price forecast with optimised LSSVM parameters and input features. So far, no literature has been found on feature and parameter selections using the method of LSSVM-GA for medium term price prediction. The model was examined on the Ontario power market; which is reported as among the most volatile market worldwide. The monthly average of Hourly Ontario Electricity Price (HOEP) for the past 12 months and month index are selected as the input features. The developed LSSVM-GA shows higher forecast accuracy with lower complexity than the existing models.**

***Index Terms*—Genetic Algorithms; Least Square Support Vector Machines; Medium Term Price Forecasting; Optimization.**

I. INTRODUCTION

Price prediction is important to market members in deregulated electricity environment to provide a better maintenance scheduling, developing investment, medium term planning, as well as decision-making.

However, forecasting electricity price is more challenging compared to predicting the load or demand due to the volatility of price series with unexpected price spikes at any point of series. In addition, unlike short-term price forecast, medium term forecast is more challenging [1]–[4]. One of the reasons is because the accessible historical data for medium term price forecast is limited. Short-term forecast usually needs only a few days of historical data to train the forecast model, but medium term forecast usually takes one year of historical data [2], [5]. Hence, medium term forecast cannot extract price trend from the immediate past.

II. LITERATURE SURVEY

Only a few researches have been conducted in the field of medium term price forecasting. Some studies on Time Series (TS) and Support Vector Machine (SVM) were reported in this area. However, some researchers found that Neural Network (NN) method is not suitable for medium term forecast as NN needs large data set for network training [6].

Author of [7] proposed two approaches for medium term forecast; TS models and generalised least squares model with auto-correlated residuals. The models were examined on Spanish market in March to September 2005. Meanwhile, Autoregressive Moving Average Exogenous (ARMAX) model was designed by [8] to forecast monthly price in Pennsylvania–New Jersey–Maryland (PJM) market during June 2009 and June 2010.

Authors of [9] proposed a hybrid model of SVM and ARMAX and compared its performances with the single SVM. The forecast models were tested on PJM market in June 2009 and June 2010. On the other hand, the findings of [1] indicated that the proposed hybrid method of SVM and ARMAX is more accurate than stand-alone ARMAX when tested on PJM market in June 2009 and June 2010.

The integration of ARMAX and LSSVM approach [8] was observed and compared with the single LSSVM by the same authors for the same test data and market. Authors in [3] compared SVM and LSSVM performances on PJM market. Results show that LSSVM outperformed SVM in June 2009. The authors also design multiple SVM [2], [10] by classifying the prices into one, two, three, four, and five price zones. The significant inputs were selected based on cross-validation technique. The results revealed that four price zones representing low, medium, high, and peak modules outperformed other types of price zones when tested on PJM market in June 2009. Meanwhile, multiple SVM models outperformed single SVM model in June 2009 and June 2010.

The authors have further investigated the medium term electricity price forecasting by applying two-stage multiple SVM [11] on the same power market and test data. The first stage was performed by a single SVM to produce initial forecast values. The outputs from the first stage were fed into the second stage according to four different price zones: low, medium, high, and peak zones. However, the proposed model does not improve the previous works when it produces lower accuracy than the single LSSVM [3], LSSVM-ARMAX [8], and multiple LSSVM for test data in June 2009. Meanwhile, the developed model shows lower performance than LSSVM-ARMAX [8] and multiple LSSVM during June 2010.

Other researchers design regression models to predict the 12-months-ahead price for the year 2009 in Nord Pool market [12]. The authors further investigated some approaches in medium term forecast and examined the forecast models on Ontario and Nord Pool electricity market [6]. The developed models tested on Ontario market show that SVM model outperformed other forecast models of RBF-NN, wavelet NN (WNN), and Navigant Consulting Company. Meanwhile, hybrid models were designed to improve forecasting accuracy of the single models. Two forecast engines were combined for each hybrid model to provide a pre-forecast and final forecast. The results revealed the superiority of SVM when the model of SVM/SVM surpassed other hybrid models of SVM/RBF-NN, RBF-NN/RBF-NN, and RBF-NN/SVM.

