Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2014, Article ID 310478, 14 pages
http://dx.doi.org/10.1155/2014/310478

Research Article

Variation-Oriented Data Filtering for Improvement in Model Complexity of Air Pollutant Prediction Model

Chi Man Vong,[1] Weng Fai Ip,[2] and Pak Kin Wong[3]

[1] *Department of Computer and Information Science, University of Macau, Macau*
[2] *Supporting Group, Faculty of Science and Technology, University of Macau, Macau*
[3] *Department of Electromechanical Engineering, University of Macau, Macau*

Correspondence should be addressed to Chi Man Vong; cmvong@umac.mo

Received 9 January 2014; Accepted 5 March 2014; Published 9 April 2014

Academic Editor: Qingsong Xu

Copyright © 2014 Chi Man Vong et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Accurate prediction models for air pollutants are crucial for forecast and health alarm to local inhabitants. In recent literature, *discrete wavelet transform* (DWT) was employed to decompose a series of air pollutant levels, followed by modeling using *support vector machine* (SVM). This combination of DWT and SVM was reported to produce a more accurate prediction model for air pollutants by investigating different levels of frequency bands. However, DWT has a significant demand in model complexity, namely, the training time and the model size of the prediction model. In this paper, a new method called *variation-oriented filtering* (VF) is proposed to remove the data with low variation, which can be considered as *noise* to a prediction model. By VF, the noise and the size of the series of air pollutant levels can be reduced simultaneously and hence so are the training time and model size. The SO_2 (sulfur dioxide) level in Macau was selected as a test case. Experimental results show that VF can effectively and efficiently reduce the model complexity with improvement in predictive accuracy.

1. Introduction

Rapid urban development if inappropriately managed may lead to increase in pollution. Many studies [1–3] reported that the amount of sulfur dioxide (SO_2) is associated with adverse human health. Moreover, WHO (World Health Organization) [4] reported that the health problems in turn may increase the burden of the health-care systems in the long run. Therefore, one possibility for reducing pollutant related sick leave is an early warning. Authorities may provide general public with an early warning using a reliable short-term prediction during adverse pollution conditions.

There are two mainstream methods for air pollution prediction: deterministic and statistical based method. Deterministic models [5] are, however, costy to develop (establishment of various inventory) and difficult to operate in real time. Even when adequate data and resources were to become available to implement the deterministic approach, Gardner and Dorling [6] and Kukkonen et al. [7] pointed out that the complexity of a problem in general increases when the spatial interactions between *systems* (regional and urban backgrounds) are ill defined. Since statistical methods usually are derived from empirical relationships between air pollution and other related parameters, these methods are simple to develop and have been widely used in short-term prediction of air pollution [8]. Statistical methods applied to air pollution prediction include *multiple linear regression*, *nonlinear regression*, *autoregressive moving average* (ARMA), and *artificial neural networks* (ANNs). Among these statistical methods, ANNs are regarded as a cost-effective and reliable method for air pollution forecasting. Comrie [9] and Ando et al. [10] showed that ANNs are better suited for air pollution forecasting than linear and nonlinear models. Recent approaches had ANNs combined with other machine learning methods, for example, *genetic algorithm* (GA) [11], in order to improve the predictability of ANN models [12–15]. However, ANNs suffer from some inherent drawbacks [16, 17] such as local minima, overfitting, poor generalization, and long training time. Recently, *support vector machine* (SVM) has attracted considerable attention for prediction

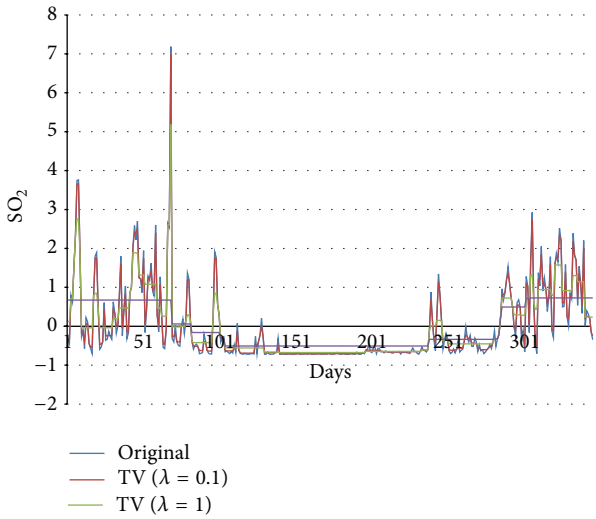


Figure 1: A sample series of SO_2 level in Macau applied after TV denoising; large λ indicates higher filtering rate.

problems [18–20]. This approach is based on the structural risk minimization to provide a good generalization capability [21]. SVM is therefore very resistant to the underfitting and overfitting that are common when ANNs are used. SVM is therefore expected to improve generalization of ANNs and obtain global solutions simultaneously. For example, Lu and Wang [22] reported that SVM model provided better prediction of air pollution than that from ANNs model.

In [23], various data filtering techniques, such as *moving average, exponential smoothing, and total variation* (TV) *denoising*, can be applied to further improve the accuracy of a prediction model by filtering or smoothing the noise in the training dataset. However, these data filtering techniques usually filtered or smoothed acute values which are considered as noise or outliers. In many categories of time-series prediction [24], such as financial data, environmental parameter estimation (current application belongs to this aspect), electric utility load, and machine reliability, the acute values are those of great significance and interest. The filtering of these acute values is likely to incur critical information loss in a warning system in particular and hence deterioration in prediction accuracy as shown in Figure 1. It can be seen that significant proportion of information with high variation become lost after TV denoising process. Therefore, the class of the above-mentioned data filtering techniques that attenuate the acute values is not applicable to current application.

Furthermore, time-frequency decomposition tool such as *discrete wavelet transform* (DWT) has been reported [25] to aid time-series prediction. For example, Osowski and Garanty [26] attempted to integrate DWT with SVM for a higher prediction accuracy for the application of air pollution forecasting. Moreover, many studies [27–30] in different applications have also reported that DWT improves the performance of prediction models by decomposing the time series into various levels of frequency bands for forecasting. However, DWT decomposes a time series into several subseries corresponding to various frequency bands. The frequency bands with relatively low variations are usually considered as insignificant because they carry little or even no information that hardly contribute to forecasting. Hence the subseries corresponding to these frequency bands can be filtered out. For every remaining subseries, a corresponding SVM submodel is constructed. Juhos et al. [31] reported that SVM model led to higher accuracy in their study for forecasting NO and NO_2 after optimizing the SVM hyperparameters through a time-consuming grid search. Nevertheless the time-consuming grid search can be replaced by efficient, direct search methods such as genetic algorithms (GA). There is, however, usually a degeneration of prediction accuracy because GA easily returns local suboptimal hyperparameters.

To achieve the best prediction accuracy, every SVM submodel requires a time-consuming optimization process of hyperparameters such that the training process using DWT and SVM can easily take several hours (e.g., in current application). The current application focuses on *daily* forecasting of SO_2 level while *hourly* forecasting in practice may become necessary for dominant diurnal activities (i.e., every hour, the dataset is updated and the prediction model is retrained) and therefore time is another critical issue. The time issue becomes even more critical when more associated factors (i.e., input variables) to the pollutant level are available. In short, DWT significantly increases the model complexity (i.e., training time and model size) of the prediction model for air pollution forecasting. From this viewpoint, the objective of this study is to design an efficient algorithm that can improve the prediction accuracy while it does not significantly increase the model complexity as in DWT.

To design such efficient algorithm, direct manipulation on the time series is preferred to the time-frequency decomposition as in DWT. A possible solution to this algorithm may involve the reduction of the number of training data by clustering similar data points into various clusters. However, the application of SO_2 level prediction has a stochastic nature, that is, two similar inputs can produce two highly different outputs. Hence, clustering of data points may even cause performance degeneration and thus does not necessarily work in the current application.

