Science of the Total Environment 651 (2019) 230–240

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

Multi-output support vector machine for regional multi-step-ahead $PM_{2.5}$ forecasting

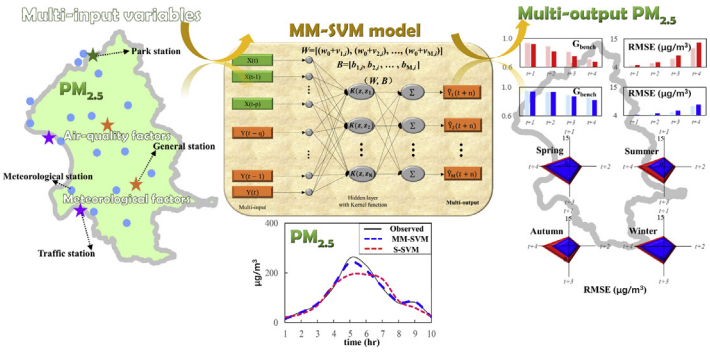
Yanlai Zhou [a], Fi-John Chang [a,*], Li-Chiu Chang [b], I-Feng Kao [a], Yi-Shin Wang [a], Che-Chia Kang [a]

[a] Department of Bioenvironmental Systems Engineering, National Taiwan University, Taipei 10617, Taiwan, ROC
[b] Department of Water Resources and Environmental Engineering, Tamkang University, New Taipei City 25137, Taiwan, ROC

HIGHLIGHTS

- A novel MM-SVM hybriding M-SVM and multi-task learning for regional $PM_{2.5}$ forecasts.
- Multi-task learning algorithm effectively optimizes parameters of the M-SVM model.
- MM-SVM model overcomes the instabilities of spatiotemporal forecasting.
- MM-SVM identifies heterogeneities in air pollutant-generating mechanisms and seasons.
- MM-SVM increases multi-step-ahead $PM_{2.5}$ forecasts accuracy and applicability.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 29 May 2018
Received in revised form 21 August 2018
Accepted 8 September 2018
Available online 13 September 2018

Editor: P. Kassomenos

Keywords:
Multi-output SVM
Multi-task learning algorithm
Multi-step-ahead forecast
$PM_{2.5}$ concentrations
Taipei City

ABSTRACT

Air quality deteriorates fast under urbanization in recent decades. Reliable and precise regional multi-step-ahead $PM_{2.5}$ forecasts are crucial and beneficial for mitigating health risks. This work explores a novel framework (MM-SVM) that combines the Multi-output Support Vector Machine (M-SVM) and the Multi-Task Learning (MTL) algorithm for effectively increasing the accuracy of regional multi-step-ahead forecasts through tackling error accumulation and propagation that is commonly encountered in regional forecasting. The Single-output SVM (S-SVM) is implemented as a benchmark. Taipei City of Taiwan is our study area, where three types of air quality monitoring stations are selected to represent areas imposed with high traffic influences, high human activities and commercial trading influences, and less human interventions close to nature situation, respectively. We consider forecasts of $PM_{2.5}$ concentrations as a function of meteorological and air quality factors based on long-term (2010–2016) observational datasets. Firstly, the Kendall tau coefficient is conducted to extract key spatiotemporal factors from regional meteorological and air quality inputs. Secondly, the M-SVM model is trained by the MTL to capture non-linear relationships and share correlation information across related tasks. Lastly, the MM-SVM model is validated using hourly time series of $PM_{2.5}$ concentrations as well as meteorological and air quality datasets. Regarding the applicability of regional multi-step-ahead forecasts, the results demonstrate that the MM-SVM model is much more promising than the S-SVM model because only one forecast model (MM-SVM) is required, instead of constructing a site-specific S-SVM model for each station. Moreover, the forecasts of the MM-SVM are found better consistent with observations than those of any single S-SVM in both training and testing stages. Consequently, the results clearly demonstrate that the MM-SVM model could be recommended as a novel integrative technique for improving the spatiotemporal stability and accuracy of regional multi-step-ahead $PM_{2.5}$ forecasts.

* Corresponding author.
E-mail address: changfj@ntu.edu.tw (F.-J. Chang).

https://doi.org/10.1016/j.scitotenv.2018.09.111
0048-9697/

1. Introduction

Air pollution is a serious environmental issue attracting more and more attention globally (Apte et al., 2015; Di et al., 2016; Li and Zhu, 2018). A great number of developing countries undergo heavy air pollution (Coelho et al., 2014; Liao et al., 2017; Lv et al., 2016). For instance, extreme air pollution events take place frequently in Taipei City, Taichung City and Kaohsiung City of Taiwan (Li et al., 2017). Fine airborne particles (e.g. $PM_{2.5}$) can penetrate into throats and even into lungs through deep breathing. Long-term exposure to these fine particles would increase the incidence of air pollution-related diseases (e.g., respiratory and cardiovascular diseases, lung function reduction, and heart attacks) and cause serious effects on human health (Xu et al., 2016; Yu and Stuart, 2017). Real-time air quality information is of great importance to air pollution control and human health protection from air pollution (Huang et al., 2014). Therefore, it is imperative to predict air quality to better govern the trend of air pollution variation so as to provide prompt and complete environmental quality information for assisting in environmental management decisions as well as avoiding serious accidents in relation to air pollution (Wang et al., 2016; Yang et al., 2018; Tong et al., 2018).

