ELECTRIC LOAD CONSUMPTION FORECASTING IN DA NANG CITY USING A HYBRID OF MOVING-WINDOW CONCEPT AND SWARM INTELLIGENCE-OPTIMIZED MACHINE LEARNING REGRESSION

DỰ BÁO SỰ TIÊU THỤ ĐIỆN Ở THÀNH PHỐ ĐÀ NẴNG SỬ DỤNG MÔ HÌNH KẾT HỢP CỬA SỔ DỊCH CHUYỂN VÀ HỒI QUY MÁY HỌC ĐƯỢC TỐI ƯU BỞI TRÍ TUỆ BẦY ĐÀN

Thi Thu Ha Truong[1], Ngoc-Tri Ngo[2], Tang Thi Khanh Vy[3]

[1]*University of Technology and Education - The University of Danang; tttha@ute.udn.vn*
[2]*University of Science and Technology - The University of Danang; trinn@dut.udn.vn*
[3]*Da Nang Power Company Limited; vyttk@cpc.vn*

Abstract - Load forecasting plays an important role in the energy management system. An accurately predictive tool supports electric utilities in making decisions on purchasing and generating electric power, load switching, and infrastructure development. This study aims to develop a load forecast model combining a moving-window concept and a least squares support vector regression (LSSVR) optimized by firefly algorithm (FA). The moving-window concept is utilized to select a size of historical data to make predictions. The FA is used to optimize hyperparameters of the LSSVR for improving forecast accuracy. A real-world load dataset collected in Da Nang city is used to validate the predictive ability of the proposed MFA-LSSVR. Experimental results show that the forecast performance of the proposed model is superior to that of the moving-window LSSVR and the moving-window autoregressive integrated moving average (ARIMA). A finding of this study provides decision-makers with a potential and effective tool in energy forecasting.

Key words - Load consumption; forecast accuracy; moving-window concept; swarm intelligence; support vector machines.

Tóm tắt - Dự báo điện năng đóng vai trò quan trọng trong hệ thống quản lý năng lượng. Một công cụ dự báo hiệu quả sẽ hỗ trợ các công ty đưa ra các quyết định về mua, sản xuất điện, truyền tải, và phát triển hạ tầng. Nghiên cứu này phát triển một mô hình dự báo điện năng kết hợp lý thuyết cửa sổ dịch chuyển và máy học véc-tơ hỗ trợ (LSSVR) được tối ưu bởi thuật toán con đom đóm (FA). Cửa sổ dịch chuyển được sử dụng để lựa chọn dữ liệu lịch sử hợp lý cho dự báo. Thuật toán con đom đóm nhằm tối ưu tham số của LSSVR để cải thiện độ chính xác dự báo. Một bộ dữ liệu thực tế được thu thập ở thành phố Đà Nẵng được sử dụng để kiểm chứng khả năng dự báo của mô hình đề xuất MFA-LSSVR. Kết quả thực nghiệm cho thấy mô hình đề xuất có khả năng dự báo tốt hơn mô hình moving-window LSSVR và moving-window ARIMA. Kết quả của nghiên cứu này cung cấp một công cụ tiềm năng và hiệu quả để dự báo sự tiêu thụ điện năng.

Từ khóa - Sự tiêu thụ điện; độ chính xác dự báo; lý thuyết cửa sổ dịch chuyển; trí tuệ bầy đàn; máy học véc-tơ hỗ trợ.

1. Introduction

Vietnam has been one of the active and fastest growing economies in the region and in the world for decades. Economic growth requires a secure and affordable supply of energy to all the society participants and economic sectors. Electricity is taking up an increasing share in the final energy consumption mixture. The electricity consumption in Vietnam grew at the average rate of 10.6% per year in the period 2011-2015 [1]. An accurately electrical forecast allows making important decisions in purchasing and generating electric power, load switching, and infrastructures development [2].

Load forecast is considered as time series forecast where future values are predicted based on historical values of the load series. Finding out an effective method for load forecast has been a challenging task since the load data is non-stationary, nonlinear, and noisy. Generally, there are two approaches for predicting the electric load namely statistical techniques and artificial intelligence (AI) techniques. An autoregressive integrated moving average model (ARIMA), introduced by Box and Jenkins, is one of the most popular statistical models. These models identify and analyze the pattern of historical data to provide future measurements.