In fact, the reduction of training data can be based on the determination of variation from a statistical viewpoint. It can be observed that a series of data points $(\mathbf{x}_i, y_i)$ carries little information if there is low variation among y_i of its successive points, where $\mathbf{x}_i$ is the vector of input variables and y_i indicates the corresponding SO_2 level for $i = 1$ to N. For example, a series of $N = 5$ data points is represented as $S = [(\mathbf{x}_1, 10), (\mathbf{x}_2, 12), (\mathbf{x}_3, 11), (\mathbf{x}_4, 10), (\mathbf{x}_5, 21)]$. The first four points of S only carry little information because y_1 to y_4 are close to each other. Hence, the second and third data points (with 12 and 11, resp.) can be discarded because they hardly contribute any extra information. Finally N can be reduced to 3 while S becomes $[(\mathbf{x}_1, 10), (\mathbf{x}_4, 10), (\mathbf{x}_5, 21)]$. Furthermore, the low variation among these data points may be even considered as noise. For a real example, Figure 2(a) shows the trend of SO_2 levels in Macau from year 2003 to year 2008. It can be seen that high daily variation of SO_2 levels occurred in some days (Figure 2(b)). The input

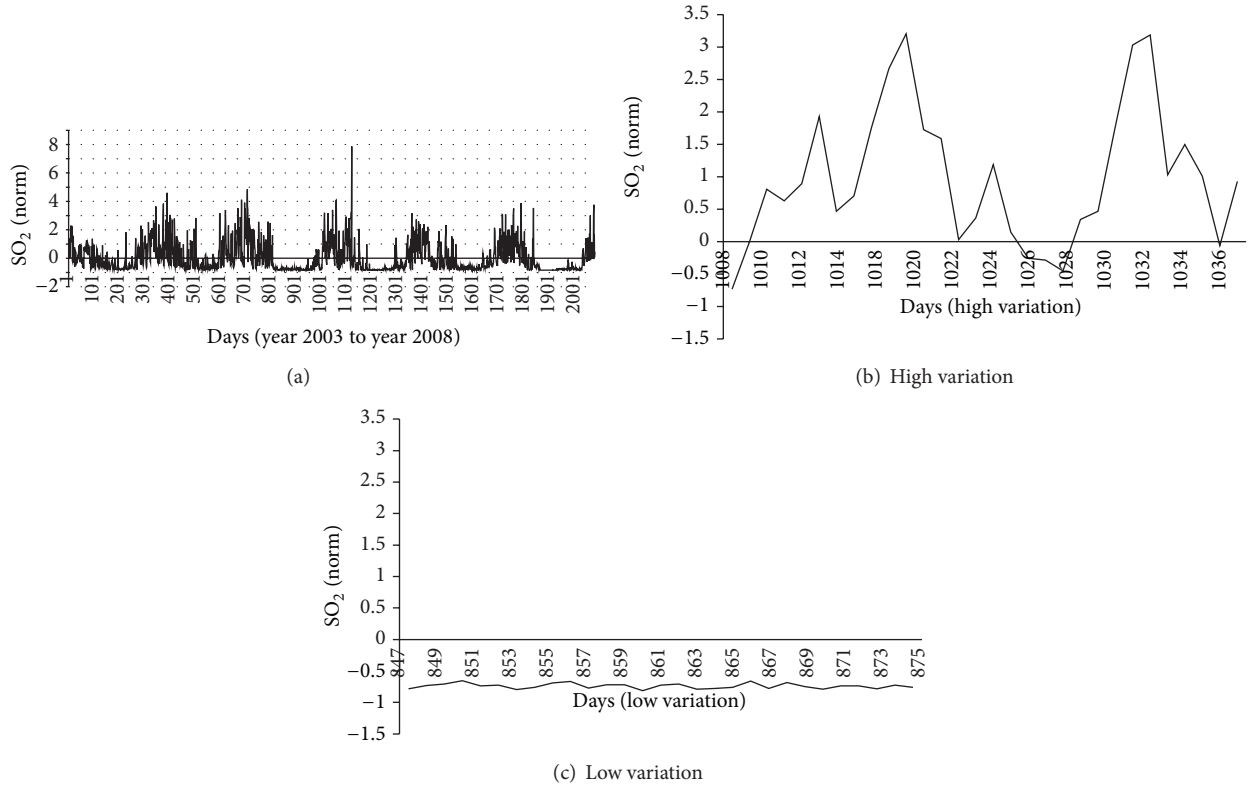


Figure 2: (a) The series of SO_2 levels in Macau from year 2003 to year 2008. (b) A representative case with high variation. (c) A representative case with low variation.

data from these days carry more valuable information for SO_2 level prediction. In Figure 2(c), some successive days are with low daily variation of SO_2 levels. Such data carry little practical information for a warning system of adverse air quality. Therefore, these data can be discarded without significantly affecting the accuracy of SO_2 prediction models under a statistical perspective, where data or dimensions with low variations can be discarded (as mentioned in the theory of *principle components analysis* (PCA) [32]). With a daily variation threshold δ of 0.15 (to be explained in Section 2), about 30% (662) of days from year 2003 to year 2008 are with low daily variation of SO_2 level and can be discarded. Based on this idea, we propose a novel algorithm called *variation-oriented filtering* (VF) which can effectively and efficiently reduce the number of data points of low variation while most of the intrinsic information (i.e., high variation and acute values) of the training data can be retained. Therefore, compared to DWT, VF can reduce the model complexity for SO_2 level prediction without obvious sacrifice of the prediction accuracy.

Moreover, since using VF the number of training data points can be reduced, the number of support vectors #SV (selected from the training data points) of the final SVM regression model [21] as shown in the following is very likely to reduce or at least remained the same:

$$y = \sum_{i=1}^{\#SV} \left(\alpha_i - \alpha_i'\right) K\left(\mathbf{x}_i, \mathbf{x}\right) + b, \quad (1)$$

where α_i and α_i' are the support values corresponding to the ith support vector, K is a kernel function, and b is a bias constant. With less training data points, the training time can be shortened while, with less support vectors, less memory is necessary for storage of the model and also faster execution time. Since DWT and SVM are well-known techniques in the past two decades, interested readers may refer to the classical textbooks [21, 25] for technical details.

In Section 2, the algorithm of VF and the framework of modeling for SO_2 level prediction are described. In Section 3, an illustrative application of SO_2 level prediction is presented, followed by the simulation results and discussion in Section 4. Finally a conclusion (and future work) is drawn in Section 5.

2. Methods

In this section, the description of the proposed algorithm is presented, followed by the framework of constructing SO_2 prediction models.

2.1. Algorithm of Variation-Oriented Filtering (VF). The trend of SO_2 levels consists of a series of data points $(\mathbf{x}_i, y_i)$, where $\mathbf{x}_i$ is the vector of normalized input variables and y_i indicates the corresponding output variable (normalized SO_2 level), for i = 1 to N (number of days). Data normalization is mentioned in Section 3.2. VF interprets a series as different segments of data points. During the interpretation of segments of data

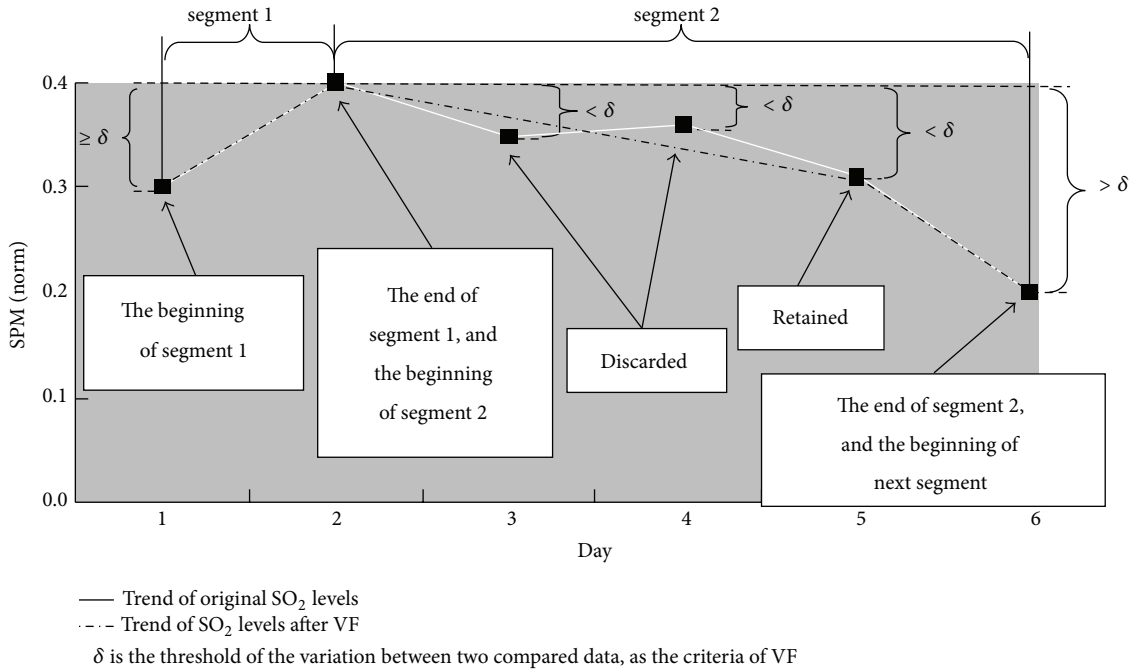


Figure 3: An example of variation-oriented filtering.

points, a data point $(\mathbf{x}_i, y_i)$ is compared with its previous point $(\mathbf{x}_{i-1}, y_{i-1})$. If the difference between y_{i-1} and y_i exceeds or equals a user-defined variation threshold δ, that is,

$$\left|y_i - y_{i-1}\right| \geq \delta, \tag{2}$$

then $(\mathbf{x}_i, y_i)$ is considered to be significant and is assigned the high variation class. Otherwise, $(\mathbf{x}_i, y_i)$ belongs to the low variation class and is discarded. After the interpretation, a reduced series of data points with only the high variation class is produced. The procedure is described in detail as follows.