Environmental prediction can be very beneficial to the protection of human health and welfare from pollutions. Many countries develop their own real-time nationwide Air Quality Forecast (AQF) Systems (e.g. http://www.nws.noaa.gov/aq) using various techniques. These techniques vary greatly in levels of sophistication. In general, there are two basic types of AQF models: deterministic and empirical models. Deterministic models commonly apply fundamental principles of atmospheric chemistry and physics involved in emission and transformation processes of air pollution to simulating and/or forecasting air quality. However, relevant studies indicate that deterministic models are less accurate than well-developed site-specific empirical air quality forecast models due to the highly complex and dynamic pollution processes of air quality and the uncertainties within models as well as pollution emission estimation (Cobourn, 2010; Lv et al., 2016). Empirical models use various statistical and/or machine learning techniques to quantify the underlying complex relationships between air pollutants and potential predictors based on large numbers of data sets under various atmospheric conditions (Cobourn, 2010; Hrust et al., 2009). Among these models, artificial neural networks (ANNs), a crucial branch of Artificial Intelligence (AI), have been frequently utilized to make $PM_{2.5}$ forecasts (e.g., Oprea et al., 2016; Nieto et al., 2018; Zhu et al., 2018). For instance, the back propagation neural networks (BPNN), radial basis function (RBF), Elman recurrent neural network, non-linear autoregressive with exogenous inputs neural network (NARX), adaptive-network-based fuzzy inference system (ANFIS) and support vector regression (SVM) models have been widely applied to modelling air quality (Voukantsis et al., 2011; Ping et al., 2015; Yamane et al., 2015; Bai et al., 2016; Prasad et al., 2016; Gong and Ordieres, 2016; Yeganeh et al., 2018; Zhai and Chen, 2018). Nevertheless, these models have a common drawback, i.e., prone to systematically underpredicting particulate matter concentrations during days with very high particulate matter concentrations. As known, these are the very events that impose the most adverse effects on human health. In order to capture the abrupt changes in particulate matter concentrations through statistical approaches, some prior knowledge and more sophisticated modelling techniques are needed. We notice that all the above-mentioned methods usually construct site-specific data-driven models for each air quality monitoring station individually and disregard the potential nonlinear spatial correlation among different air quality monitoring stations. Bearing this in mind as a motivation, the multi-task learning (MTL) algorithm has been designed to share correlation information across tasks, and each task might benefit from the others (Baxter, 1997). When relations exist between tasks, the MTL algorithm can learn all tasks simultaneously, instead of learning each task independently as traditional approaches do. The MTL algorithm has been widely applied with satisfactory results to multi-output forecast issues (Zhang et al., 2014; Liu et al., 2015; Zhu et al., 2016; Xu et al., 2018).

The multi-output data-driven model adapted in forecasting is generally the instance that the underlying nonlinear correlation among different output variables could be identified and contribute to forecast accuracy, such as spatiotemporal ANNs (Nguyen et al., 2012) and deep learning architectures for air quality predictions (Li et al., 2016). These multi-output data-driven models primarily focused on the mid-long term forecasting of $PM_{2.5}$ concentration but did not involve real-time multi-step-ahead $PM_{2.5}$ concentration forecasting. The demand for multi-step-ahead $PM_{2.5}$ forecasts has increased the implementation difficulty of single-output data-driven models, especially for providing a good representative of regional air quality features. Hence, it is essential to conduct in-depth research with multi-input and multi-output data-driven models to conquer the complexity and challenges encountered in regional air quality forecasts for effectively enhancing forecast stability and accuracy.

Given the severity of health and economic impacts of $PM_{2.5}$, reliable and precise $PM_{2.5}$ forecasting is urgently needed to implement early municipal warnings, crisis responses and emergency planning for mitigating health risks. Inspired by multi-task learning, a framework (MM-SVM) combining the Multi-output Support Vector Machine (M-SVM) and the Multi-Task Learning (MTL) algorithm is proposed for modelling regional multi-step-ahead $PM_{2.5}$ forecasts in this study. Meanwhile, the innovative nature of this study is indebted to the hybrid of the M-SVM and the MTL algorithm as well as its application for the first time to regional multi-step-ahead $PM_{2.5}$ forecasting. This study is explored with two main foci: (1) propose a multi-input and multi-output support vector machine model (M-SVM) for making regional $PM_{2.5}$ forecasts simultaneously; and (2) apply a multi-task learning algorithm to train the multi-output SVM model for sharing correlation information across related tasks. The proposed M-SVM data-driven model, trained by the multi-task algorithm, is implemented not only to identify the complex spatiotemporal pattern between regional meteorological inputs, air quality inputs and $PM_{2.5}$ multi-outputs at multiple air quality monitoring stations but to assess model reliability and accuracy via a detailed case study on the regional ground-level $PM_{2.5}$ forecasting in Taipei City of Taiwan. The original single-output SVM (S-SVM) is examined for comparative purpose. The remainder of the study is structured as follows: Section 2 presents the methods used in the analysis; Section 3 introduces the study area and materials; Section 4 presents discusses and results, and conclusions are then drawn in Section 5.

2. Methodology

The purpose of this paper is to verify the aforementioned argument and to propose a multi-input and multi-output data-driven model under the SVM architecture for improving forecast accuracy, where a multi-task learning algorithm is implemented with weight adjustment in training stages. The architectures of the S-SVM model (Fig. 1(a)) and the proposed MM-SVM model (Fig. 1(b)) are illustrated in Fig. 1, respectively. Multiple independent S-SVM models are constructed as a benchmark. The proposed MM-SVM model that aims at making regional multi-output forecasts and its overall effectiveness are shown in the following sections.

