The Forecast of PM10 Pollutant by Using a Hybrid Model

Ronnachai Chuentawat, Nittaya Kerdprasop, and Kittisak Kerdprasop

***Abstract*—This research aims to study the forecasting model to predict the 24-hour average PM10 concentration in the Northern region of Thailand. This research presents a hybrid model that combines the autoregressive part of the Autoregressive Integrated Moving Average (ARIMA) model with the support vector regression technique. The data used in this study are the 24-hour average PM10 concentration from 3 locations. Each of the data sets is the daily univariate time series during 1st January to 3^{1th} May 2016. We evaluate predictive performance of our hybrid model using the two measurements: Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE). The performance of our hybrid model has been compared against the ARIMA model. From the experimental results, we found that a hybrid model has lower RMSE and MAPE than the ARIMA model for all three data sets. Therefore, we concluded that our hybrid model can be used to forecast the 24-hour average PM10 concentration in the Northern region of Thailand.**

***Index Terms*—PM10, ARIMA model, support vector regression, hybrid model.**

I. INTRODUCTION

The PM10 is particulate matter of size 10 micrometers or less in diameter. It is the impure matter that can cause air pollution. If the PM10 concentration in the air exceeds the standard criterion, it will have a negative effect on the respiratory and may cause serious respiratory illness to death. Therefore, the Pollution Control Department of Thailand defines the standard concentration of PM10 in the air that the average 24-hour should not exceed 120 micrograms per cubic meter ($\mu g/m^3$). Due to the negative effect of PM10 on health, many researchers have tried to forecast the concentration of PM10 with various techniques. Some prominent works are summarized as follows.

Chen and Pai [1] studied about predicting hourly particulate matter (PM) including PM10 and PM2.5 concentrations in Dali area of Taichung City, Taiwan. They studied by comparing one-variable grey differential equation model (GM(1,1)) and the back–propagation artificial neural network (BPNN) model. They evaluated the result by using the mean absolute percentage error (MAPE), mean squared error (MSE) and root mean squared error (RMSE). The result indicated that the GM (1,1) model could predict the hourly PM variation precisely even compared with the BPNN model.

Lin *et al.* [2] studied and developed a support vector regression with logarithm preprocessing procedure and immune algorithms (SVRLIA) model to forecast the concentrations of three air pollutants, namely particulate matter (PM10), nitrogen oxide, (NOx), and nitrogen dioxide (NO2). They applied data smoothing as preprocessing procedure before optimizing with the immune algorithm in order to forecast more accurately the air pollutants. Experimental results reveal that the SVRLIA model can accurately forecast concentrations of air pollutants.

Ul-Saufie *et al.* [3] studied to improve prediction of Multiple Linear Regression (MLR) and Feedforward backpropagation (FFBP) by combining them with principal component analysis (PCA) for predicting future (next day, next two days, and next three days) PM10 concentration in Negeri Sembilan, Malaysia. Prediction Accuracy (PA), Coefficient of Determination (R2) and Index of Agreement (IA) were used as metrics to assess the accuracy of the models. Normalized Absolute Error (NAE) and RMSE were also used to evaluate the performance of the models. The results show that PCA combined with MLR and PCA combined with FFBP can improve accuracy on predicting PM10 concentrations three days in advance.

Wongsathan and Seedadan [4] studied and developed the hybrid model that combined the ARIMA model with the neural networks (NNs) model to forecast the PM10 in Chiang Mai city moat area of Thailand. The errors of the ARIMA model were used to generate the NNs model. After that, the predictive values of the NNs model were merged with the predictive values of the ARIMA model to output the final predictive values of the hybrid model. The experimental results demonstrated that the hybrid model outperformed the single NNs and the single ARIMA.

This research also aims to develop the forecasting model to predict the PM10 concentration in the North region of Thailand. Data are collected from the PM10 measurement centers. These data were published via the website of the Pollution Control Department of Thailand in the page of data archives for air and noise pollution (available from: http://aqnis.pcd.go.th/). This daily data set is a univariate time series that has only one observed variable and the values of the variable are the average 24-hours of the PM10 concentration. Some values are missing, therefore, we choose the three data sets that have the less missing value. Three data sets are the data in the area of Sripoom district in Chiangmai province, Bandong district in Lampang province, and Muang district in Lampoon province. We solve the missing value problem by using the mean between previous day and the next day and we use the data from 1st January until 31th May 2016, which be an amount of 152 observed values.