A data point $(\mathbf{x}_i, y_i)$ is assumed to be the beginning of a segment. It is compared with the next data point $(\mathbf{x}_{i+1}, y_{i+1})$ using (1) under an optimized threshold of daily variation (δ^*). If (2) is false, the data point $(\mathbf{x}_{i+1}, y_{i+1})$ belongs to the low variation class and is discarded. The comparison then continues with $(\mathbf{x}_{i+2}, y_{i+2})$ and so on until either (2) is true with a data point $(\mathbf{x}_j, y_j)$ for $i < j < N$, or the end of the series of data points is reached; that is, $j = N$. In order to retain the trend of a segment, the preceding and current data points, namely, $(\mathbf{x}_{j-1}, y_{j-1})$ and $(\mathbf{x}_j, y_j)$, are both retained instead of $(\mathbf{x}_j, y_j)$ alone. Finally $(\mathbf{x}_j, y_j)$ is marked as the end point of the current segment and simultaneously the beginning of next segment, where the interpretation of next segment begins.

An example of applying VF is shown in Figure 3. Assume the data point $(\mathbf{x}_1, y_1)$ at Day 1 is the beginning of Segment 1. It is compared with its following data point $(\mathbf{x}_2, y_2)$ at Day 2 using (2) under δ. Since (2) is true, $(\mathbf{x}_2, y_2)$ is retained and marked as the end of Segment 1 and the beginning of Segment 2. For Segment 2, the data point $(\mathbf{x}_3, y_3)$ at Day 3 is compared with $(\mathbf{x}_2, y_2)$ and (2) is false. Then, $(\mathbf{x}_3, y_3)$ is discarded. Similarly, the data points at Day 4 and Day 5 are discarded. The comparison with $(\mathbf{x}_2, y_2)$ continues until the data point $(\mathbf{x}_6, y_6)$ at Day 6 is reached where (2) becomes true. In order to maintain the trend of the segment, the data point $(\mathbf{x}_6, y_6)$ and its previous one $(\mathbf{x}_5, y_5)$ are retained. Then $(\mathbf{x}_6, y_6)$ is marked as the end of Segment 2 and also the beginning of Segment 3. The dash dot line in Figure 3 shows the trend of SO_2 levels, which is not significantly deformed, after applying VF. In addition, the data points of Day 3 and Day 4 carry little information and can be considered as noise. The algorithm of VF described above is outlined as in Algorithm 1.

2.2. Workflow of Modeling. The workflow of modeling employs the techniques of VF, SVM, and GA. For accurate SO_2 level prediction, VF is proposed to filter out the noise in the input data (i.e., the data points with low variation). In fact, the variation threshold δ for VF is difficult to define and is subject to different training data. If δ is set too high, some informative data points for prediction will be discarded. On contrary, if δ is set too low, the redundant data points cannot be effectively filtered out. Hence, an optimization for δ is required.

SVM was employed as the modeling technique in this study. Mostly radial basis function (RBF) kernel is selected in SVM for modeling problems. Under this setup, SVM includes two hyperparameters c and g, which are the regularization factor and the RBF kernel parameter, respectively. Therefore, there are three hyperparameters (δ, c, and g) in total to be optimized in the current modeling framework. Although there are numerous optimization techniques in the literature, GA is regarded as one of the most powerful optimization techniques and has been widely used in many problems [12, 13, 33] to optimize the model performance. In this study, GA is employed as the optimization tool and the details of GA setup can be found in Section 3.4. Other optimization techniques may be tried in the future.

```
int i = 1;                    //initialize the beginning of the segment
int j = 2;                    //initialize the end of the segment
select (x_1, y_1);            //select the data at the first day of the first segment
for j = 2 to N do             //loop from 2 to N, i.e., the end of the series
    if i = (j − 1) and (2) is true {
                  //if the ith day and the jth day are successive days and the variation of
                  the SO2 levels in these two days is equal to or greater than δ
        select (x_j, y_j)     //select the data from the jth day
        i = j;                //change the end of the segment to the beginning of the segment
        j++;                  //change the end of the segment to the next day
}
elseif (2) is true {//if the ith day and the jth day are not on successive days but
                    their variation is equal to or greater than δ
    select (x_{j−1}, y_{j−1}) //select data at the (j − 1)th day
    select (x_j, y_j)         //select data at the jth day
    i = j;                    //set the beginning of the next segment
    j++; }                    //change the end of the segment to the next day
else j++; }                   //discard the data point and move to next segment
```

Algorithm 1

The workflow of modeling is depicted in Figure 4. The first objective is to determine the optimal hyperparameters (δ^*, c^*, and g^*) using GA, as shown in Figure 4(a). An initial population of hyperparameters (δ, c, and g) is randomly generated. For every instance of (δ, c, and g) from the population, the steps of VF and SVM are applied as follows. Given a training dataset $\mathbf{D}$, a set of data points of high variation class $\mathbf{D}_{\mathrm{VF}}$ is selected by applying VF with the variation threshold δ. Then SVM with the c and g is used to build a provisional SVM model. If there are 100 instances of (δ, c, and g) in the population, there are 100 provisional SVM models. Every provisional SVM model is then evaluated using a fitness function over an independent validation set *VALID* (more details can be found in Section 3.4). The above procedure repeats until the stopping criterion is satisfied. Finally the optimal hyperparameters (δ^*, c^*, and g^*) are returned. The second step is simply the construction of a prediction model of SO_2 levels from $\mathbf{D}$ using the optimal hyperparameters as shown in Figure 4(b).

3. Application

In this section, the environment and monitoring sites in Macau are briefly introduced. The data preparation and representation in this study are described. In addition, the detail of experimental setup is described.

3.1. The Environment and Monitoring Sites in Macau. Macau, located on the southern coast of China with merely 26.8 square miles land area, comprises three land zones: Macau peninsula, Taipa, and Coloane (Figure 5). Similar to many coastal cities in Mainland China, Macau has experienced rapid urban development over the past decades. Therefore, the air pollution data and meteorological data in Macau were used as a case study.

Macau government meteorological center (DSMG) [34] has established general and roadside meteorological stations for collecting air pollution data, such as suspended particulate matters (SPM), nitrogen dioxide (NO_2), sulfur dioxide (SO_2), and ozone (O_3), and meteorological data, such as temperature, wind direction, wind speed, and relative humidity. Since the land area of Macau is relatively small, the data obtained at the general meteorological station at Taipa Grande (at an elevation of approximately 150 m above sea level) may be considered as representative for the entire region of Macau. Therefore, air pollution data and meteorological data form Taipa Grande general meteorological station were used in this study.

3.2. Data Preparation and Representation. The daily average values for air pollution data and meteorological data from year 2003 to year 2008 were extracted from the website of DSMG. These data were considered representative measures and were adopted in this study of SO_2 modeling. The choice of the study periods was based primarily on the completeness of both air pollution data and meteorological data available at the time of the experiment. In the study periods, the percentage of missing data, possibly due to maintenance or calibration work, is less than 3%. Outside of these periods, either air pollution data or meteorological data are unavailable.

It is necessary to provide commensurate data ranges so that the SO_2 model will not be dominated by variables with large values. Moreover, the normalization procedure usually leads to more stable and accurate prediction results. In this study, the training data sets (*TRAIN*) and test data sets (*TEST*), including input variables $\mathbf{x}$ and output variable y, were normalized based on zero mean and unit variance, as shown in the following equation:

$$a \longleftarrow \frac{a - \bar{a}}{\sigma_a}, \tag{3}$$

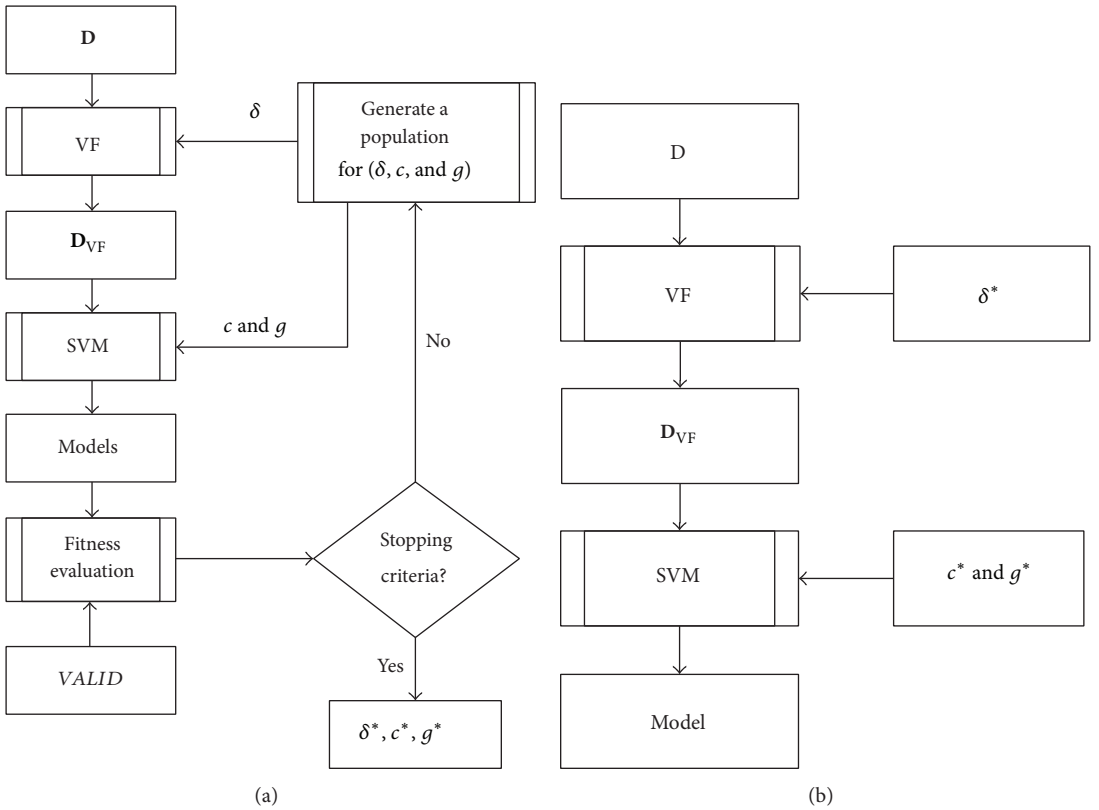


Figure 4: Workflow of modeling in the current study: (a) determination of optimal hyperparameters; (b) SVM model construction.

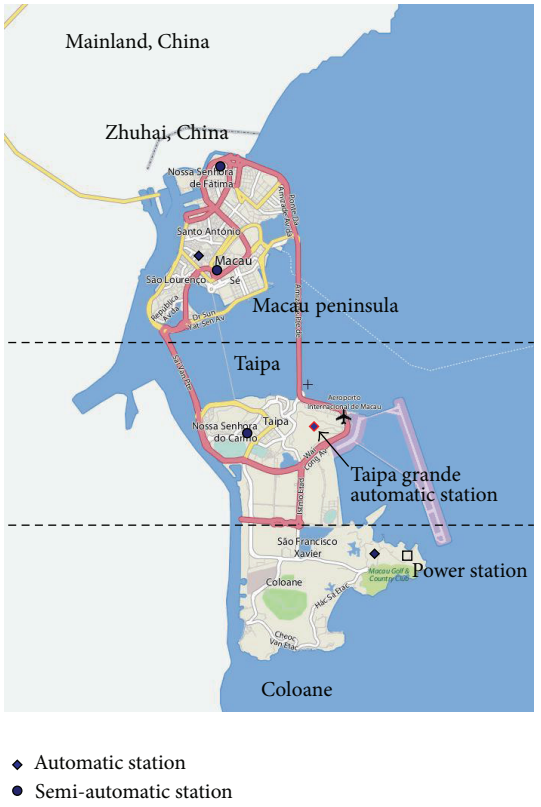


Figure 5: The three land zones of Macau and the location of the meteorological stations and the power station.

Table 1: Pearson correlation coefficients between the output variables $SO_2(d+1)$ and input variables candidates on pervious days in the time period from year 2003 to year 2008.

Relation to $SO_2(d+1)$	Day				
Input variables	d	$d-1$	$d-2$	$d-3$	$d-4$
SPM	0.50	0.40	0.37	0.35	0.36
SO_2	0.72	0.57	0.50	0.45	0.44
NO_2	0.64	0.52	0.48	0.44	0.43
O_3	−0.05	−0.03	−0.03	−0.01	0.00
AP	0.57	0.56	0.54	0.52	0.49
TEMP	−0.60	−0.59	−0.55	−0.52	−0.48
mRH	−0.36	−0.31	−0.27	−0.23	−0.21
WS	0.21	0.27	0.26	0.22	0.20
RF	−0.15	−0.14	−0.14	−0.14	−0.14
SHr	−0.08	−0.08	−0.07	−0.08	−0.07

where a is either an input variable in $\mathbf{x}$ or the output variable y, and $\bar{a}$ and σ_a are the mean and the standard deviation of the variable, respectively.

In order to maintain parsimony for the input variables in the model so that the model does not become overgeneralized, it is necessary to sort out and select the highly related input variables only. The selection of input variables is based on Pearson correlation coefficient [35]. In short, the Pearson correlation coefficient is a measure of the linear dependence between two variables X and Y, giving a value between +1 and −1 inclusive. A value close to +1 indicates positive correlation while a value close to −1 indicates negative correlation. A value close to zero indicates no dependence between the two variables.

Pearson's correlation coefficient between two variables X and Y is commonly denoted by $r_{X,Y}$:

$$r_{X,Y} = \frac{\sum_{i=1}^{N}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sqrt{\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^2 \sum_{i=1}^{N}\left(Y_i - \bar{Y}\right)^2}}. \quad (4)$$

In this study, variables X and Y are time series, containing a series of data point. X_i and Y_i are the ith data points of X and Y, respectively, $i = 1$ to N, where N is the number of data. $\bar{X}$ and $\bar{Y}$ are the means of X and Y, respectively.

In this case study, the candidates of input variables are the air pollutants (SPM, SO_2, NO_2, and O_3) and available meteorological data (atmospheric pressure (AP), temperature (TEMP), mean relative humidity (mRH), wind speed (WS), rainfall (RF), and sunshine hour (SHr)) on previous days (e.g., current day d, previous day $d-1$); the output variable is the SO_2 level on the following day $d+1$. Pearson's correlation coefficients among the output variables $SO_2(d+1)$ and the candidates of input variables on previous days are calculated based on (4) and summarized in Table 1. For example, in order to find Pearson's correlation coefficient ($r_{X,Y} = 0.64$) between X and Y, namely, $NO_2(d)$ and $SO_2(d+1)$, respectively, d is set as 1 January 2003. Then the series of NO_2 levels from 1 January 2003 to 30 December 2008 is retrieved and N is set to the number of days in this period. Each NO_2 level in this period is considered as X_i, $i = 1$ to N. Similarly, the series of SO_2 levels from 2 January 2003 to 31 December 2008 is selected because one day is shifted for $d+1$, and each SO_2 level in the period is Y_i. The corresponding data are put into (3) to calculate $r_{X,Y}$.

The level of significance is assumed to be 0.50; that is, two variables have positive correlation if the corresponding Pearson correlation coefficient is greater than 0.50 and negative correlation if the coefficient is smaller than −0.50; otherwise, the two variables have no relationship. Hence, with the level of significance set to 0.50, the followings can be summarized from Table 1.

(i) $SO_2(d+1)$ seems positively correlated with SPM, SO_2, NO_2, and AP, and negatively correlated with TEMP, on different days (d, $d-1$, and so on). Therefore the variables SPM, SO_2, NO_2, AP, and TEMP on different days were selected as input variables in this study.

(ii) $SO_2(d+1)$ seems poorly correlated with O_3, mean relative humidity (mRH), wind speed (WS), rainfall (RF), and sunshine hours (SHr) on all days. These input variables candidates have weak influence on the prediction of SO_2 level at the following day $d+1$ and thus are not selected.

Furthermore, Macau is characterized as a subtropical monsoon climate. Air-borne industrial pollutants may be carried by north-ward prevailing wind from Mainland China through Macau in winter, while the south-eastern prevailing wind from the sea in summer usually carries pollutants away. Wind direction (WD), a significant factor to SO_2 levels in Macau, is therefore included in the SO_2 prediction model. In the current study, WD is separated into 16 discrete directions {N, NNE, NE, ENE, E, ...}. In order to make WD more applicable for the generalization and to avoid unnecessary input variables, only the relevant wind directions were selected. Out of the 16 wind directions, those from North- and South-ward, namely, {NNW, N, NNE, NE, E, ESE, SE, SSE}, are related to SO_2 levels prediction in Macau and thus selected as input variables as well. To represent these wind directions without incurring any bias, a set of Boolean variables $WD_i \in \{0, 1\}$ was used for $i = 1$ to 8, instead of a number $WD \in \{1, 2, \ldots, 8\}$. In addition, the wind direction from the current day and the

Table 2: Combinations of SVM models.

Model	SVM	Grid Search	GA	VF	DWT
SVM	√	√			
SVM-GA	√		√		
SVM-VF	√		√	√	
SVM-DWT	√	√			√

previous day, namely, $\mathrm{WD}_i(d)$ and $\mathrm{WD}_i(d-1)$, can adequately be represented in the input, rather than using wind direction for more than one day ago.