2.1. Single-output support vector machine (S-SVM)

The S-SVM is known to extract the pattern between multi-input variables and the single-output from an observed dataset with independent and identical distribution. Multiple independent S-SVM

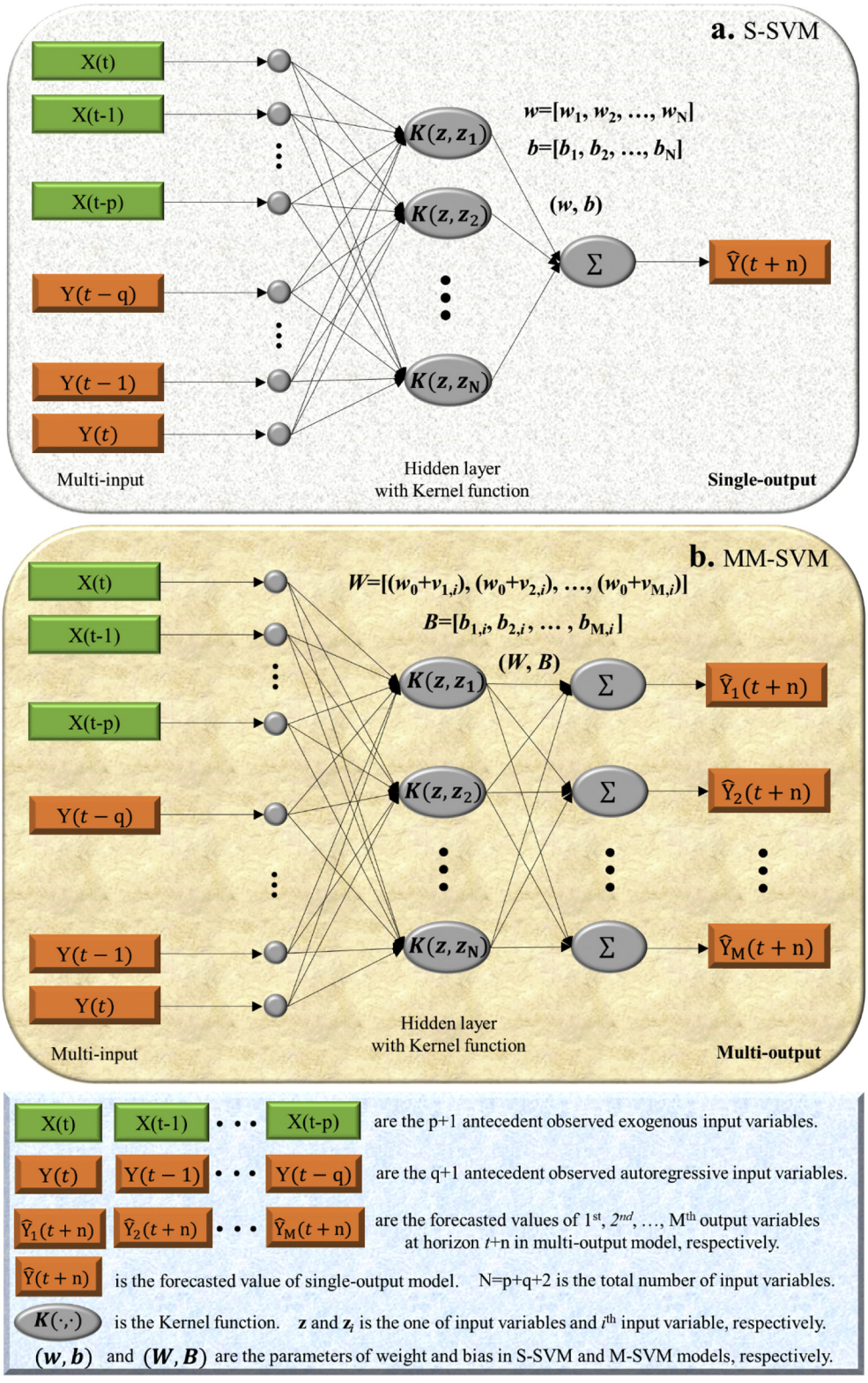


Fig. 1. Architecture of Support Vector Machine (SVM). a. Single-output SVM (S-SVM). b. Hybrid multi-output SVM with multi-task learning (MM-SVM).

models (Fig. 1(a)) are constructed to make air-quality forecasts at various monitoring stations, described below.

$$\hat{Y}(t+n) = \sum_{i=1}^{N} (\omega_i \cdot K(z, z_i) + b_i) \quad (1)$$

where $\hat{Y}(t+n)$ is the forecasted value of the output variable at horizon t + n. $K(\cdot,\cdot)$ is the Kernel function. z is one of variables in the dataset of observed exogenous (X(t), X(t − 1), ..., X(t − p)) and autoregressive (Y(t), Y(t − 1), ..., Y(t − q)) input variables. p and q are the time-lags of observed exogenous and autoregressive input variables, respectively. z_i is the ith input variable, and i = 1,2, ..., N. N is the total number of input variables, and N = p + q + 2.

ω_i and b_i are the parameters of the weight and the bias for the ith input variable in the S-SVM model, respectively.

The S-SVM solves the problem of single-output forecasting (Eq. (1)) by searching the optimal parameters of the weight (ω) and the bias (b) that minimize the objective function subject to constraints, described as follow.

$$\min g(w,\xi) = \frac{1}{2}\omega^{T}\omega + \gamma\frac{1}{2}\xi^{T}\xi \tag{2a}$$

subject to

$$\hat{Y}(t+n) = \sum_{i=1}^{N}(\omega_i \cdot K(z,z_i) + b_i) + \xi \tag{2b}$$

where $g(\cdot,\cdot)$ is the objective function in the single-output forecast problem. ξ is the slack variable. γ is the positive real regularized parameter for the slack variable. The conjugate gradient algorithm considered as one of fast and efficient optimization techniques could be used to handle the optimization problem shown in Eqs. (2a) and (2b) (Liu et al., 2015).