To predict the univariate time series, we firstly deploy the ARIMA method. ARIMA was a popular forecasting model to predict a univariate time series due to its high accuracy [5], [6]. The ARIMA model is good at capturing linear patterns, but it cannot easily capture the non-linear pattern [4], [6], [7]. For

Manuscript received February 15, 2017; revised April 10, 2017.

Ronnachai Chuentawat is with the Nakhonratchasima Rajabhat University, Thailand (e-mail: c_ronnachai@hotmail.com).

Nittaya Kerdprasop and Kittisak Kerdprasop are with the Suranaree University of Technology, Thailand (e-mail: nittaya@sut.ac.th, kerdpras@sut.ac.th).

doi: 10.18178/ijfcc.2017.6.3.504

capturing non-linear patterns, the Support Vector Regression (SVR) model, developed from the machine learning field, can do a better job than ARIMA. However, SVR may not easily capture the linear pattern [6], [8]. Therefore, we propose to use the hybrid model that integrates the autoregressive of the ARIMA with the SVR that has been searched for its optimized parameters with the genetic algorithm. The method to hybrid the autoregressive and SVR is a novel technique for predicting the average-24 hour of PM10 concentration. We finally evaluate the accuracy of the hybrid model based on the RMSE and MAPE measurements.

II. Methodology

The purpose of this research is to develop a hybrid model that integrates the autoregressive of the ARIMA for capturing linear patterns and the SVR technique for capturing non-linear patterns. Therefore, the assumption of this research is that the hybrid model can capture both linear and non-linear patterns and thus the forecasting performance should be more accurate than the ARIMA model that is solely good at linear pattern capture. We evaluate our assumption by measuring RMSE and MAPE of the hybrid model, and compare with the ARIMA model.

To create the hybrid forecasting model using the SVR technique, it is necessary to specify proper parameters. This research uses a genetic algorithm (GA) to find a suitable value for each of the SVR parameters. This model is thus called the hybrid GASVR model. The conceptual framework can be shown as in Fig. 1.

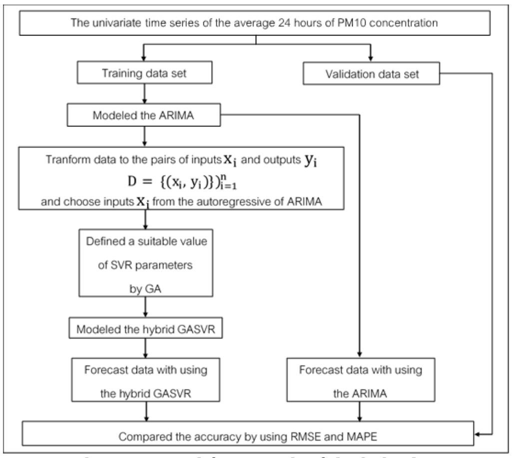


Fig. 1. The conceptual framework of the hybrid GASVR.

We use R programming as a research tool to develop the univariate time series forecasting model, called hybrid GASVR, composing of two main phases: the first phase is modeling with ARIMA and the second phase is the generation of final model with SVR that has been optimized its parameters with GA. The ARIMA is a model derived from Box and Jenkins method [9] that has been widely used in statistical analysis, whereas the SVR is the machine learning technique that was introduced by Vapnik [10]. The two forecasting models from two different paradigms require different steps to model. These steps can be explained as follows.