Recently, AI-based models like artificial neural networks (ANNs) have been widely applied in electric load forecast [3, 4]. These models are data-driven and non-parametric because they do not require strong model assumptions and can map any nonlinear function without a prior assumption about the properties of the data like conventional models [5]. For instance, Mordjaoui et al. (2017) [3] used a dynamic neural network to predict the daily power consumption in Algeria. Experimental results indicated that performance of their proposed model was superior to conventional models like Holt-Winters exponential smoothing and seasonal ARIMA models.

Introduced by Vapnik [6], support vector machines (SVMs) have been widely used and studied in power load forecasting [7, 8]. Compared to the ANNs, the SVMs exhibits a better generalization performance in real-world applications [9]. Zhang (2005) [8] adopted the support vector regression (SVR) to forecast one-day ahead power consumption. Results showed that the SVR achieved a greater forecast accuracy and a faster speed than the backpropagation neural network. In a study of Vrablecová et al. (2018) [7], an online SVR was used to forecast the power consumption of Irish households and enterprises every 30 minutes.

However, performance of the SVR depends on tuning its hyperparameters which are the regularization parameter and kernel function parameter. An appropriate selection of these two hyperparameters is an optimization problem. To date, many evolutionary algorithms, such as genetic algorithm (GA), particle swarm optimization (PSO), have been adopted to tune the SVR's hyperparameters. The firefly algorithm (FA), a swarm-based intelligent algorithm, has proved effective in solving optimization problems. Studies indicated the superiority of the FA against some metaheuristics including GA, PSO, differential evolution, ant colony optimization, and simulated annealing [10].

This paper develops an electric load forecast model combining the moving-window concept and the least squares support vector regression optimized by a firefly algorithm (MFA-LSSVR). The moving-window concept is adopted to select the size of historical data for training the forecast model. The FA substantially enhances the efficiency of and reduces the computational burden on the LSSVR. A load dataset collected from Da Nang city is used to validate the forecast performance of the proposed model.

The remainder of this paper is organized as follows. Section 2 elucidates the proposed load forecast model and performance evaluation method. Section 3 describes data preparation and model application. Section 4 shows performance evaluations and discussion. Finally, Section 5 provides concluding remarks and future researches.

2. Methodologies

2.1. Moving-window concept

In practice, the volume of historical time series data is large because new observations are constantly made. The latest data has a smaller impact than the older data on the training model because they are fewer than older ones [4]. Additionally, a large volume of historical data may cause difficulty in management and results in high computational complexity. Thus, training a model using very many historical data is infeasible and expensive.

To solve this problem, this study proposes a moving-window concept for forecasting time series. This concept captures the importance of recent data. The most recent data is considered, and the oldest data is neglected. A window is used to select a range of data of interest. New data is added while some old data is dropped from the window as it moves forward in time [4, 11]. The length of the moving-window is kept constant whenever the window is moved. Figure 1 displays the moving-window concept. This method, therefore, limits the volume of data that is used to train the model while retaining the efficiency and general applicability of the model.

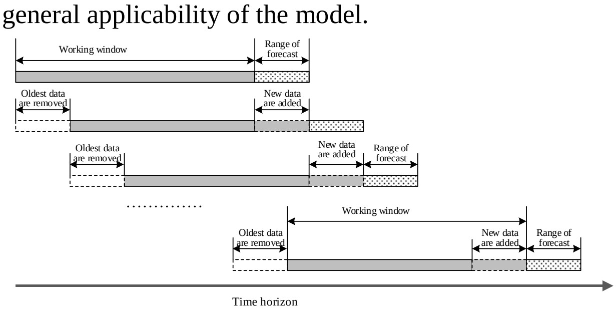

Figure 1. *The moving-window concept.*

Consider a univariate time series $\vec{x} = \{x_1, x_2, \dots, x_p, \dots, x_N\}$ where p is a fixed length of the moving-window. Generally, historical time series are transformed into three or more dimensions to explicate information that implicits in the series. This transformation process, widely known as state reconstruction [12], depends on an embedding dimension (ED) as shown in Eq. (1). The ED greatly influences the forecasting performance of the model. In this study, the optimal ED is determined by performing a sensitivity analysis.