2.2. Multi-output support vector machine (M-SVM)

Under the condition of the same multiple inputs with the S-SVM model, the M-SVM model is proposed to extract the mapping relation from multi-input variables to multi-output variables. That is to say, it requires only one M-SVM model to forecast regional multi-outputs, described below.

$$\hat{Y}_1(t+n) = \sum_{i=1}^{N}((w_0 + \upsilon_{1,i}) \cdot K(z,z_i) + b_{1,i}) \tag{3}$$

$$\hat{Y}_2(t+n) = \sum_{i=1}^{N}((w_0 + \upsilon_{2,i}) \cdot K(z,z_i) + b_{2,i}) \tag{4}$$

$$\hat{Y}_m(t+n) = \sum_{i=1}^{N}((w_0 + \upsilon_{m,i}) \cdot K(z,z_i) + b_{m,i}) \tag{5}$$

$$\hat{Y}_M(t+n) = \sum_{i=1}^{N}((w_0 + \upsilon_{M,i}) \cdot K(z,z_i) + b_{M,i}) \tag{6}$$

$$w_0 = \frac{\lambda}{M}\sum_{m=1}^{M}\upsilon_m \tag{7}$$

where $\hat{Y}_1(t+n)$, $\hat{Y}_2(t+n)$, $\hat{Y}_m(t+n)$ and $\hat{Y}_M(t+n)$ are the forecasted values of the 1st, 2nd, mth and Mth output variables at horizon t + n, and m = 1,2, …,M. M is the total number of multi-output variables. $\upsilon_m = [\upsilon_{m,1}, \upsilon_{m,2}, \ldots, \upsilon_{m,N}]$ is the vector of the weight parameters for the mth output variable. $\upsilon_{1,i}$ $\upsilon_{2,i}$, $\upsilon_{m,i}$, $\upsilon_{M,i}$ and $b_{1,i}$, $b_{2,i}$, $b_{m,i}$ $b_{M,i}$ are the parameters of the weights and the biases for the ith input variable upon the 1st, 2nd, mth and Mth multi-output variables, respectively, where the $\upsilon_{m,i}$ is a small value close to 0 (→0) if the multi-outputs are very similar to each other. w_0 is the mean weight of the multi-output variables in the M-SVM model and represents the correlation among multi-output variables, where w_0 is a small value close to 0 (→0) if the multi-outputs are significantly distinct from each other. λ is the positive real regularized parameter for mean weight.

The M-SVM solves the problem of multi-output forecasts (Eqs. (3)–(7)) by searching the optimal parameters of the weight vector $V = (\upsilon_{1,i}, \upsilon_{2,i}, \ldots, \upsilon_{M,i})$ and the bias vector $B = (b_{1,i}, b_{2,i}, \ldots, b_{M,i})$ that minimizes the objective function subject to constraints, described as follow.

$$\min G(w_0, V, \varepsilon) = \frac{1}{2}\omega_0{}^{T}\omega_0 + \frac{1}{2}\frac{\lambda}{M}V^{T}V + \gamma\frac{1}{2}\varepsilon^{T}\varepsilon \tag{8}$$

subject to

$$\hat{Y}_1(t+n) = \sum_{i=1}^{N}((w_0 + \upsilon_{1,i}) \cdot K(z,z_i) + b_{1,i}) + \xi_1 \tag{9}$$

$$\hat{Y}_2(t+n) = \sum_{i=1}^{N}((w_0 + \upsilon_{2,i}) \cdot K(z,z_i) + b_{2,i}) + \xi_2 \tag{10}$$

$$\hat{Y}_m(t+n) = \sum_{i=1}^{N}((w_0 + \upsilon_{m,i}) \cdot K(z,z_i) + b_{m,i}) + \xi_m \tag{11}$$

$$\hat{Y}_M(t+n) = \sum_{i=1}^{N}((w_0 + \upsilon_{M,i}) \cdot K(z,z_i) + b_{M,i}) + \xi_M \tag{12}$$

where $G(\cdot,\cdot,\cdot)$ is the objective function in the multi-output forecast problem. $\varepsilon = (\xi_1, \xi_2, \ldots, \xi_M)$ is the vector of slack variables.

Eqs. (8)–(12) are reformulated into a linear matrix equation system, described as follow.

$$\begin{bmatrix} S & 0 \\ 0 & H \end{bmatrix}\begin{bmatrix} B \\ H^{-1}PB + w_0 \end{bmatrix} = \begin{bmatrix} P^{T}H^{-1}Y \\ Y \end{bmatrix}$$

$$S = P^{T}H^{-1}P$$

$$H = \left[Z^{T}Z, M, M\right] + \frac{1}{\gamma}I + \frac{M}{\lambda}Z^{T}Z \tag{13c}$$

where S is the positive definite matrix. P is the identity matrix with M columns. H is the transformation matrix. I is the identity matrix with M columns and M rows. Z is the matrix of input variables, and $Z = [z_1, z_2, \ldots, z_N]$. Y is the matrix of multi-output variables, and $Y = [\hat{Y}_1(t+n), \hat{Y}_2(t+n), \ldots, \hat{Y}_M(t+n)]$.