A. The Steps to Develop the ARIMA Model

The ARIMA model can predict the future data from two information: the first one is the autoregressive (AR), which is the data prediction at any time depending on the previous data, and the second one is the moving average (MA), which is the data prediction that depends on the previous errors. The general term of ARIMA can be presented by the backward shift operator (B) in the following equation [5].

$$\theta_P(B)(1-B)^d Y_t = w_q(B)a_t \tag{1}$$

where

$$BY_t = Y_{t-1}\ ,\ B^k Y_t = Y_{t-k}$$

and

$$w_q(B) = 1 - w_1 B - w_2 B^2 - \cdots - w_q B^p$$

$$\theta_p(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_p B^p$$

The development of the ARIMA model needs analyzing step to define a suitable form of ARIMA(p, d, q), in which p, d and q are integer. The parameter p is for autoregressive. The parameter d is for the transformation from non-stationary to be a stationary time series. The parameter q is the moving average. The steps to develop the ARIMA model can be shown as in Fig. 2. To model the ARIMA, we use the R package named "forecast" with the development details described as follows.

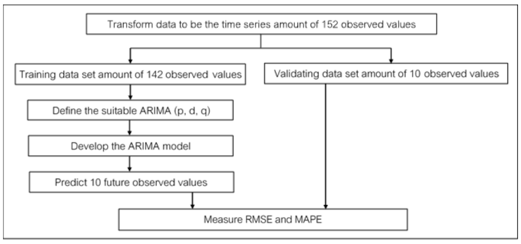


Fig. 2. The steps to develop the ARIMA model.

1) Transform the daily data during 1st January to 31th May 2016 to be time series by using "ts()" function in R.
2) Split the time series into 2 parts. The first part is the training data: 1st January to 21th May 2016. The second part is the validating data: 22th to 31th May 2016.
3) Define suitable parameters of ARIMA (the parameters p, d and q) by using "auto.arima()" function.
4) Generate ARIMA model by using "arima()" function.
5) Predict 10 observed values in validating data set, 22th to 31th May 2016, by using "predict()" function.
6) Measure the forecasting accuracy with RMSE and MAPE metrics that can be computed as in equations 2 and 3 [5].

$$\text{RMSE} = \sqrt{\tfrac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2} \tag{2}$$

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n} \left|\, 100\,\%\ \frac{y_t - \hat{y}_t}{y_t} \right| \tag{3}$$

where y_t is the observed value at time t, $\hat{y}_t$ is the predictive value at time t, and n is the amount of the predicted time period.

B. The Steps to Develop the Hybrid GASVR Model

The SVR is a modification of the support vector machine (SVM), which is a classification method. The SVR is used to forecast numeric values instead of the categorical classification as traditionally been done by SVM. Therefore, the SVR focuses on finding a linear relationship mapping the input vector X in n-dimensions to the output y by using the linear regression of the SVR that can be shown as equation 4 [11].

$$f(x) = w^T x + b \quad (4)$$

where w and b are the slope and offset of the regression line, respectively. We can define w and b by using Lagrange multipliers (α_i and α_i^*) and equation 4 can be transformed in term of Lagrange multipliers as shown in equation 5 [11].

$$f(x) = w_0^T x + b = \sum_{i=1}^{l} (\alpha_i - \alpha_i^*)\, x_i^T x + b \quad (5)$$

The equation 5 is the formation of linear regression. For the non-linear case, we can map the original input space to high dimensional feature space by using kernel function. The mapping can be the multiplication of the vectors x_i and x_j. Therefore, equation 5 can be reformulated as non-linear regression by using kernel function as shown in equation 6 [11].

$$f(x) = \sum_{i=1}^{l} (\alpha_i - \alpha_i^*)\, k(x_i, x) + b \quad (6)$$

The SVR modeling requires 3 important parameters: the cost (C), the epsilon (ε), and the parameter of kernel function. This research uses a linear kernel to train the SVR model for all of the data sets because we found from experimentation that it is the most suitable kernel. This linear kernel does not need any parameter. Therefore, the parameters of SVR in this research are only C and ε. A proper setting of C and ε can significantly increase the model accuracy. Thus, this research deploys the genetic algorithm (GA), which uses the evolution theory to find the optimal solution [12], to search for the optimal C and ε parameters. The GA is initiated by random a solution to be a member of the initial population. Each of the members is encoded and called a Chromosome. Each chromosome is consisted of genes and each gene contains a set of suitable parameter values for SVR. A solution of parent population is used to generate the next generation, called the offspring, by using the genetic operations including the crossover, mutation, and selection. The final step is the replacement operation such that the parent chromosomes are replaced by those of the offspring and the whole process is repeated until the termination condition is met.