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{p-1} \\ x_p \end{bmatrix} \xrightarrow{m} X = \begin{bmatrix} x_1 & x_2 & \cdots & x_{m-1} & x_m \\ x_2 & x_3 & \cdots & x_m & x_{m+1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_{p-m-1} & x_{p-m} & \cdots & x_{p-3} & x_{p-2} \\ x_{p-m} & x_{p-m+1} & \cdots & x_{p-2} & x_{p-1} \end{bmatrix}, \; Y = \begin{bmatrix} x_{m+1} \\ x_{m+2} \\ \vdots \\ x_{p-1} \\ x_p \end{bmatrix} \quad (1)$$

where m is the embedding dimension; X and Y are input matrix and output matrix that are tranformed from the series $\vec{x}$, respectively.

2.2. Machine learning regression optimized by swarm intelligence

2.2.1. Least squares support vector regression

The LSSVR, an advanced machine learning technique, was proposed by Suykens et al [13]. To solve regression problem, the LSSVR nonlinearly maps the input space into a high-dimensional feature space, and then run linear regression in the feature space. The LSSVR finds the solution by solving a set of linear equations rather than a quadratic programming problem, as in the standard SVR. In the LSSVR training process, a least square cost function is used to obtain linear set of equations in the dual space. By this way, the LSSVR has a lower computational burden while enabling good generalization capacity [14].

In a function estimation of the LSSVR, given a training dataset $\{x_k, y_k\}_{k=1}^{N}$, the optimization problem is formulated as

$$\min_{(\omega,b,e)} J(\omega,e) = \frac{1}{2}\|\omega\|^2 + \frac{1}{2}C\sum_{i=1}^{N} e_k^2 \quad (1)$$

subject to $y_k = \langle \omega, \varphi(x_k) \rangle + b + e_k, \quad k = 1, \dots, N$

where *J(ω,e)* is the optimization function; *ω* is the parameter of the linear approximator; $e_k \in R$ is error variables; $C \geq 0$ is a regularization constant that represents the trade-off between the empirical error and the flatness of the function; x_k is input patterns; y_k is prediction labels; and *N* is the sample size.

Since Eq. (1) is a typical optimization problem of a differentiable function with constraints, it can be solved by using Lagrange multipliers (α_k). The resulting LSSVR model for function estimation can be expressed as

Eq. (2). $$y(x) = \sum_{k=1}^{N} \alpha_k K(x, x_k) + b \quad (2)$$

where α_k, b are Lagrange multipliers and the bias term, respectively; and $K(x, x_k)$ is the kernel function. In the feature space, the kernel function can be described as Eq. (3).

$$K(x, x_k) = \sum_{k=1}^{m} g_k(x) g_k(x_k) \quad (3)$$

Typical examples of kernel function are polynomial kernel and radius basis function (RBF) kernels. The RBF is selected as a kernel function of the LSSVR in this study because the RBF kernel often yields better results compared to other proposed kernels in highly nonlinear spaces [14].

$$K(x, x_k) = \exp(-\|x - x_k\|^2 / 2\sigma^2) \quad (4)$$

where σ is the kernel parameter which controls the kernel width used to fit the training data.

2.2.2. Optimization using firefly algorithm

Although the LSSVR has been demonstrated to be effective in solving prediction problems, its predictive accuracy depends on the setting of its hyperparameters, including regularization constant (C) and the RBF kernel function (σ). In this study, the FA, a swarm intelligence-based optimization algorithm, is utilized to optimize parameters C and σ.

The FA, which was developed by Yang (2008) [15], is a stochastic and nature-inspired metaheuristic algorithm. It imitates the social behavior of fireflies in the summer sky in tropical regions. The FA has three main rules: (i) All fireflies are unisex, so any firefly is attracted to all other fireflies; (ii) Attractiveness is proportional to brightness and decreases as distance increases; a firefly moves randomly if no firefly is brighter than itself; (iii) The brightness of a firefly is determined by the landscape of the objective function.