Finally the data representation in this study is defined as a pair $(\mathbf{x}, y)$, where $\mathbf{x}$ is the vector of input variables and y indicates the corresponding SO_2 level. According to Table 1, the representation of $\mathbf{x}$ is defined as

$$\begin{aligned}\mathbf{x} = \langle & \mathrm{SPM}(d), \mathrm{SO}_2(d-2), \mathrm{SO}_2(d-1), \mathrm{SO}_2(d), \\ & \mathrm{NO}_2(d-1), \mathrm{NO}_2(d), \mathrm{AP}(d-3), \mathrm{AP}(d-2), \\ & \mathrm{AP}(d-1), \mathrm{AP}(d), \mathrm{TEMP}(d-3), \mathrm{TEMP}(d-2), \\ & \mathrm{TEMP}(d-1), \mathrm{TEMP}(d), \mathrm{WD}_i(d-1), \mathrm{WD}_i(d)\rangle \\ & \text{for } i = 1 \text{ to } 8. \end{aligned} \tag{5}$$

Note that the output $y = \mathrm{SO}_2(d+1)$ is defined as the SO_2 level at the following day $d + 1$.

3.3. Construction of Prediction Models. In this study, four SO_2 models using combinations of SVM, VF, GA, and DWT (shown in Table 2) were constructed to evaluate the effectiveness and efficiency of VF against DWT. The hyperparameters c and g for SVM model and SVM-DWT model were optimized using exhaustive grid search, while the optimization of c and g for SVM-GA and SVM-VF was done using GA. The ranges of both c and g are defined as 0.0~9.9 for a demonstrative trial. Practically wider ranges of these hyperparameters are preferred. Since VF employs GA to search for hyperparameters, SVM-VF works much faster than SVM, which uses exhaustive grid search. In order to illustrate fair comparison between the models SVM and SVM-VF, a model SVM-GA is therefore required to assess the efficiency of VF.

Each model was independently trained 10 times to generate more reliable results. For a more comprehensive comparison for model stability, experiment data (from year 2003 to year 2008) were divided into three groups of *TRAIN* and *TEST* (as shown in Table 3). Each group of data was employed to construct four SO_2 models. Therefore, there are totally 12 SO_2 models in this study.

3.4. GA Setup. As mentioned in Section 2.2, the hyperparameters δ, c, and g are optimized using GA. The ranges of δ, c, and g are defined as 0.00~0.99, 0.0~9.9, and 0.0~9.9, respectively. Hence, it is enough to employ 2 digits (0 to 9) for each of the hyperparameters; that is, an individual of GA is a real-code string of 6 digits. An example of an individual with a string of "865407" is shown in Figure 6, which means $\delta = 0.86$, $c = 5.4$, and $g = 0.7$, respectively.

Table 3: Three groups of experiment data.

Group	*TRAIN*	*TEST*
1	Year 2003 to 2005	Year 2006
2	Year 2004 to 2006	Year 2007
3	Year 2005 to 2007	Year 2008

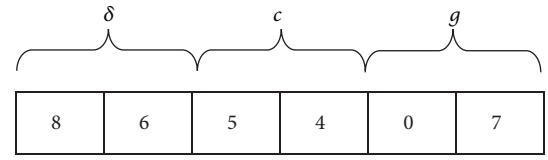


Figure 6: An example of an individual.

The generation of populations of the hyperparameters is controlled by the operations of selection, crossover, and mutation. The selection procedure in this study was divided into two parts. Firstly, in order to imply the elitism strategy [36] which retains the best individuals, two individuals with the best fitness were directly selected from the current population without further processing. Secondly, the other 98 individuals were selected from the current population by using tournament selection and passed to crossover and mutation. The predefined probability of crossover and mutation is 0.8 and 0.05 according to [37].

The evaluation of the fitness of an individual is the accuracy of the SVM model over an independent validation dataset (*VALID*) using the individual (i.e., the hyperparameters). The accuracy of the SVM model is evaluated using *complementary* Willmott's index of agreement (CWIA), which is further discussed in Section 4.1. *VALID* consists of 36 data points randomly selected from *TEST*. The stopping criterion for GA process is that no better fitness can be achieved for 50 successive iterations. Finally, the individual with the best fitness is returned as the optimal hyperparameters (δ^*, c^*, and g^*).

3.5. DWT Setup. As a comparison with VF, DWT was also applied to the training dataset. The family of Daubechies (db) wavelets is one of the most popular mother wavelets and was employed in this study. Ten kinds of Daubechies wavelets (from db1 to db10) were applied to construct different SVM models as illustrated in Figure 7. Given a training dataset $TRAIN = \{(\mathbf{x}_i, y_i)\}$ for $i = 1$ to N, DWT decomposes *TRAIN* into several levels $(TRAIN_1, TRAIN_2, \ldots, TRAIN_{J+1})$. For $k = 1$ to $J + 1$, every $TRAIN_k = \{(\mathbf{x}_i, y_{ik})\}$ is used to construct a submodel SVM_k, respectively, where y_{ik} is the kth decomposed value of y_i. In other words, the series of SO_2 levels, $i = 1$ to N, is decomposed into $J+1$ series $\{y_{ik}\}$ for $k = 1$ to $J + 1$. The final SO_2 prediction is simply the summation of the prediction by all submodels SVM_k.

The SVM model accuracy using different kinds of Daubechies wavelets can be evaluated through CWIA (discussed in Section 4.1) over an independent validation set.

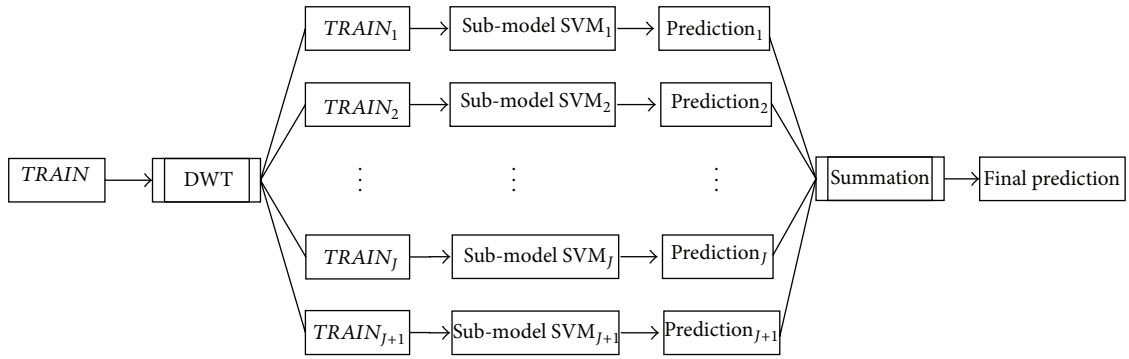


Figure 7: Representation of SVM-DWT model.

From simulation results (not shown here), db7 was found to be the most suitable for this study. In addition, it is also important to decide the decomposition level J. The number of the decomposition levels of each model can be estimated by using the function "wmaxlev" provided in the MATLAB wavelet toolbox. In this study, J was found to be 6.

4. Performance Evaluation and Results

4.1. Error Measures. In order to effectively measure the accuracy of SVM models, three measures for statistic error are used in this study: mean absolute error (MAE) for value difference, root mean squared error (RMSE) for sensitivity, and complementary Willmott's index of agreement (CWIA) for curve fitting. Both MAE and RMSE are common measures to estimate the average error of models. However, neither of them provides information about the relative size of the average difference or the nature of the differences. Willmott [38] reported that CWIA is a standardized measure that can be easily interpreted and provides cross-comparisons of its magnitudes for a variety of models, regardless of units. Owing to its dimensionless nature, relationships described by CWIA tend to complement the information contained in RMSE. Therefore, CWIA was also used to evaluate the accuracy of SVM models in this study. The range of CWIA is from 0 to 1, where a value closer to 0 indicates a better performance. The three measures are described by the following formulas, where P_i and O_i, respectively, represent the predicted and observed value of $SO_2(d+1)$ in the ith day and N is the size of *TEST*:

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left|P_i - O_i\right|, \tag{6}$$

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(P_i - O_i\right)^2}, \tag{7}$$

$$\text{CWIA} = \frac{\sum_{i=1}^{N}\left(P_i - O_i\right)^2}{\sum_{i=1}^{N}\left(\left|P_i'\right| + \left|O_i'\right|\right)^2}, \tag{8}$$

where $\bar{O} = (1/N)\sum_{i=1}^{N} O_i$, $P_i' = P_i - \bar{O}$, and $O_i' = O_i - \bar{O}$.

4.2. Stability and Errors of Models. In addition to the statistic errors MAE, RMSE, and CWIA, model stability is another important criterion to evaluate the performance of models. Therefore, standard deviation (σ) as shown in (8) is employed to evaluate the stability of models in this study, where M is the number of runs, ε_i represents the statistic error for a model in the ith run, and $\bar{\varepsilon}$ represents the mean of all runs:

$$\sigma = \sqrt{\frac{1}{M}\sum_{i=1}^{M}\left(\varepsilon_i - \bar{\varepsilon}\right)^2}. \tag{9}$$

Based on (6) to (9), the statistic errors and standard deviations of four combinations of SVM models can be obtained. The experimental results in Figure 8 show that the standard deviation σ of each statistic error for all models is very small (or even close to zero). It indicates that the results produced in each run for all models were very close to their corresponding means. Therefore, the performance of these models is stable and hence the statistic errors of these models are reliable.