Inasmuch as the common techniques for optimizing the parameters of the standard M-SVM model, including the gradient descent algorithm and the conjugate gradient algorithm, are easily trapped in the bottlenecks of falling into local minima and are time-consuming, they could not effectively solve the joint optimization of parameters of the M-SVM model. That is to say, the M-SVM model possesses not only a more complex model structure but also much more parameters than the S-SVM model, which implies the M-SVM model demands for more sophisticated techniques to greatly contribute to model stability and generalizability.

2.3. Hybriding the M-SVM model and multi-task learning algorithm (MM-SVM)

The MTL proposed by Baxter (1997) has the capability to synergistically optimize the parameters of multi-output models. Many studies empirically as well as theoretically demonstrated that learning multiple related tasks simultaneously, with a hope of appropriate information sharing across tasks, would significantly improve model performance, as compared to learning each task independently (e.g., Chandra et al., 2017; Zhao et al., 2017; Shireen et al., 2018).

For S-SVM and M-SVM models, the radial basis function is set as the Kernel function, described below.

$$K(z,z_i) = e^{-\theta\|z-z_i\|^2} \tag{14}$$

where θ is the parameter of the radial basis function, and the radial basis

function is the Gaussian function if the θ is equal to $\frac{1}{2\sigma^2}$ (σ is the standard deviation).

For the MM-SVM model constructed in this study, the MTL algorithm used to train the M-SVM implements the following computation steps.

Initialization: initialize the radial basis function parameter ($\theta_1 > 0$), the positive real regularized parameter ($\gamma_1 > 0$) of the slack variable, the positive real regularized parameter ($\lambda_1 > 0$) of the weight based on the grid dataset $\{2^{-k}, 2^{-k+2}, \ldots, 2^{k}\}$ (Evgeniou and Pontil, 2004), as well as the maximal generation number (G_{max}).
For $j = 1, \ldots, G_{max}$
Step 1: input $(\theta_i, \gamma_i, \lambda_i)$ and solve η and β of matrix equations $H\eta = P$ and $H\beta = Y$ according to the linear matrix equation system, respectively (Eqs. (13a), (13b), (13c)).
Step 2: calculate the positive definite matrix $S = P^T\eta$.
Step 3: search solution: $B = S^{-1}\eta^T Y$ and $V = \beta - \eta B$.
Step 4: terminate the computation according to the stop criteria by evaluating the solution (B and V) based on the following error function through Steps 1–3.

$$MSE = \frac{1}{2T}\sum_{t=1}^{T}\left(\hat{Y}_m(t+n) - Y_m(t+n)\right)^2 \quad (15)$$

where MSE is the mean square error. $\hat{Y}_m(t+n)$ and $Y_m(t+n)$ are the forecasted and the observed value of the mth output variable at horizon t + n, respectively. T is the number of time steps.
If the iteration number is less than the G_{max}, then repeat Steps 1–3. Otherwise, stop and output the optimized parameters corresponding to the minimal error function MSE.
End for
Output: Save the optimized parameters of the fully trained M-SVM model, including the radial basis function parameter (θ^*), the positive real regularized parameter (γ^*) of the slack variable, the positive real regularized parameter (λ^*) of weights, the weight vector (V^*) and the bias vector (B^*).

The comparison between S-SVM and MM-SVM models is summarized as: (1) the former is a single-output model while the latter is a multi-output data-driven model in the perspective of model architecture, hence the former needs to construct multiple models while the latter needs to construct only one model for regional multi-output forecasting; (2) there are N × M weight parameters for M independent S-SVM models while there are N × M weight parameters and one positive real regularized parameter (λ) of mean weight for the MM-SVM model. In other words, the mean weight with a positive real regularized parameter (λ) in the MM-SVM model is utilized to share the potential nonlinear spatial correlation among air quality monitoring stations; and (3) the former could utilize many mature optimization methods (e.g., the conjugate gradient method) for optimizing model parameters while the latter demands for more sophisticated techniques, such as the multi-task learning algorithm, to optimize model parameters.

3. Study area and materials

With the fast-growing economy and population, air quality deterioration in Taiwan has become a hot topic in recent years. Taipei City is the center of politics, commerce, and culture in Taiwan and covers an area of 272 km^2 with a population of 2.69 million in 2016. People across Taipei City nowadays are forced to deal with the high-level invasion of $PM_{2.5}$. Air pollution is not just about sore throats and respiratory diseases but a matter of life or death. Therefore, healthy and green urban development demands for accurate multi-step-ahead $PM_{2.5}$ forecasts that adequately deal with the high variability of regional air quality.

Fig. 2 shows the locations of Taipei City, 16 meteorological monitoring stations and five air quality monitoring stations in the study area. Among them, Stations A1 (Yonghe) and A2 (Sanchong) are traffic stations (i.e., stations located in areas of heavy traffic), Stations A3 (Songshan) and A4 (Shilin) are general stations, and Station A5 (Yangming) is a park station (i.e., a station located in the Yangmingshang Park). Traffic stations are the representative of traffic loads for monitoring the primary air pollutant mechanism. General stations are the representative of human activities and commercial trading for monitoring the secondary air pollutant mechanism. Park stations are the representative of natural situations, with less human intervention. More description about the functions of the five stations could be found in the official statement released by the Environmental Protection Administration (EPA), ROC (https://taqm.epa.gov.tw/taqm/en/b0101.aspx). It is noted that park stations selected by our EPA for monitoring air quality are in use for conservation purpose. The locations of the monitoring sensors installed at park stations would keep away from automobile roads, parking lots, combustion sources and industrial factories. As compared with traffic, industrial and general stations, air