The proposed forecasting model was generated by integrating the autoregressive of ARIMA and the SVR that parameters are optimized with GA (called GASVR), and hence the model is named the hybrid GASVR. The autoregressive form was used to define the set of the lag time observed values to be inputs for the GASVR and steps to generate the hybrid GASVR can be shown as in Fig. 3. The hybrid GASVR was generated by calling the"svm()" function which is available in the "e1071" package of the R language. According to steps in Fig. 3, the implementation can be described as follows.

1) Divide the data into a training data set and a validating data set. After that, transform a training set format to be a pair of input vectors and corresponding targets ($D = \{(x_i, y_i)\}_{i=1}^{n}$). This is the autoregressive form of the ARIMA model.
2) Generate the ARIMA model from a training data set.
3) Define the optimal C and ε parameters by applying the GA. In the R language, it has "rgba()" function in "genalg" package for finding the optimal solution by GA. The operation to find the optimal solution of GA can be described as follows.
   - Random the chromosomes to be a member of the initial population with 200 chromosomes (default number).
   - Evaluate fitness value. The evaluation of fitness in each chromosome is for choosing the offspring that will be the next generation. In our research, we use MAPE as an evaluation function.
   - Check the condition to terminate the GA operation. In this research, we set the GA to terminate its operation when it generates 100 generations (default number).
   - When the termination condition has not been met, perform the crossover operation. This step is to build the offspring by using 2 parent chromosomes for the single-point crossover operation and set the rate of crossover to be 0.8 (default number).
   - Perform the mutation operation. This step is to mutate the offspring by using only single parent chromosome. The mutation is used to avoid the problem of local optimum and set the rate of mutation to be 0.01 (default number).
   - Perform the selection operation. This step is for selecting the offspring that will be the next generation. This research uses the Roulette wheel method for the selection.
   - Perform the replacement operation. This step is the replacement over the existing population by using the new set of population that has the fitness value better than the old population set. We replace the parent by the offspring at the rate of 20% (default number).
   - Generate the next generation of GASVR model and repeat the steps from 3.2 to 3.7 until the termination condition is met.
4) When we receive the optimal value of C and ε from GA, we use them as parameters for SVR and train the SVR with the training data set to generate the hybrid GASVR model.
5) Predict 10 forecasting values by using the hybrid GASVR.
6) Measure RMSE and MAPE forecasting errors. These two measured errors will be compared with the errors of the ARIMA model to evaluate the accuracy of the hybrid GASVR model.

III. Experimental Results

This research uses 3 data sets according to three different locations in Northern Thailand. These district locations are Sripoom in Chiangmai, Bandong in Lampang, and Muang in Lampoon. All data sets are the univariate time series that has an observed value to be the average 24 hours of PM10 concentration in the air. Each data set has 152 observed values between 1st January to 31th May 2016. The data are divided into 2 parts: the first part is the first 142 observed value to be used as training data set, and the second parts is the 10 remaining values to be used as validating data set. This research experiments with 3 data sets following the same steps of the conceptual framework shown in Fig. 1. The results of the experiments are as follows.

A. The Result of the ARIMA Model

After exploring with the auto.arima() function to find suitable parameters for ARIMA(p, d, q), we have found that the first data set of a district in Chaingmai has an ARIMA(2, 1, 2) formation, the second data set of a district in Lampang has an ARIMA(0, 1, 2) formation and the last data set of main district in Lampoon province has an ARIMA(1, 0, 2) formation. We then use a suitable formation of each data set to generate the ARIMA model to forecast the 10 daily PM10 concentration during 22th to 31th May 2016. After that, we compare the 10 predictive values with the actual values to measure an error in terms of RMSE and MAPE. The results of error measurement are shown in Table I.