In Figure 8, the statistic errors MAE, RMSE, and CWIA of the four models SVM, SVM-GA, SVM-VF, and SVM-DWT are shown. Each model was trained with three different datasets according to the grouping in Table 3 so that 12 different models were obtained. Note that the statistic errors in Figure 8 are the mean of the errors predicted by each model in 10 independent runs. From the results, SVM-VF model (Figure 8) produced the lowest statistic errors in three groups of experimental data among the four models. Despite being generated from fewer training data, VF can improve the accuracy of the prediction by filtering unimportant low variation data points. Compared to SVM and SVM-GA models, SVM-VF model has a relative improvement of about 5% and 9% in all statistic errors, respectively. Conversely, SVM-DWT model produced significantly worse accuracy than the other models. It seems that DWT does not show its superiority for improving the accuracy of the SVM in this study. The decline in accuracy may arise from the fact that DWT decomposes a time series into different subseries, each of which has an independent prediction model as shown in Figure 7. Each of these models incurs an error and hence the accumulated error for the final SVM-DWT model becomes relatively large. Furthermore, SVM-DWT model took the longest training time (almost 5 hours) among the three

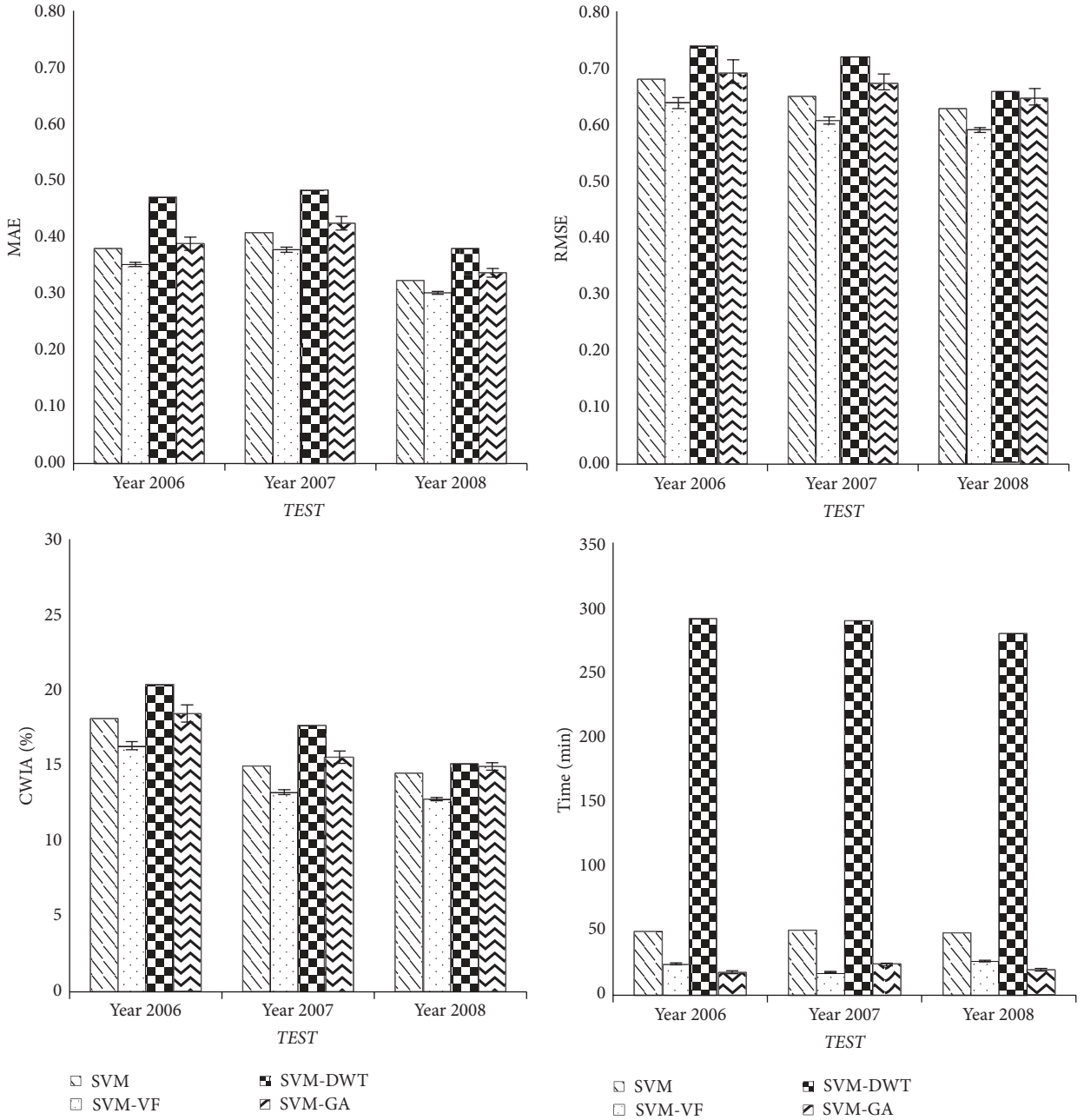


Figure 8: Statistic errors and standard deviations of four combinations of SVM models. *TRAIN* are 2003–2005, 2004–2006, and 2005–2008, respectively.

compared models, while the training time of SVM-VF model was the shortest (<0.5 hour). The time issue will be further discussed in Section 4.4.

4.3. Seasonal Variation of SO_2 Levels. In order to visualize the performance of the studied methods, the trends of the predicted SO_2 levels were compared with the observed SO_2 levels from 2006 to 2008. Figure 9 shows that the trends of predicted SO_2 levels of each model are generally close to the trend of observed levels. It suggests that these four models have a good capability of prediction.

In addition, representative cases of predicted and observed SO_2 levels can clearly show a better comparison of the prediction capability of the four models. According to the climatic characteristics of Macau (mentioned in Section 3.2), the trends of SO_2 levels in winter and summer can represent the general yearly trend in Macau. Hence, the predicted and observed SO_2 levels in winter and summer for each experiment were shown in Figures 10 and 11, respectively. These two figures illustrated that the predicted SO_2 level produced by SVM-VF model is closer to the observed SO_2 level when comparing with those of the other three models.

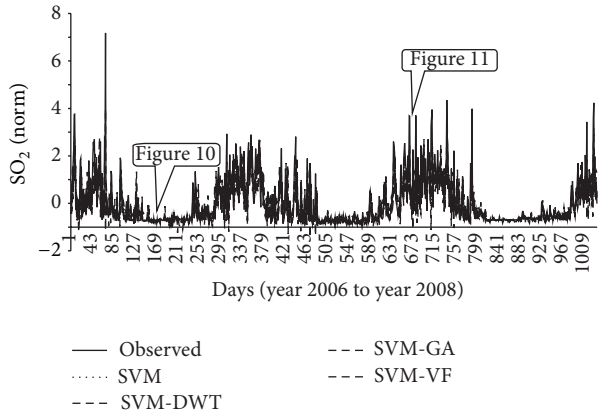


Figure 9: Representative predicted and observed SO_2 levels using the proposed four combinations of SVM models from 2006 to 2008.

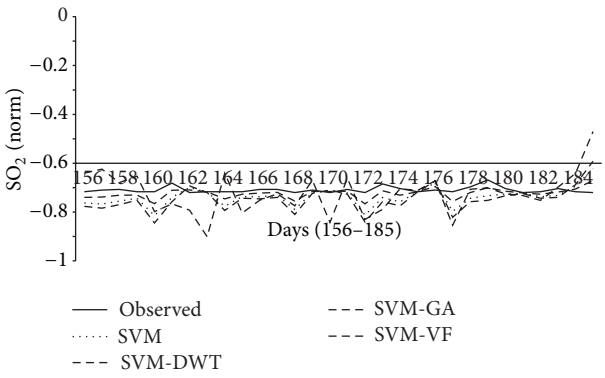


Figure 10: Representative predicted and observed SO_2 levels using the proposed four combinations of SVM models from Day 156 to Day 185 (29 days).

In particular, SVM-VF appears to fit the more prominent high SO_2 level during winter seasons (Figure 11), when health warning is most needed in this case study.

4.4. Model Complexity. The model complexities of the three combinations of SVM models are shown in Table 4, where SVM-DWT model is observed to have relatively large number of support vectors (SV) (about 6800) and very long training time (about 290 minutes). Therefore, the cost of DWT decomposition for SO_2 level forecasting is at the relatively high model complexity (i.e., training time and model size) in the current application of air pollution prediction.

For SVM-VF models, the size of the training dataset *TRAIN* was reduced to different extents after applying VF for data filtering. The highest data reduction rate by VF is up to 13.33% (for *TRAIN* in year 2003 to 2005). In addition, the number of SV for SVM-VF is the fewest among the three models, especially when compared with SVM-DWT. Furthermore, the execution time of SVM-VF model is much shorter than those of SVM model (about 50 minutes) and SVM-DWT model (about 290 minutes) (Table 4).