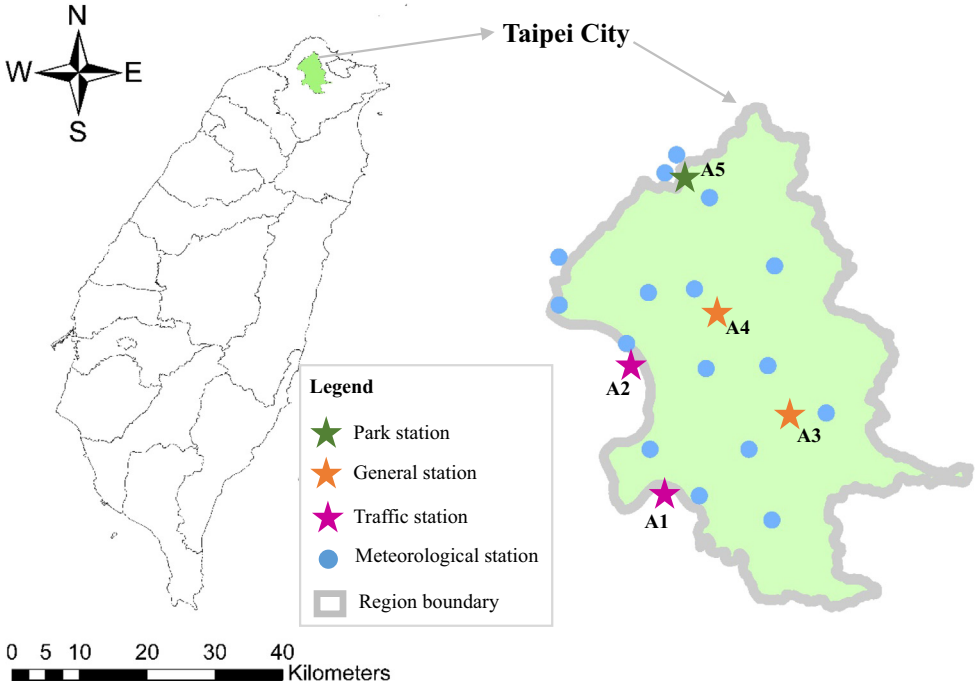


Fig. 2. Distribution of air quality and meteorological monitoring stations in Taipei City. Stations A1 (Yonghe) and A2 (Sanchong) are traffic stations (i.e. stations located in areas of heavy traffic). Stations A3 (Songshan) and A4 (Shilin) are general stations. Station A5 (Yangming) is a park station (i.e. a station located in a park).

quality at the park station (i.e., located in the Yangmingshang Park with elevations ranging from 200 to 1120 m) in this study is considered representative of the situation close to nature

Data of five meteorological factors (rainfall, temperature, wind speed, wind direction and relative humidity) and eight air quality factors ($PM_{2.5}$, PM_{10}, O_3, NO_x, NO_2, NO, SO_2, CO) collected from 2010 to 2016 (7 years) in the study area are available. A total of 61,368 (=[(2 × 366) + (5 × 365)] × 24) hourly datasets are used in this study, where 35,064 data (4 years) are used for model training while the remaining 26,304 data (3 years) are used for model testing.

To reduce the negative effect of the different scales of input data on model's learning ability, all thirteen input variables are standardized to the same scale. To obtain stable convergence of the developed model, the normal standardization is applied to data pre-processing. The standardization formula is defined as follows.

$$X^*(t) = \frac{X(t) - \overline{X}}{\sigma} \tag{16}$$

where $X^*(t)$ is the normal standardization for input data in the tth time. $\overline{X}$ and σ are the average and standard deviation of input data, respectively.

The Root-Mean-Square-Error (RMSE) and the goodness-of-fit with respect to the benchmark (G_{bench}) are carried out for comparison purpose. The RMSE and G_{bench} are defined as follows.

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(\hat{Y}(t) - Y(t)\right)^2}, \quad RMSE \geq 0 \tag{17}$$

$$G_{bench} = \left(1 - \frac{\sum_{i=1}^{T}\left(\hat{Y}(t) - Y(t)\right)^2}{\sum_{i=1}^{T}\left(Y(t) - Y_{bench}(t)\right)^2}\right) \times 100\%, \quad G_{bench} \leq 100\% \tag{18}$$

where $\hat{Y}(t)$ and $Y(t)$ is the forecasted and observed values of the output variable at the tth time, respectively. $Y_{bench}(t)$ is the observed data shifted backwards by one or more time lags, e.g., for the nth-step-ahead forecast, $Y_{bench}(t) = Y(t - n)$.

Table 1 presents the statistic indexes of seasonal $PM_{2.5}$ concentration at five air quality monitoring stations. We notice that the statistic indexes of the maximum, average and standard derivation at traffic stations (A1 and A2) are the highest while those in the park station (A5) are the lowest, which could be due to the primary source of particulate matter of a station. For instance, vehicle exhaust emission is the primary source of particulate matter at traffic stations; air pollutant emission from residential and commercial activities is the primary source of particulate matter at general stations; and atmospheric transport is the primary trigger of particulate matter at the park station. In other words, vehicle transportation is a stronger driving force of air pollutant than human activities in Taipei City. It is noted that the five monitoring stations do represent three situations (traffic, general and park stations) and significant differences show up by the characteristics of the monitoring data. The statistical analysis shown in Table 1 provides persuasive evidence to support the division of the air quality monitoring stations released by the official statement of our EPA.