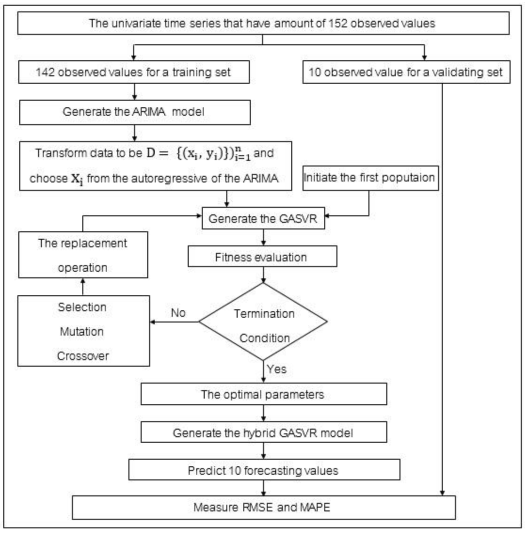


Fig. 3. The steps to generate the hybrid GASVR model.

TABLE I: THE RMSE AND MAPE OF THE ARIMA MODEL

Data sets	RMSE	MAPE
First data set: Sripoom, Chiangmai	7.884052	22.39237
Second data set: Bandong, Lampang	20.52467	27.95113
Third data set: Muang, Lampoon	18.90313	67.58202

B. The Result of the Hybrid GASVR Model

Due to a series of steps to generate the hybrid GASVR model, we use the autoregressive of the ARIMA model to define an input of GASVR, and from the result of the ARIMA model, we know the suitable ARIMA formation of each data set. Thus we can analyze the autoregressive term of each data set by using equation 1 as follows.

1) The first data set, Sripoom district in Chiangmai, has an ARIMA (2, 1, 2) formation. Therefore, we use p=2 and d=1 in equation 1 and ignore the q because it is the order of moving average, which plays no role in our modeling method. The derived equation can be shown as follow.

$$(1-\theta_1 B-\theta_2 B^2)(1-B)Y_t = Moving\ average\ term$$
$$(1-\theta_1 B-\theta_2 B^2-B+\theta_1 B^2+\theta_2 B^3)Y_t = Moving\ average\ term$$
$$Y_t = (1+\theta_1)Y_{t-1}+(\theta_2-\theta_1)Y_{t-2}-\theta_2 Y_{t-3} + Moving\ average\ term \quad (7)$$

From equation 7, we can conclude that the inputs of GASVR model for the first data set are the lag time observed value at t-1, t-2 and t-3 (Y_{t-1}, Y_{t-2}, Y_{t-3}).

2) The second data set, Bandong district in Lampang, has an ARIMA(0, 1, 2) formation. Therefore we use p=0 and d=1 in equation 1 and ignore the q. The derived equation is as follow.

$$(1-B)Y_t = Moving\ average\ term$$
$$Y_t = Y_{t-1} + Moving\ average\ term \quad (8)$$

From equation 8, we can conclude that the input of GASVR model for the second data set is the only lag time observed value at t-1 (Y_{t-1}).

3) The third data set, Muang district of Lampoon, has an ARIMA(1, 0, 2) formation. Therefore, we use p=1 and d=0 in equation 1. The derived equation can be shown as follow.

$$(1-\theta_1 B)Y_t = Moving\ average\ term$$
$$Y_t = \theta_1 Y_{t-1} + Moving\ average\ term \quad (9)$$

From equation 9, we can conclude that the input of GASVR model for the third data set is the only lag time observed value at t-1 (Y_{t-1}).

When we know the input of GASVR model, we can then use the genetic algorithm to find the optimal C and ε**.** Their values for each data set can be shown in Table II.

TABLE II: THE OPTIMAL PARAMETERS OF EACH DATA SET

Data sets	Optimal C	Optimal ε
First data set: Sripoom, Chiangmai	4.7738	0.9625
Second data set: Bandong, Lampang	17.6048	0.2256
Third data set: Muang, Lampoon	17.4739	0.1639

When we know the optimal C and ε, we can thus use them to generate the optimal hybrid GASVR model. After that, we apply the model to predict the 10 observed values and measure the RMSE and MAPE (shown in Table III).