More work should be carried out to produce better forecast accuracy by properly selecting the significant features and network parameters. To the best of the authors' review, no literature has been found on the application of LSSVM and Genetic Algorithm in medium term electricity price forecast. In addition, the approach of feature selection and parameter optimisation using a single optimisation technique has not reported yet. Thus, this study developed a forecasting technique to improve medium term electricity price forecasting using a hybrid model of LSSVM and GA. With a single optimisation method of GA, the input features and LSSVM parameters are simultaneously optimised. This method is proven to give better forecast accuracy as compared to other existing models, which can contribute for decision-making and medium-term planning in electricity power market.

III. THEORY

This section discusses the theory of the main forecast engine (LSSVM) and the optimisation algorithm (GA) implemented in this study.

A. Fundamental of SVM and LSSVM

SVM as presented by [13], is a supervised learning model that supports data analysis and pattern recognition for classification and estimation. Support Vector Regression solves for quadratic programs which involve inequality constraint. However, SVM has a high computational problem. SVM can reduce over-fitting, local minima problems [14], and able to deal with high dimensional input spaces splendidly [15]. Nevertheless, the main disadvantage of SVM is its high computational complexity due to constrained optimisation programming. Hence, Least Squares Support Vector Machine (LSSVM) was proposed to diminish the computational burden of SVM, which applies with equality instead of inequality constraints [16]. LSSVM solves a system of linear equations instead of quadratic programming (QP) problem that improves the computational speed [15], [17]. The linear system, namely as Karush- Kuhn-Tucker (KKT), is more straightforward than QP system. LSSVM also maintains the principle of SVM, which possess good generalisation capability. LSSVM reduces the sum square errors (SSEs) of training data sets while concurrently diminishing margin error. Meanwhile, in contrast to SVM, LSSVM uses the least squares loss function instead of the ε-insensitive loss function.

B. Fundamental of Genetic Algorithm (GA)

GA that was first introduced by [18] is based on the 'survival of the healthiest' and natural evolution mechanism via reproduction. It can find the optimal solution after some iterative computations. The solution is represented by a string, named 'chromosome', comprising of a set of components, named 'genes', which consist of a set of values for the optimisation variables. The objective functions are often referred to as fitness functions. Three main operations in GA are selection, crossover, and mutation.

The optimisation process is started with a random initial population of chromosomes, followed by fitness evaluation. The next step is a selection of fittest individuals or parents for reproduction, where chromosomes with better fitness values have more potential to yield children during subsequent generation. In order to mimic the natural survival of the fittest progression, the best chromosomes exchange genes via crossover and mutation to create children chromosomes during the reproduction process. With the size of the population is preserved, the highly fit parent perform crossover with another parent in a population where parts of two genotypes are swapped. The crossover rate usually ranges from 0.6 to 1.0 [19].

After crossover, mutation is performed for any parent chromosome to maintain the variety of the solution candidates by bringing small and random changes into them. Mutations are accomplished randomly by changing a "1" bit into a "0" bit or a "0" bit into a "1" bit. In contrast to crossover, mutation is an unusual process, but by introducing new genetic material to the evolutionary progress, possibly thus avoiding chromosomes from being trapped in local minima. The mutation rate is usually 0.001 [20] or less than 0.1 [19].

The flowchart of GA operation is also illustrated in Figure 1 in Section V. Four core elements influencing the performance of GAs; population size, a number of generations, crossover rate, and mutation rate. Chances of obtaining global optimum can be increased by having a larger size of population (i.e. hundreds of chromosomes) and generations (thousands), but considerably increasing the computational time [19].

IV. THE ONTARIO POWER MARKET

In Ontario, electricity power market is conducted by Independent Electricity System Operator (IESO), which controls power system operation, forecasting short-term demand and supply of electricity, and managing the real-time spot market electricity price. The Ontario electricity market is a single settlement market, which applies real-time system while the day-ahead system is in progress. Due to the single settlement real-time power market, Ontario was reported as one of the most volatile markets in the world [21] and hence gives a big challenge for electricity price forecaster. Proper selection of features influences the efficiency and accuracy of forecasting. The important features for electricity price forecasting are analysed and being selected in the next section.