As a firefly's attractiveness is proportional to the light intensity seen by adjacent fireflies, the attractiveness β of a firefly is defined as $\beta = \beta_0 e^{-\gamma r^2}$ (5)

where β is the attractiveness of the firefly; β_o is the attractiveness of the firefly at $r = 0$; r is the distance between the firefly of interest and any other, e is a constant coefficient, and γ is the absorption coefficient.

The distance between any two fireflies i and j at x_i and x_j, respectively, is calculated as follows

$$r_{ij} = \left\| x_i - x_j \right\| = \sqrt{\sum_{k=1}^{d} (x_{i,k} - x_{j,k})^2} \quad (6)$$

where r_{ij} is the distance between any two fireflies i and j at x_i and x_j, respectively; $x_{i,k}$ is the kth component of spatial coordinate x_i of the ith firefly; $x_{j,k}$ is the kth component of spatial coordinate x_j of the jth firefly, and d is the number of dimensions of the search space.

The movement of the ith firefly when attracted to a brighter jth firefly is determined as

$$x_i^{t+1} = x_i^t + \beta_0 e^{-\gamma r_{ij}^2} (x_j^t - x_i^t) + \alpha^t [\text{rand} - 0.5] \quad (7)$$

where x_i^{t+1} is the coordinate of the ith firefly in the $(t+1)$th iteration; x_i^t is the coordinate of the ith firefly in the tth iteration; x_j^t is the coordinate of the jth firefly in the tth iteration; γ is the absorption coefficient and was set to explore global optima, γ varies from 0 to 10. The best result obtained in the sensitivity analysis of γ is $\gamma = 1$; $\beta_o = \beta_{\min}$ is the attractiveness at $r_{ij} = 0$; α^t denotes a trade-off constant to determine the random behavior of movement; rand is a random-number generator uniformly distributed within [0, 1].

2.3. Proposed load forecast model

This section describes the proposed MFA-LSSVR model. Herein, the FA improves forecast performance of the LSSVR by tuning its parameters (C and σ). The moving-window concept limits the volume of historical data that is used to train the model while retaining the efficiency of the forecast model. The proposed MFA-LSSVR model is implemented in the MATLAB programming language, and its flowchart is presented in Figure 2.