As compared with other correlation analysis techniques (e.g., Pearson coefficient, Spearman coefficient, or principal component analysis), the Kendall tau coefficient (Maidment, 1993) has advantages including: (1) the investigative data need not to satisfy the hypothesis relevant to the requirement for a normal distribution; (2) it is commonly used to analyze the non-linear correlation characteristics between two datasets; and (3) it has wider applicability owing to its ability of non-parametric statistical analysis (Chang et al., 2012; Méheust et al., 2012; Chen et al., 2013). Therefore, the Kendall tau coefficient is conducted in this study to identify the highest correlation regarding the time lags between input variables (meteorological and air quality factors) and output variables ($PM_{2.5}$). As known, longer time lags imply a demand for more model inputs, which would increase model complexity and raise model uncertainty (Jiang et al., 2017; Zhai et al., 2018). For achieving a reduction of white noise in multi-input and multi-output data-driven models applied to regional multi-step-ahead $PM_{2.5}$ forecasting, the threshold of Kendall tau coefficient is set as 0.5 in this study so that only time lags with higher correlation values (i.e., >0.5) would be selected as model inputs to produce more stable and reliable forecasting results. Table 2 shows the results (mean values) of the correlation analysis between input variables and output variables using the Kendall tau coefficient. According to the highest values of the Kendall tau coefficients (≥0.5), the time lags of air-quality factors are set as 1 h–4 h for traffic stations (A1 and A2) and 1 h–2 h for general and park stations (A3, A4, and A5) while the time lags for meteorological factors are set as 1 h–4 h at 16 meteorological monitoring stations, respectively. Five independent S-SVM models are established individually for five air quality monitoring stations, but only one MM-SVM model is established for the same five air quality monitoring stations. The parameters of the S-SVM models and the MM-SVM model are the same, except for the positive real regularized parameter (λ) of mean weight in the MM-SVM model. S-SVM and MM-SVM models are evaluated by the RMSE and G_{bench}.

Table 1
Statistic indexes of seasonal $PM_{2.5}$ concentration at five air quality monitoring stations in Taipei City.

Season	Statistic index	Air quality monitoring stations				
		A1[a]b	A2	A3	A4	A5
Spring	Maximum	377 [b]	358	259	278	147
	Mean	25	27	21	18	13
	Minimum	0	0	0	0	0
	Standard deviation	18	16	13	12	8
Summer	Maximum	226	215	155	167	88
	Mean	15	16	13	11	8
	Minimum	0	0	0	0	0
	Standard deviation	11	10	8	7	5
Autumn	Maximum	264	251	181	195	103
	Mean	18	19	15	13	9
	Minimum	0	0	0	0	0
	Standard deviation	13	11	9	8	6
Winter	Maximum	358	340	246	264	140
	Mean	24	26	20	17	12
	Minimum	0	0	0	0	0
	Standard deviation	17	15	12	11	8
Annual	Maximum	377	358	259	278	147
	Mean	22	24	17	15	10
	Minimum	0	0	0	0	0
	Standard deviation	16	15	10	10	6

[a] Stations A1 (Yong-He) and A2 (San-Chong) are traffic stations (i.e. stations located in areas of heavy traffic). Stations A3 (Song-Shan) and A4 (Shi-Lin) are general stations. Station A5 (Yangming) is a park station (i.e. a station located in a park).
[b] unit: $\mu g/m^3$

Table 2
Results (mean values) of the correlation analysis between input variables and the output variable using the Kendall tau coefficient.

Station name	Eight air quality factors at 5 air-quality monitoring stations				Five meteorological factors at 16 meteorological monitoring stations			
	t + 1	t + 2	t + 3	t + 4	t + 1	t + 2	t + 3	t + 4
A1	0.84	0.73	0.59	0.54	0.75	0.68	0.61	0.52
A2	0.83	0.75	0.62	0.51	0.81	0.72	0.64	0.53
A3	0.77	0.58	0.45	0.37	0.83	0.74	0.67	0.55
A4	0.67	0.61	0.41	0.29	0.86	0.79	0.67	0.55
A5	0.72	0.58	0.39	0.18	0.78	0.67	0.59	0.54

4. Results and discussion

4.1. Spatial stability of SVM models

To show the merit of the proposed MM-SVM model (M-SVM incorporated with MTL), the results of S-SVM and MM-SVM models in training stages at the horizon of t + 4 shown in Fig. 3 are assessed. It indicates that the error function values of the MM-SVM model are smaller than those of the S-SVM models, where a sharp drop occurs around the 200th generation. The reason is that the MM-SVM model utilizes more weight parameters to fit the potentially non-linear interrelationships of air quality among five $PM_{2.5}$ monitoring stations, which reduces error accumulation and propagation during modelling. It is noted that the error function values of the MM-SVM model show a significant decreasing trend with less fluctuation, which implies the S-SVM would easily trigger forecast instability. These results clearly show that the proposed MM-SVM model can overcome the instability shortcomings caused by multiple independent S-SVM models for regional multi-step-ahead $PM_{2.5}$ forecasting.

In addition, we test the spatial stability of these constructed models. Fig. 4 shows the performance of S-SVM and MM-SVM models for regional and site-specific multi-step-ahead $PM_{2.5}$ forecasts at horizons t + 1 up to t + 4 at five monitoring stations in Taipei City, respectively, which clearly demonstrates the following findings.