V. RESEARCH METHODOLOGY

This section provides the methodologies for medium term forecasting. In contrast with the short-term forecast, this medium term forecast has limited data, and only monthly average HOEP data are publicly available for the analysis. Section A presents the analysis for monthly average HOEP over the previous few years. Section B presents the proposed hybrid model of LSSVM-GA where LSSVM is the main forecast engine while GA is the optimisation algorithm that optimises the LSSVM parameters of gamma (γ) and sigma (σ) and selects the significant features to be fed into the LSSVM.

A. Analysis on monthly average HOEP

Monthly average HOEP from the year 2003 to 2010 were analysed to observe the price behaviour throughout the years. The monthly average HOEP are publicly available at http://www.ieso.ca/. Table 1 shows the monthly HOEP for each month, yearly HOEP, and the standard deviation of each month and year. It is clearly indicated that the monthly average HOEP fluctuates every year, which also proven by high standard deviation during each month.

Table 1
Monthly Average HOEP for 2003-2010

Month	Year								Standard deviation
	2003	2004	2005	2006	2007	2008	2009	2010	
January	59.62	66.22	57.9	55.54	44.48	40.74	53.22	37.4	9.39
February	86.46	52.74	49.58	48.12	59.12	52.38	47.24	35.9	13.75
March	81.49	48.9	59.87	49.01	54.85	56.84	28.88	28.22	16.10
April	58.88	45.92	61.93	43.52	46.05	48.98	18.4	30.83	13.25
May	43.17	48.06	53.05	46.32	38.5	34.56	27.77	38.77	7.54
June	41.64	46.69	65.99	46.08	44.38	57.44	22.84	40.36	11.83
July	40.08	45.58	76.05	50.52	43.9	56.58	18.99	50.83	15.01
August	48.97	43.51	88.24	52.72	53.62	46.57	26.07	44.41	16.35
September	48.56	49.57	93.7	35.42	44.63	49.09	20.76	32.91	20.07
October	57.09	49.11	75.92	40.2	48.91	45.27	29.22	29.39	14.24
November	40.45	52.28	58.25	49.71	46.95	51.78	26.54	31.89	10.21
December	44.42	50.82	79.77	39.25	49.08	46.34	35.05	33.83	13.58
Standard deviation	14.91	5.57	13.54	5.63	5.42	6.54	10.35	6.33	13.44 / 8.54

Therefore, it indicates that HOEP for the same month of every year is less suitable as input prediction. The average standard deviation of monthly HOEP over these eight years is 13.44, while the average standard deviation of yearly HOEP is 8.54. These standard deviations show that the HOEPs of the same month deviate more extensively from year to year, compared to the deviation of monthly HOEPs in a year. Hence, monthly HOEP of previous months is more suitable as input prediction, rather than applying the HOEP of the same months in previous years.

B. The Proposed Hybrid Model

A hybrid model of LSSVM-GA is developed with the training data from July 2004 to October 2009 and the testing period is from November 2009 to October 2010. Mean Absolute Percentage Error (MAPE) is selected as the objective function to measure the forecast accuracy. MAPE is formulated as in Equation (1):

$$\text{MAPE} = \frac{100}{N} \times \sum_{t=1}^{N} \frac{\left| P_{actual_t} - P_{forecast_t} \right|}{P_{actual_t}} \tag{1}$$

P_{actual} and P_{forecast} are the actual and forecasted HOEP at month t, respectively, while N is the number of the month. Meanwhile, Mean Absolute Error (MAE) is also calculated as in Equation (2):

$$\text{MAE} = \frac{1}{N} \times \sum_{t=1}^{N} \left| P_{actual\,t} - P_{forecast_t} \right| \tag{2}$$

Monthly average HOEP for the past 12 months and month index are selected as the input features. Month index is the index of the targeted month, which numbered from 1 to 12 to represent January to December. Hence, each training sample has 13 features, which were trained to produce one month-ahead. The flowchart of hybrid LSSVM-GA is illustrated as in Figure 1. GA optimises the 13 features and LSSVM parameters simultaneously. The optimisation process is initiated with a random population of chromosomes or solutions. The selected parameters and features are trained in LSSVM to produce a fitness value or MAPE. The following phase involves GA processes of selection, crossover, and mutation.