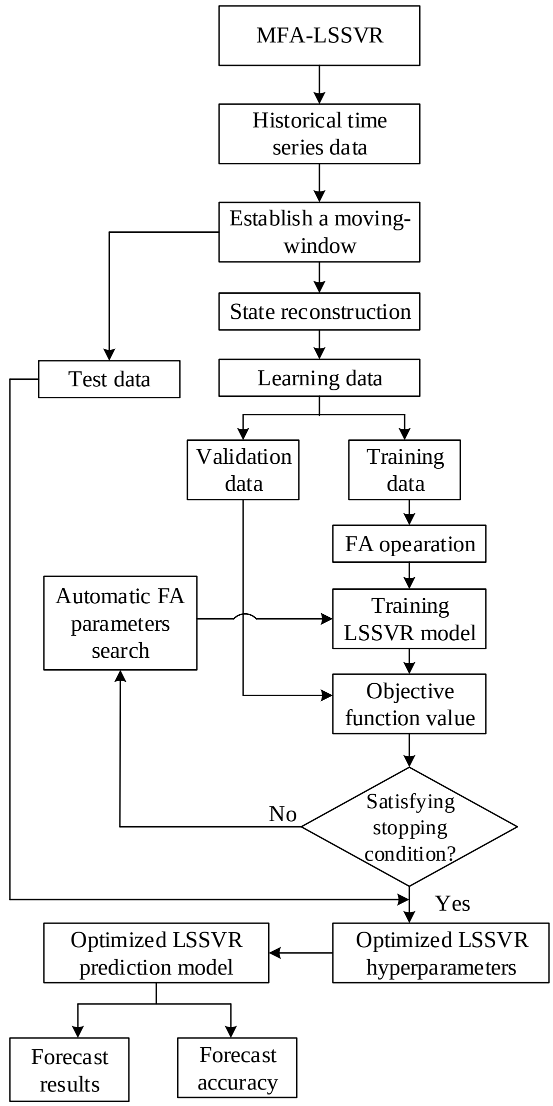

Figure 2. *The MFA-LSSVR proposed model flowchart*

The moving-window is firstly set up from the historical time series data. Its length (p) must be shorter than total dataset (N) which means ($p < N$). The moving-window contains the learning data, so the amount of the learning data is p and the number of test data is (N-p). This study performs one-step ahead forecast, so the window moves forward in (N-p) times. With a specific value of ED (m), the state reconstruction is made. This process results in an input matrix and an output matrix [refer to Eq. (1)]. Then, the learning data is divided into the training data and validation data. The training data is used to train the FA-LSSVR model while the validation data is used to optimize the MFA-LSSVR model. The objective function of the MFA-LSSVR is established based on the validation data as shown in Eq. (8). In this study, root means square error (RMSE) is used to designate the objective function. Herein, the FA is applied to simultaneously and automatically identify the optimal values of LSSVR's parameters (C and σ) by minimizing the RMSE value. The optimization process ends when the stopping condition is satisfied and the optimal values of C and σ are determined. Finally, test data is used to test the performance of the optimized MFA-LSSVR forecast model.

$$f(C,\sigma) = RMSE_{Val} = \sum_{i=1}^{n} \sqrt{\frac{1}{n}(y' - y)^2} \tag{8}$$

where $RMSE_{Val}$ is the root mean square error calculated according to the predicted (y') and actual (y) values, respectively, based on the validation data; n is the sample size of validation data.

2.4. Forecast accuracy evaluation

To assess forecast accuracy of forecast models, criteria are used including root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and synthesis index (SI). The lower values of RMSE, MAE and MAPE indicate the better forecast accuracy. Their corresponding equations are as follows.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y' - y)^2} \tag{9}$$

where y is the actual observation; y' is the predicted value; and n is the number of predictions.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y - y'| \tag{10}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y - y'}{y}\right| \tag{11}$$

$$SI = \frac{1}{m}\sum_{i=1}^{m}\frac{P_i - P_{\min,i}}{P_{\max,i} - P_{\min,i}} \tag{12}$$

where m is the number of performance measures and P_i is the ith performance measure. SI is used to rank performance of each model. SI ranges from 0 to 1 and an SI value of close to 0 indicates a highly accurate predictive model.

3. Data preparation and model application

3.1. Data preparation

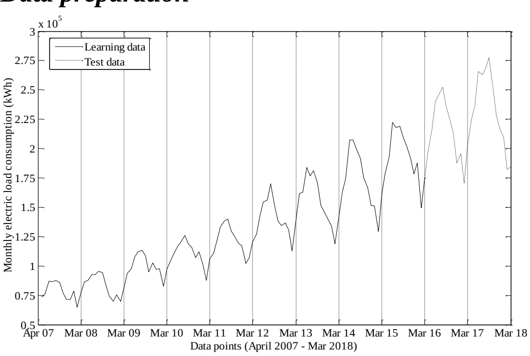

Figure 3. *The monthly load consumption of households*

In this study, the performance of the MFA-LSSVR is validated by a real-world dataset collected in Da Nang city. The dataset includes 132 observations of monthly load consumption of households in Da Nang city ranging from April 2007 to March 2018. 108 observations of the dataset are used as learning data and the remaining 24 observations are used for test data. The actual load values are displayed in Figure 3.

3.2. Model application

Two main purposes in this section are to determine the optimal length of the moving-window and compare the performance of the proposed model with the moving-window least squares support vector regression (MLSSVR) and the moving-window autoregressive integrated moving average (MARIMA). The implementation consists of two stages as follows.

Stage 1 – Determine the optimal length of moving-window – The moving-windows with different lengths are used to validate the performance of the MFA-LSSVR. For each length of the moving-window, 12 observations in 2015 were used as test data to evaluate the forecast performance.