Although SVM-GA model requires even shorter training time than SVM-VF model, it suffers from performance

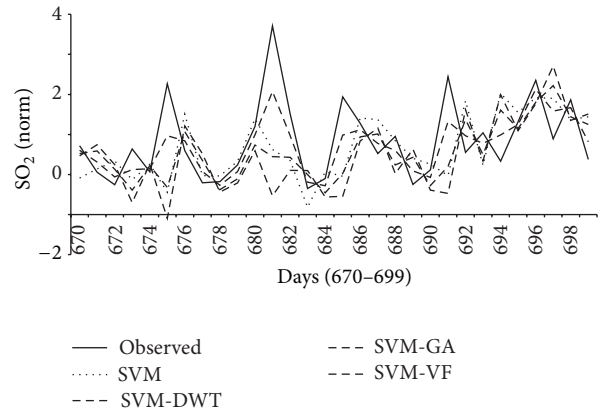


Figure 11: Representative predicted and observed SO_2 levels using the proposed four combinations of SVM models from Day 670 to Day 699 (29 days).

degeneration as shown in Figure 8 because GA can easily locate the suboptimal SVM hyperparameters. Hence, the proposed algorithm VF may resolve the high complexities of model construction in both time and model size while it can even improve the prediction accuracy. In summary, SVM-VF model not only effectively reduces the model complexity of prediction model for SO_2 level, but also improves the accuracy of the prediction models by filtering the unimportant low variation data points.

4.5. Discussion. From the experimental results, VF was verified to effectively reduce the size of training dataset *TRAIN* and also the number of SV of the prediction models. The reductions of training data points and number of SV lead to shorter training time and smaller model size that can benefit for real large-scale modeling using hundreds of input variables and millions of data points. Despite using fewer training data, SVM-VF model can produce higher accuracy than SVM model with significantly shorter training time. It is because VF only discards the unimportant data points with low variation and hence does not affect the accuracy of the SO_2 prediction model. Moreover, the unimportant low variation in the discarded data points may be a noise to the modeling. By filtering out these noises, the accuracy of SO_2 prediction model can be improved as shown in the results. In addition, the training time of SVM-VF model is significantly reduced as compared to SVM.

Using the accuracy of SVM as a standard, when comparing with SVM-GA model, SVM-VF model takes longer training time but produces relatively up to 9% better prediction accuracy. This is also credited to the reduction of modeling noise. Since the optimization of SVM hyperparameters by GA is based on the SVM prediction accuracy, the hyperparameters become more "suboptimal" under a noisy training dataset (i.e., with unimportant low variation data) so that the prediction accuracy of SVM-GA model is further deteriorated.

Although many literatures reported that DWT decomposition can improve the performance of prediction models for many applications of signal or image processing, DWT does

Table 4: Model complexities of the four combinations of SVM models.

	SVM	SVM-DWT	SVM-VF	SVM-GA
	TRAIN: year 2003 to 2005, *TEST*: year 2006			
Size of *TRAIN*	1055	1055	914	1055
Reduction ratio of *TRAIN*	—	—	13.33%	—
Numbers of SV	1012	6882	881	1012
Training time (mins)	49	291	24	18
	TRAIN: year 2004 to 2006, *TEST*: year 2007			
Size of *TRAIN*	1056	1056	941	1056
Reduction ratio of *TRAIN*	—	—	10.88%	—
Numbers of SV	991	6759	883	991
Training time (mins)	50	290	24	18
	TRAIN: year 2005 to 2007, *TEST*: year 2008			
Size of *TRAIN*	1043	1043	955	1043
Reduction ratio of *TRAIN*	—	—	8.46%	—
Numbers of SV	985	6747	912	987
Training time (mins)	48	280	27	20

Table 5: A comparison among the models in current study.

Model	Accuracy	Training time	Size
SVM	Good	Fair	Good
SVM-VF	**Very good**	Good	**Very good**
SVM-GA	Good	**Very good**	Good
SVM-DWT	Fair	Very poor	Very poor

not necessarily work in time-series prediction. At least, DWT did not establish its superiority in this study. The cause may arise from DWT decomposition and accumulated error as explained in the following. DWT decomposes a time series into several subseries as shown in Figure 7. Each subseries is employed to train a submodel SVM_k, for $k = 1$ to $J + 1$. Every SVM_k produces a prediction of the output for the subseries, which may incur a certain (small but measurable) amount of error e_k. Since the final prediction is simply the accumulation of the predictions of all SVM_k, the total error of the final prediction is also the accumulation of e_k which may become large. Moreover, DWT significantly increases the training time and model size because $J + 1$ submodels, SVM_k, are produced. Hence, SVM-DWT model did not outperform other models in this study. As a comparison to SVM-DWT model, SVM-VF model not only resolves the issue of high model complexity, but also produces better performance. In a nutshell, Table 5 shows the summary and a quick comparison among the models in the current application. It is concluded that our proposed method VF produces better performance, in particular the training time, in current application.

5. Conclusion

The current application of air pollutant prediction aims to predict the significant acute pollutant levels as a warning message. However, most of the existing methods of data filtering or smoothing techniques are likely to attenuate the acute pollutant levels in such a time series (Figure 1) and hence a deterioration in prediction accuracy for time-series prediction. Therefore, these existing methods are not necessarily applicable to the current application. In the literatures about air pollutant level prediction, DWT was used to decompose the series of pollutant levels into different subseries for modeling in order to obtain higher prediction accuracy. However, DWT incurs the issue of high complexity in training time and model size in addition to possible performance degeneration.

This research proposes a new algorithm of variation-oriented filtering (VF) to filter only the unimportant low variation data while the acute ones (i.e., with high variation) are retained. VF can resolve the issue of high model complexity without sacrifice (or even with improvement) of the prediction accuracy. The SO_2 level in Macau was used as a case study; and four different combinations of SVM models and 12 scenarios using GA, DWT, and VF were compared. Experimental results reveal that the models using VF outperform the other models using GA and DWT for SO_2 level prediction in Macau. Moreover, with VF, the number of training data and model construction time can be significantly reduced so that *hourly* (current application is *daily*) prediction of pollutant level can become more feasible. The proposed method VF can also be applied to the prediction of other air pollutants or even other time-series prediction (such as financial data, environmental parameter estimation, electric utility load, machine reliability, etc.) where acute values are of interest and significance, for filtering data of low variations and performance improvement.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The research is supported by the University of Macau Research Grant, Grant no. MYRG141 (Y1-L2)-FST11-IWF.

References

[1] C. A. Pope III, D. V. Bates, and M. E. Raizenne, "Health effects of particulate air pollution: time for reassessment?" *Environmental Health Perspectives*, vol. 103, no. 5, pp. 472–480, 1995.

[2] B. Brunekreef, D. W. Dockery, and M. Krzyzanowski, "Epidemiologic studies on short-term effects of low levels of major ambient air pollution components," *Environmental Health Perspectives*, vol. 103, supplement 2, pp. 3–13, 1995.

[3] M. Almeida-Silva, H. T. Wolterbeek, and S. M. Almeida, "Elderly exposure to indoor air pollutants," *Atmospheric Environment*, vol. 85, pp. 54–63, 2014.

[4] WHO, "Air quality guidelines: global update 2005: particulate matter, ozone, nitrogen dioxide and sulfur dioxide," Tech. Rep., World Health Organization, 2006.

[5] J. N. Carras, M. Cope, W. Lilley, and D. J. Williams, "Measurement and modelling of pollutant emissions from Hong Kong," *Environmental Modelling and Software*, vol. 17, no. 1, pp. 87–94, 2002.

[6] M. W. Gardner and S. R. Dorling, "Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences," *Atmospheric Environment*, vol. 32, no. 14-15, pp. 2627–2636, 1998.

[7] J. Kukkonen, L. Partanen, A. Karppinen et al., "Extensive evaluation of neural network models for the prediction of NO_2 and PM_{10} concentrations, compared with a deterministic modelling system and measurements in central Helsinki," *Atmospheric Environment*, vol. 37, no. 32, pp. 4539–4550, 2003.

[8] P. Zanetti, *Air Pollution Modeling. Teories, Computational Methods and Available Software*, Computational Mechanics Publications, New York, NY, USA, 1990.

[9] A. C. Comrie, "Comparing neural networks and regression models for ozone forecasting," *Journal of the Air & Waste Management Association*, vol. 47, no. 6, pp. 653–663, 1997.

[10] B. Ando, G. Cammarata, A. Fichera, S. Graziani, and N. Pitrone, "A procedure for the optimization of air quality monitoring networks," *IEEE Transaction on Systerms, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 29, no. 1, pp. 157–163, 1999.

[11] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley, 1989.

[12] Z. Hao, C. Kefa, and M. Jianbo, "Combining neural network and genetic algorithms to optimize low NO_x pulverized coal combustion," *Fuel*, vol. 80, no. 15, pp. 2163–2169, 2001.

[13] U. Kesgin, "Genetic algorithm and artificial neural network for engine optimisation of efficiency and NO_x emission," *Fuel*, vol. 83, no. 7-8, pp. 885–895, 2004.