The optimisation process ends when a pre-defined number of generations have been achieved. Instead, the termination can also be executed when an acceptable solution has been found. Nevertheless, when no improvement is observed over a number of generations, the searching process should be stopped.

VI. Result and Discussion

Table 2 tabulates the network configuration and performance of LSSVM-GA. The optimised features, gamma and sigma, are case dependent, which is optimised by GA to produce the best MAPE.

From Table 2, it can be observed that GA selects eight features including month index and monthly HOEP of past tenth ($p_{(m-10)}$), ninth ($p_{(m-9)}$), eighth ($p_{(m-8)}$), seventh ($p_{(m-7)}$), sixth ($p_{(m-6)}$), fifth ($p_{(m-5)}$), and a month prior to the forecasted month ($p_{(m-1)}$). It can be observed that the selected features demonstrate short-term trend due to the selection of a month prior to the forecasted month ($p_{(m-1)}$). Meanwhile, the neighboring features (monthly HOEP of past tenth ($p_{(m-10)}$), ninth ($p_{(m-9)}$), eighth ($p_{(m-8)}$), seventh ($p_{(m-7)}$), sixth ($p_{(m-6)}$) and fifth ($p_{(m-5)}$) also exhibit short-term trend. Regression (R) is a correlation between target and output, which lies between 0 to 1. The target is highly correlated with the output when the regression value closes to 1 and hence leads to the more accurate forecast.

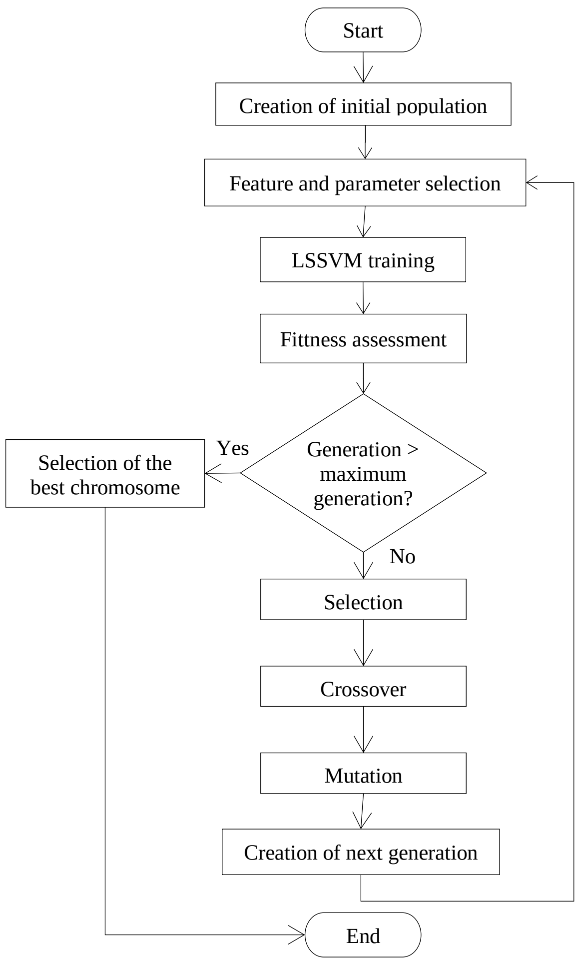

Figure 1: Flowchart of hybrid LSSVM-GA

Table 2
LSSVM-GA Performance for Medium Term Forecast

GA configuration	No. of population: 50 No. of generation: 80
Gamma	100
Sigma	15.65
Selected Features	8 features: $p_{(m-10)}$, $p_{(m-9)}$, $p_{(m-8)}$, $p_{(m-7)}$, $p_{(m-6)}$, $p_{(m-5)}$, $p_{(m-1)}$, month index
Regression	0.69
MAPE (%)	9.43
MAE	3.49

Meanwhile, the plot of actual HOEP against the forecasted HOEP is as shown in Figure 2. Between the period of May to October 2010 is summer period with the average HOEP of $39.45/MWh [22]. It was reported that this summer period has an increase in average HOEP by 62.5% from last summer period. In addition, the monthly average HOEP for any month during this summer period is above $30.00/MWh except for October 2010. The monthly average HOEP for each month during last summer is below $30.00/MWh. It can be noticed that generally, the predicted HOEP can track the actual price for most of the months except for the fifth and ninth month, which is March and July 2010; respectively. In fact, this spike price of $50.83/MWh is the first time the monthly average HOEP exceeded $50.00/MWh since January 2009 [22].