Step 2 - Compare the performance of predictive models – The performance of predictive models including MFA-LSSVR, MLSSVR, and MARIMA are compared using the same optimal length of moving-window. The test data includes 24 observations in the period of April 2016-March 2018. The initial settings of the proposed MFA-LSSVR is presented in Table 1.

Table 1. *The MFA-LSSVR model parameters*

Component	Name	Values/ Setting
Moving-window concept	Embedding dimension	≥ 3
FA	No. of fireflies	60
	Max. generation	30
	Attractiveness	0.1
	Absorption coefficient	1
	Objective function	RMSE
	Training partition	70%
	Validation partition	30%
LSSVR	Rang of C	$[10^{-3};10^{12}]$
	Range of σ	$[10^{-3};10^{12}]$

Table 2. *Forecast performance of different lengths of moving-window*

Moving-window (month)	RMSE (kWh)	MAE (kWh)	MAPE (%)	SI (Rank)
48	9,234.723	7,411.854	4.30%	0.827 (7)
54	9,090.986	7,774.420	4.51%	0.972 (8)
60	8,908.451	7,386.482	4.32%	0.765 (6)
66	8,445.002	7,408.289	4.27%	0.660 (5)
72	8,192.981	7,170.801	4.15%	0.508 (3)
78	8,381.772	6,985.845	4.10%	0.478 (2)
84	8,497.339	7,062.112	4.15%	0.540 (4)
90	7,502.142	6,344.792	3.72%	0.000 (1)

As mentioned in Section 2.1, the ED influences the forecast accuracy of a model. By using a particular moving-window length of 60 and taking the ED value in [3;18], a sensitivity analysis shows that the optimal ED is 14. The performance of different moving-window lengths with the optimal ED is showed in Table 2. The SI is zero indicating that the optimal length is 90 (months). Table 3 compares the performance of predictive models using the optimal ED and the optimal moving-window length.

Table 3. *Performance comparison among predictive models*

Model	Test phase			Improved by the MFA-LSSVR (%)		
	RMSE (kWh)	MAE (kWh)	MAPE (%)	RMSE	MAE	MAPE
MARIMA	29,081.395	21,978.633	9.93	65.23	67.50	66.16
MLSSVR	67,801.103	60,250.023	25.51	85.09	88.14	86.83
MFA-LSSVR	10,110.680	7,143.743	3.36	-	-	-

4. Performance evaluation and discussion

Table 2 shows the forecast performance using various lengths of moving-window. When the length is 90, the MFA-LSSVR has the smallest values of RMSE, MAE, and MAPE. The MFA-LSSVR yields an MAE of 6,344.792 kWh, which is significantly lower than that obtains by using other moving-window lengths. Thus, decision-makers should be suggested taking 90 historical observations to make a next prediction.

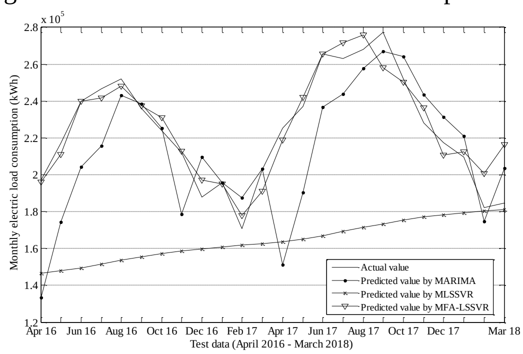

Figure 4. *Actual and predicted values using test data*

With the same optimal ED value (14) and the same moving-window length (90 months), the performance measures obtained by the proposed MFA-LSSVR are superior to those obtained by the MSSVR and MARIMA. The proposed model has the lowest MAPE (3.36%) and the lowest MAE (7,143.743 kWh) compared to other models. The performance measures of the MARIMA model are better than those of the MLSSVR model. The ARIMA model yields a MAPE of 9.93% which is significantly lower than the MLSSVR (25.51%). Overall, the error rates of the MFA-LSSVR model are 65.23% - 88.14% better than those of the MARIMA and the MLSSVR. Figure 4 displays actual values and predicted values of the load series. The predicted values exhibited by the MFA-LSSVR capture and are closer to the actual values than those obtained by the MARIMA and the MSSVR.