[14] G. Grivas and A. Chaloulakou, "Artificial neural network models for prediction of PM_{10} hourly concentrations, in the Greater Area of Athens, Greece," *Atmospheric Environment*, vol. 40, no. 7, pp. 1216–1229, 2006.

[15] C. Zanchettin, T. B. Ludermir, and L. M. Almeida, "Hybrid training method for MLP: optimization of architecture and training," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 41, no. 4, pp. 1097–1109, 2011.

[16] C.-M. Vong, P.-K. Wong, and Y.-P. Li, "Prediction of automotive engine power and torque using least squares support vector machines and Bayesian inference," *Engineering Applications of Artificial Intelligence*, vol. 19, no. 3, pp. 277–287, 2006.

[17] P. K. Wong, L. M. Tam, K. Li, and C. M. Vong, "Engine idle-speed system modelling and control optimization using artificial intelligence," *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, vol. 224, no. 1, pp. 55–72, 2010.

[18] J. Y. Yang, W. F. Ip, C. M. Vong, and P. K. Wong, "Effect of choice of kernel in support vector machines on ambient air pollution forecasting," in *Proceedings of the International Conference on System Science and Engineering (ICSSE '11)*, pp. 552–557, June 2011.

[19] W.-Y. Lin, Y.-H. Hu, and C.-F. Tsai, "Machine learning in financial crisis prediction: a survey," *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, vol. 42, no. 4, pp. 421–436, 2012.

[20] Y. Lv and Z. Gan, "Robust ε-support vector regression," *Mathematical Problems in Engineering*, vol. 2014, Article ID 373571, 5 pages, 2014.

[21] V. Vapnik, *Satistical Learning Theory*, John Wiley & Sons, New York, NY, USA, 1998.

[22] W.-Z. Lu and W.-J. Wang, "Potential assessment of the "support vector machine" method in forecasting ambient air pollutant trends," *Chemosphere*, vol. 59, no. 5, pp. 693–701, 2005.

[23] W. Wettayaprasit, N. Laosen, and S. Chevakidagarn, "Data filtering technique for neural networks forecasting," in *Proceedings of the 7th WSEAS International Conference on Simulation, Modelling and Optimization*, Beijing, China, 2007.

[24] N. Sapankevych and R. Sankar, "Time series prediction using support vector machines: a survey," *IEEE Computational Intelligence Magazine*, vol. 4, no. 2, pp. 24–38, 2009.

[25] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674–693, 1989.

[26] S. Osowski and K. Garanty, "Forecasting of the daily meteorological pollution using wavelets and support vector machine," *Engineering Applications of Artificial Intelligence*, vol. 20, no. 6, pp. 745–755, 2007.

[27] S.-T. Li and L.-Y. Shue, "Data mining to aid policy making in air pollution management," *Expert Systems with Applications*, vol. 27, no. 3, pp. 331–340, 2004.

[28] L. He, G.-H. Huang, G.-M. Zeng, and H.-W. Lu, "Wavelet-based multiresolution analysis for data cleaning and its application to water quality management systems," *Expert Systems with Applications*, vol. 35, no. 3, pp. 1301–1310, 2008.

[29] K. Siwek and S. Osowski, "Improving the accuracy of prediction of PM_{10} pollution by the wavelet transformation and an ensemble of neural predictors," *Engineering Applications of Artificial Intelligence*, vol. 25, no. 6, pp. 1246–1258, 2011.

[30] F. Murtagh, M. Spagat, and J. A. Restrepo, "Ultrametric Wavelet regression of multivariate time series: application to Colombian conflict analysis," *IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans*, vol. 41, no. 2, pp. 254–263, 2011.

[31] I. Juhos, L. Makra, and B. Tóth, "Forecasting of traffic origin NO and NO_2 concentrations by Support Vector Machines and neural networks using Principal Component Analysis," *Simulation Modelling Practice and Theory*, vol. 16, no. 9, pp. 1488–1502, 2008.

[32] I. T. Jolliffe, *Principal Components Analysis*, Wiley Online Library, 2002.

[33] P.-F. Pai and W.-C. Hong, "Forecasting regional electricity load based on recurrent support vector machines with genetic algorithms," *Electric Power Systems Research*, vol. 74, no. 3, pp. 417–425, 2005.

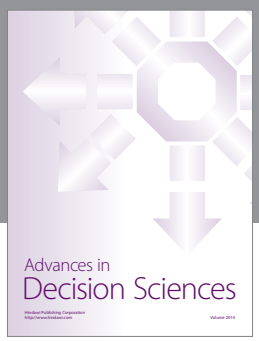
[34] DSMG, http://www.smg.gov.mo/www/c_index.php.

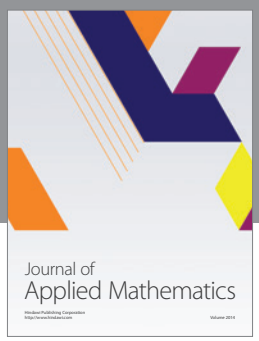
[35] S. M. Stigler, "Francis Galton's account of the invention of correlation," *Statistical Science*, vol. 4, no. 2, pp. 73–79, 1989.

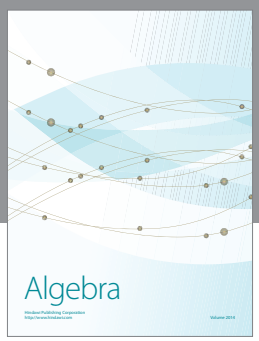
[36] M. Srinivas and L. M. Patnaik, "Genetic algorithms: a survey," *Computer*, vol. 27, no. 6, pp. 17–26, 1994.

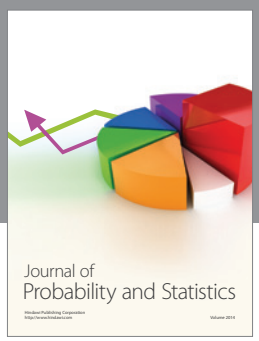
[37] M. Aleixandre, I. Sayago, M. C. Horrillo et al., "Analysis of neural networks and analysis of feature selection with genetic algorithm to discriminate among pollutant gas," *Sensors and Actuators, B: Chemical*, vol. 103, no. 1-2, pp. 122–128, 2004.

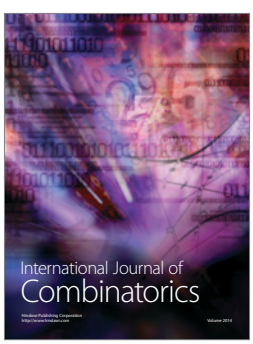
[38] C. J. Willmott, "On the validation of models," *Physical Geography*, vol. 2, no. 2, pp. 184–194, 1981.

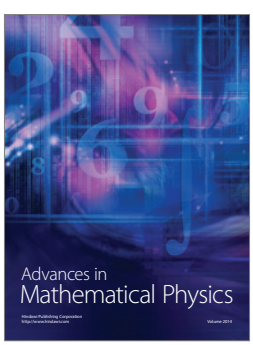


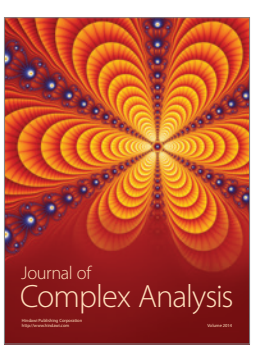


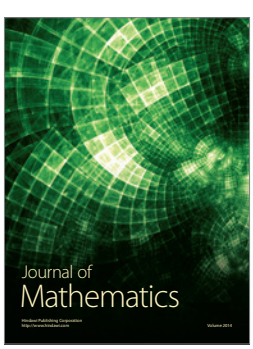


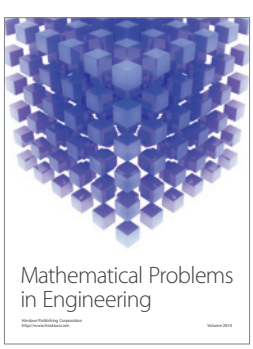


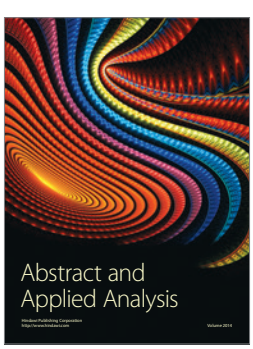


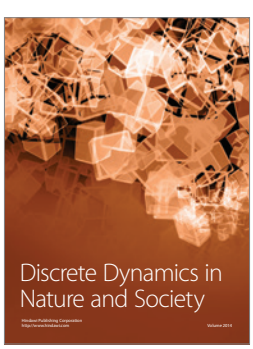


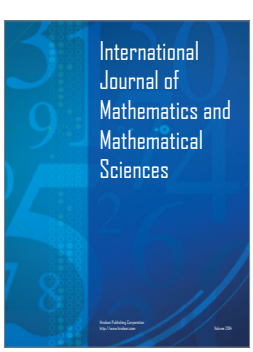












Hindawi

Submit your manuscripts at
http://www.hindawi.com