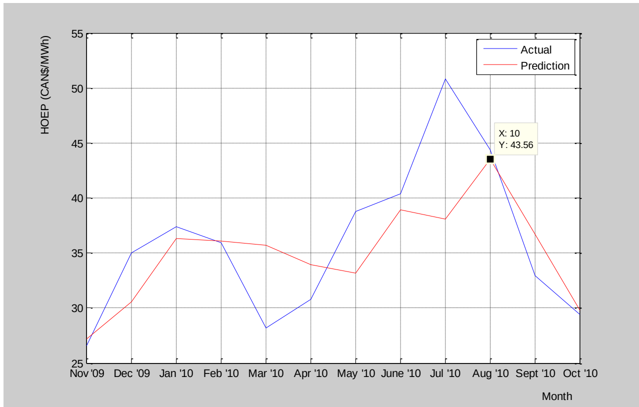

Figure 2: Actual and forecast prices of LSSVM-GA

For the sake of fair comparison, the developed model of LSSVM-GA was compared with other existing methods in Ontario for the same testing periods. Due to less research in medium term forecast, only one reference has been found for the comparison. The summary of the comparison is shown in Table 3. The result proves that the hybrid models of LSSVM-GA outperformed other models as well as the forecast produced by the Navigant Consulting Ltd. (Navigant). Navigant is engaged by the Ontario Energy Board (OEB) to provide price forecast for the Ontario electricity market. The price forecast will be used as one of the inputs to set price for the market participants.

Authors of [6] proposed methods which are based on SVM, RBF-NN, Weighted Nearest Neighbor (WNN), and Moving Average (MA). Hybrid models are also developed to improve the forecasting error. As an overall, LSSVM-GA model outperforms other existing models with MAPE of 9.43%.

Table 3
MAPE for Medium Term Forecast in the Ontario Electricity Market

Ref	Year	Method	MAPE (%)
Proposed		LSSVM+GA	9.43
[6]	2012	SVM	14.25
		RBF-NN	17.65
		WNN	32.99
		MA	32.58
		SVM/SVM	12.97
		SVM/RBF-NN	13.2
		RBF-NN/RBF-NN	14.33
		RBF-NN/SVM	16.09
		Navigant Co.	33.04

VII. Conclusion

Medium term electricity price forecasting is essential for maintenance scheduling, resources reallocation, developing investment, as well as medium term planning. Until recently, no study has investigated the application of LSSVM-GA in medium term price prediction. Hence, a hybrid model of LSSVM-GA for month-ahead electricity forecast was developed in this study to produce a month-ahead price forecast. By using the most recent features, GA optimises the input features and LSSVM parameters simultaneously. This approach minimises significant features for forecasting while optimising LSSVM parameters. The developed models of LSSVM-GA outperform other existing models for the same market and test period. Due to the lack of studies in medium term electricity price forecasting, the developed model will provide a significant contribution to the field of electricity price forecasting.

ACKNOWLEDGEMENT

This study is supported in part by the Fundamental Research Grant Scheme (FRGS) provided by Ministry of Higher Education Malaysia (FRGS/1/2017/TK04/FKE-CERIA/F00331). We also would like to dedicate our appreciation to Universiti Teknikal Malaysia Melaka (UTeM) and National Energy University (UNITEN) for providing technical and moral support throughout conducting this study.

REFERENCES

[1] X. Yan and N. a. Chowdhury, "Mid-term electricity market clearing price forecasting utilizing hybrid support vector machine and auto-regressive moving average with external input," *Int. J. Electr. Power Energy Syst.*, vol. 63, pp. 64–70, Dec. 2014.

[2] X. Yan and N. a. Chowdhury, "Mid-term electricity market clearing price forecasting: A multiple SVM approach," *Int. J. Electr. Power Energy Syst.*, vol. 58, pp. 206–214, Jun. 2014.