5. Conclusions and recommendations

This study proposes a load forecast model that integrates a moving-window and a LSSVR model optimized by the FA for predicting monthly electric load consumption of households in Da Nang city. The performance of the MFA-LSSVR model is compared to that of the MARIMA and the MLSSVR. A sensitivity analysis is performed to select the optimal embedding dimension and the optimal length of moving-window.

Compared to the MARIMA and the MLSSVR, the MFA-LSSVR achieves the lowest error rates in terms of RMSE, MAE, and MAPE. The error rates improved by the proposed model is 65.23% - 88.14%. By integrating the moving-window into the forecast model, decision-makers could save time when making predictions. Therefore, the proposed model could be used as an efficient and dynamic forecast tool.

For future research, some factors affecting the load consumption like outdoor temperature, population, seasonality need to be considered in the forecast model. Moreover, a model which can make multiple step ahead prediction should be developed.

Acknowledgment: This research is funded by Funds for Science and Technology Development of the University of Danang under grant number 57/HĐ-KHCN-2017. The authors would like to thank the Da Nang Power Company Limited for data acquisition.

REFERENCES

[1] E.o. Denmark, M.o.I.a. Trade, *Vietnam energy outlook report,* Vietnam, 2017.

[2] J.H. Chow, F.F. Wu, J.A. Momoh, *Applied Mathematics for Restructured Electric Power Systems: Optimization*, Control, and Computational Intelligence, Springer 2005.

[3] M. Mordjaoui, S. Haddad, A. Medoued, A. Laouafi, Electric load forecasting by using dynamic neural network, *International Journal of Hydrogen Energy* 42(28) (2017) 17655-17663.

[4] J. Yang, H. Rivard, R. Zmeureanu, On-line building energy prediction using adaptive artificial neural networks, *Energy and Buildings 37(12)* (2005) 1250-1259.

[5] C.-J. Lu, T.-S. Lee, C.-C. Chiu, Financial time series forecasting using independent component analysis and support vector regression, *Decision Support Systems 47(2)* (2009) 115-125.

[6] V.N. Vapnik, The nature of statistical learning theory, Springer-Verlag, New York, 1995.

[7] P. Vrablecová, A. Bou Ezzeddine, V. Rozinajová, S. Šárik, A.K. Sangaiah, Smart grid load forecasting using online support vector regression, *Computers & Electrical Engineering 65* (2018) 102-117.

[8] M.-G. Zhang, Short-term load forecasting based on support vector machines regression, International Conference on Machine Learning and Cybernetics, *IEEE, Guangzhou,* China, 2005.

[9] S.K. Pal, C.S. Rai, A.P. Singh, Comparative Study of Firefly Algorithm and Particle Swarm Optimization for Noisy Non-Linear Optimization Problems, *International Journal of Intelligent Systems and Applications 4(10)* (2012) 50-57.

[10] I. Fister, I. Fister Jr, X.-S. Yang, J. Brest, A comprehensive review of firefly algorithms, *Swarm and Evolutionary Computation 13* (2013) 34-46.

[11] R.-P. Mundani, J. Frisch, V. Varduhn, E. Rank, A sliding window technique for interactive high-performance computing scenarios, *Advances in Engineering Software* 84 (2015) 21-30.

[12] N.-T. Ngo, T.T.H. Truong, Time Series Analysis for Improving Accuracy of Stock Price Forecasts Using Least Squares Support Vector Regression, *International Conference for Young Researchers in Economics and Business*, Da Nang, Vietnam, 2017, pp. 637-643.

[13] J.A.K. Suykens, T.V. Gestel, J.D. Brabanter, B.D. Moor, *J. Vandewalle, Least squares support vector machines*, World Scientific, Singapore, 2002.

[14] Z. Min, T. Huanq, Short Term Load Forecasting with Least Square Support Vector Regression and PSO, in: J. Zhang (Ed.), *Communications in Computer and Information Science*, Springer Heidelberg Dordrecht London NewYork2011, pp. 124-132.

[15] X.-S. Yang, Firefly algorithm, Luniver Press, Bristol, U.K, 2008.

(The Board of Editors received the paper on 28/9/2018, its review was completed on 19/10/2018)