TABLE III: THE RMSE AND MAPE OF THE ARIMA MODELS

Data sets	RMSE	MAPE
First data set: Sripoom, Chiangmai	4.525136	11.91260
Second data set: Bandong, Lampang	12.39131	18.82038
Third data set: Muang, Lampoon	5.504721	18.12998

C. A Comparison between the Hybrid GASVR and the ARIMA Models

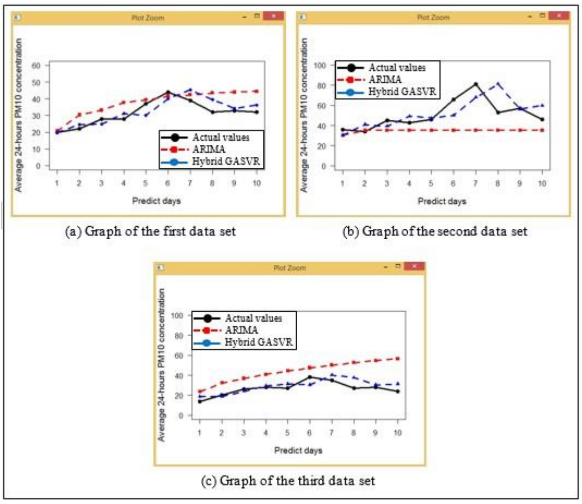
From the comparison regarding the accuracy of the hybrid GASVR model and the ARIMA model, we have found that the hybrid GASVR model yields the lower RMSE and MAPE than the ARIMA model for all three data sets as summarized in Table IV. From all three data sets in Chiangmai, Lampang, and Lampoon provinces, the accuracy performance of the hybrid GASVR is better that ARIMA model about 46.80%, 32.67%, and 73.17%, respectively. When we plot the 10 predictive values of ARIMA and hybrid GASVR against the actual values for visual comparison, the graphs can be shown as in Fig. 4. From the graphs, we found that the forecasting trends of the hybrid GASVR is more similar and align to the actual values than the ARIMA for all data sets. Therefore, we can conclude that the hybrid GASVR model is more accurate than the ARIMA model for the forecasting of the average 24-hour PM10 concentration in the Northern region of Thailand.

The high performance of hybrid GASVR may due to the fact that the PM10 concentration time series data probably consist of both complex linear and non-linear patterns. It has been known that ARIMA model is only good at capturing linear patterns [4], [6], [7]. On the contrary, the GASVR that is based on the machine learning technique tends to better capturing non-linear patterns than the linear patterns [7], [8].

Neither ARIMA nor GASVR technique alone is adequate in modeling and predicting time series data that consisting of linear and non-linear patterns. Therefore, when we integrate both techniques to generate the hybrid model, it can capture both kinds of patterns and thus yield accuracy higher than the ARIMA model.

TABLE IV: THE COMPARISON OF RMSE AND MAPE EVALUATED FROM THE HYBRID GASVR AND THE ARIMA MODELS

Data sets	ARIMA		Hybrid GASVR		Accuracy performance
	RMSE	MAPE	RMSE	MAPE	
First data set: Chiangmai	88.7	22.39	525.4	11.91	+46.80%
Second data set: Lampang	52.20	27.95	39.12	18.82	+32.67%
Third data set: Lampoon	90.18	67.58	50.5	18.13	+73.17%



(a) Graph of the first data set (b) Graph of the second data set (c) Graph of the third data set

Fig. 4. The comparison graphs of all three data sets.

IV. CONCLUSION AND FUTURE WORK

A. Conclusion

This research present the development of the hybrid forecasting model, called the hybrid GASVR, by combining the autoregressive and support vector regression optimized with genetic algorithm. This hybrid model has been applied to forecast the 24-hour average PM10 concentration in the Northern region of Thailand. The performance of the hybrid GASVR has been tested with three sample data sets that are the univariate time series containing the observed values of the 24-hour average concentration of PM10 in three locations in the Northern provinces of Thailand. These values are to be recorded daily from 1st January until 31th May 2016. From the experimental results, we found that our hybrid GASVR model is more accurate than the ARIMA model for predicting daily PM10 values in all three locations. The high performance of a hybrid method is consistent to the findings reported by other researchers [2], [7], [8] that tried a different combination of hybrid scheme. Therefore, we can conclude that our hybrid GASVR model can be an accurate model to forecast the 24-hour average concentration of PM10.