[3] X. Yan and N. A. Chowdhury, "A comparison between SVM and LSSVM in mid-term electricity market clearing price forecasting," in *2013 26th IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, 2013, pp. 1–4.

[4] X. Yan, S. Member, and N. a Chowdhury, "Mid-term Electricity Market Clearing Price Forecasting Using Multiple Support Vector Machine," *J. energyy*, vol. 2015, pp. 1–5, 2013.

[5] N. a. Chowdhury, "Electricity market clearing price forecasting in a deregulated electricity market," in *2010 IEEE 11th International Conference on Probabilistic Methods Applied to Power Systems*, 2010, no. 1, pp. 36–41.

[6] S. S. Torbaghan, A. Motamedi, H. Zareipour, and L. A. Tuan, "Medium-term electricity price forecasting," in *2012 North American Power Symposium (NAPS)*, 2012, pp. 1–8.

[7] C. Nunes, A. Pacheco, and T. Silva, "Statistical models to predict electricity prices," in *2008 5th International Conference on the European Electricity Market*, 2008, pp. 1–6.

[8] X. Yan and N. a. Chowdhury, "Mid-term electricity market clearing price forecasting: A hybrid LSSVM and ARMAX approach," *Int. J. Electr. Power Energy Syst.*, vol. 53, pp. 20–26, Dec. 2013.

[9] X. Yan and N. a. Chowdhury, "Hybrid SVM & ARMAX based mid-term electricity market clearing price forecasting," in *2013 IEEE Electrical Power and Energy Conference, EPEC 2013*, 2013, pp. 0–4.

[10] X. Yan and N. A. Chowdhury, "Mid-term Electricity Market Clearing Price Forecasting Using Multiple Support Vector Machine," in *2013 4th IEEE PES Innovative Smart Grid Technologies Europe (ISGT Europe)*, 2013, pp. 1–5.

[11] X. Yan and N. a Chowdhury, "Midterm Electricity Market Clearing Price Forecasting Using Two-Stage Multiple Support Vector Machine," *J. Energy*, vol. 2015, pp. 1–11, 2015.

[12] S. S. Torghaban, H. Zareipour, and L. A. T. L. A. Tuan, "Medium-term electricity market price forecasting: A data-driven approach," in *North American Power Symposium (NAPS), 2010*, 2010, pp. 1–7.

[13] V. N. Vapnik, *Statistical Learning Theory*. new York: Wiley, 1998.

[14] G. Xie, S. Wang, Y. Zhao, and K. K. Lai, "Hybrid approaches based on LSSVR model for container throughput forecasting: A comparative study," *Appl. Soft Comput.*, vol. 13, no. 5, pp. 2232–2241, May 2013.

[15] H. Wang and D. Hu, "Comparison of SVM and LS-SVM for Regression," in *2005 International Conference on Neural Networks and Brain*, 2005, no. 5, pp. 279–283.

[16] J. A. K. Suykens and J. Vandewalle, "Least Squares Support Vector Machine Classifiers," *Neural Process. Lett.*, vol. 9, no. 3, pp. 293–300, 1999.

[17] S. Li and L. Dai, "Classification of gasoline brand and origin by Raman spectroscopy and a novel R-weighted LSSVM algorithm," *Fuel*, vol. 96, pp. 146–152, Jun. 2012.

[18] J. H. Holland, *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. 1975.

[19] E. Elbeltagi, T. Hegazy, and D. Grierson, "Comparison among five evolutionary-based optimization algorithms," *Adv. Eng. Informatics*, vol. 19, no. 1, pp. 43–53, Jan. 2005.

[20] D. Zhijie, Li; Xiangdong, Comparative Research on Particle Swarm Optimization and Genetic AlgorithmLiu; Xiangdon, "Comparative Research on Genetic Algorithm, Particle Swarm Optimization and Hybrid GA-PSO," in *Computer and Information Science*, 2010, vol. 3, pp. 120–127.

[21] H. Zareipour, K. Bhattacharya, and C. a. Cañizares, "Electricity market price volatility: The case of Ontario," *Energy Policy*, vol. 35, no. 9, pp. 4739–4748, Sep. 2007.

[22] Ontario Energy Board, "Monitoring Report on the IESO-Administered Electricity Markets," 2015.