B. Future Work

Based on our results, although our hybrid GASVR is more accurate than the ARIMA model, the forecasting error of our hybrid model is at high level in some data set. Therefore, our future work is the model improvement by applying other techniques to increase the forecasting accuracy. The finding of new factors highly correlating with the PM10 concentration to be used as inputs of the forecasting model is also our plan for future research for the main purpose of reducing the forecasting error.

ACKNOWLEDGMENT

We are grateful to the Pollution Control Department of Thailand for publicizing the data that had been used in this study.

REFERENCES

[1] L. Chen and T. Y. Pai, "Comparisons of GM (1, 1), and BPNN for predicting hourly particulate matter in Dali area of Taichung City, Taiwan," *Atmospheric Pollution Research*, vol. 6, pp. 572-580, 2015.

[2] K. P. Lin, P. F. Pai, and S. L. Yang. "Forecasting concentrations of air pollutants by logarithm support vector regression with immune algorithms," *Applied Mathematics and Computation*, vol. 217, pp. 5318–5327, 2011.

[3] A. Z. Ul-Saufie, A. S. Yahaya, N. A. Ramli, N. Rosaida, and H. A. Hamid, "Future daily PM10 concentrations prediction by combining regression models and feedforward backpropagation models with principle component analysis (PCA)," *Atmospheric Environment*, vol. 77, pp. 621-630, 2013.

[4] R. Wongsathan and I. Seedadan, "A hybrid ARIMA and neural networks model for PM-10 pollution estimation: The case of Chiang Mai city moat area," *Procedia Computer Science*, vol. 86, pp. 273–276, 2016.

[5] Y. Wang, J. Wang, G. Zhao, and Y. Dong, "Application of residual modification approach in seasonal ARIMA for electricity demand forecasting: A case study of China," *Energy Policy*, vol. 48, pp. 284-294, 2012.

[6] X. Wang and M. Meng. "A hybrid neural network and ARIMA model for energy consumption forecasting," *Journal of Computers*, vol. 7, no. 5, pp. 1184-1190, 2012.

[7] P. F. Pai and C. S. Lin, "A hybrid ARIMA and support vector machines model in stock price forecasting," *Omega*, vol. 33, pp. 497-505, 2005.

[8] F. Zhang, C. Deb, S. E. Lee, J. Yang, and K. W. Shah. "Time series forecasting for building energy consumption using weighted support vector regression with differential evolution optimization technique," *Energy and Buildings*, vol. 126, pp. 94-103, 2016.

[9] G. E. P. Box and G. Jenkins, *Time Series Analysis, Forecasting and Control*, 3rd ed. San Francissco: Holden-Day, 1990.

[10] V. Vapnik, *The Nature of Statistical Learning Theory*, 1st ed. Springer-Verlag, 1995.

[11] P. Bagheripour, A. Gholami, M. Asoodeh, and M. V. Asadi. "Support vector regression based determination of shear wave velocity," *Journal of Petroleum Science and Engineering*, vol. 125, pp. 95-99, 2015.

[12] J. H. Holland, *Adaptation in Natural and Artificial Systems*, 1st ed. Michigan: University of Michigan Press, 1975.

Ronnachai Chuentawat received the PhD in computer engineering from Suranaree University of Technology, Thailand in 2017. He is currently an assistant professor with the Faculty of Science and Technology in the Nakhonratchasima Rajabhat University, Thailand. Dr. Ronnachai Chuentawat's research interests lie in the areas of data mining, machine learning and artificial intelligence.

Nittaya Kerdprasop received the PhD in computer science from Nova Southeastern University, U.S.A. in 1999. She is currently an associate professor with the School of Computer Engineering, Suranaree University of Technology, Thailand. Her research interests lie in the areas of data mining, artificial intelligence, knowledge discovery in databases, logic programming and intelligent databases.

Kittisak Kerdprasop received the PhD in computer science from Nova Southeastern University, USA in 1999. He is currently an associate professor and chair of the School of Computer Engineering, Suranaree University of Technology, Thailand. His research interests include data maining, artificial intelligence, functional and logic programming languages and computational statistic.