1) The S-SVM model produces an unstable and inferior performance for $PM_{2.5}$ forecasting in the whole study region (Taipei City) and at each five air quality monitoring stations. It indicates that the S-SVM model cannot provide a stable and precise regional multi-step-ahead forecast results when multiple independent S-SVM is individually constructed.
2) The MM-SVM model has the best forecast performance not only at individual air quality monitoring station but also for the whole region (Taipei City), in terms of RMSE and G_{bench} values. It demonstrates that the proposed MM-SVM model that adequately considers underlying non-linear spatial relationships among five $PM_{2.5}$ monitoring stations can effectively adjust synaptic weights to provide reliable and accurate regional multi-step-ahead $PM_{2.5}$ forecasts.
3) The MM-SVM model has the best testing performance of all the cases in terms of RMSE and G_{bench} values. It appears that the MM-SVM model produces much higher G_{bench} values but much smaller RMSE values than the S-SVM model in both training and testing stages. It is noted that the RMSE (G_{bench}) values of the S-SVM model significantly increase (decrease) in both stages at time steps t + 2 up to t + 4 while those of the MM-SVM model remain relatively stable. The MM-SVM model also has the best performance at all stations according to RMSE values. For traffic stations (A1 and A2), the performance of the MM-SVM model is significantly better than that of the S-SVM model, whereas the performance of MM-SVM and S-SVM models makes less difference at general and park stations (A3, A4 and A5). As shown in Table 3, taking horizon t + 4 for example, the improvement rates in terms of G_{bench} and RMSE values reach 22.39% and 38.43% at Station A1, respectively, but remain only 4.52% and 13.22% at Station A3, respectively. In addition, it is an interesting finding that the improvement rates in terms of G_{bench} and RMSE values significantly increase overtime from t + 2 to t + 4 at all stations. In other words, the proposed MM-SVM model is able to produce more stable and precise multi-step-ahead forecasts by identifying the heterogeneities of different air quality monitoring stations. The reason is that the correlation between traffic stations not only stems from traffic volumes but also stems from meteorological factors (e.g. rainfall and wind speed), while the correlation between general stations only stems from meteorological factors in the perspective of spatial relationship. The Kendall tau correlation for $PM_{2.5}$ between traffic stations (coefficient = 0.87) is stronger than that between the other stations (average coefficient = 0.73). Therefore, the MM-SVM model produces superior performance at traffic stations than the other stations. The results indicate that the multi-output model does provide valuable information for regional $PM_{2.5}$ forecasting.

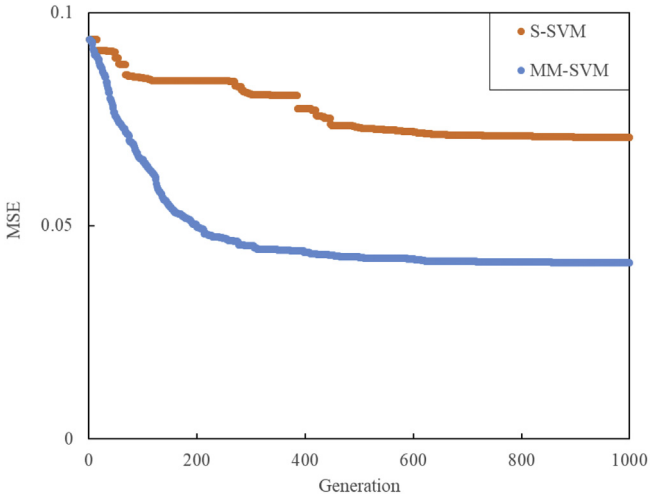


Fig. 3. Error function values of S-SVM and MM-SVM models in training stages at horizon t + 4 (the error function is the Mean Squared Error with normalized dataset). The MSE of the S-SVM model is the mean value of the MSE of five independent S-SVM models corresponding to five air quality monitoring stations in Taipei City. The MM-SVM denotes the hybrid of the multi-output SVM and the multi-task learning.

From the perspective of spatial stability, the MM-SVM model is very beneficial to regional air quality forecasting since the proposed multi-output data-driven model not only can enhance model reliability and accuracy but also can provide smarter applications than the single-output data-driven model for regional multi-step-ahead air quality forecasts.

4.2. Temporal stability of SVM models

Next, we evaluate the temporal stability of these constructed models. Taking the indicator RMSE for example, Fig. 5 presents the testing performance between the S-SVM models and the MM-SVM model for regional $PM_{2.5}$ forecasts in four seasons. For a regional scale (Taipei City), the MM-SVM model shows superior performance in four seasons according to RMSE values. It is noted that the performance of the M-SVM is significantly better than that of the S-SVM in summer and autumn, while the performance between MM-SVM and S-SVM models is less distinct in spring and winter. The reason is that the meteorological relationship in summer and autumn (average Kendall tau coefficient for meteorological factors = 0.75) is stronger than that in spring and winter (average Kendall tau coefficient for meteorological factors = 0.62). Therefore, the stronger correlation between air-quality monitoring stations is the trigger for enhancing the performance of the MM-SVM model. In short, these findings are also very beneficial to the data-driven modeler because an independent model is not essential for each season, in terms of temporal stability comparison between S-SVM and MM-SVM models.

From a statistical standpoint, data-driven models could identify and learn input-output patterns based on a majority (>75%) of observations (i.e. above the lower quartile) while a common technical bottleneck would occur in most of the data-driven models, which means it is easily inclined to systematically underpredict particulate matter concentrations at the occasions of very high concentrations (higher than the value of the upper quartile), especially for the maximal $PM_{2.5}$ concentration events (Li et al., 2016; Lv et al., 2016; Nieto et al., 2018). As a consequent, we select the maximal $PM_{2.5}$ events corresponding to the three

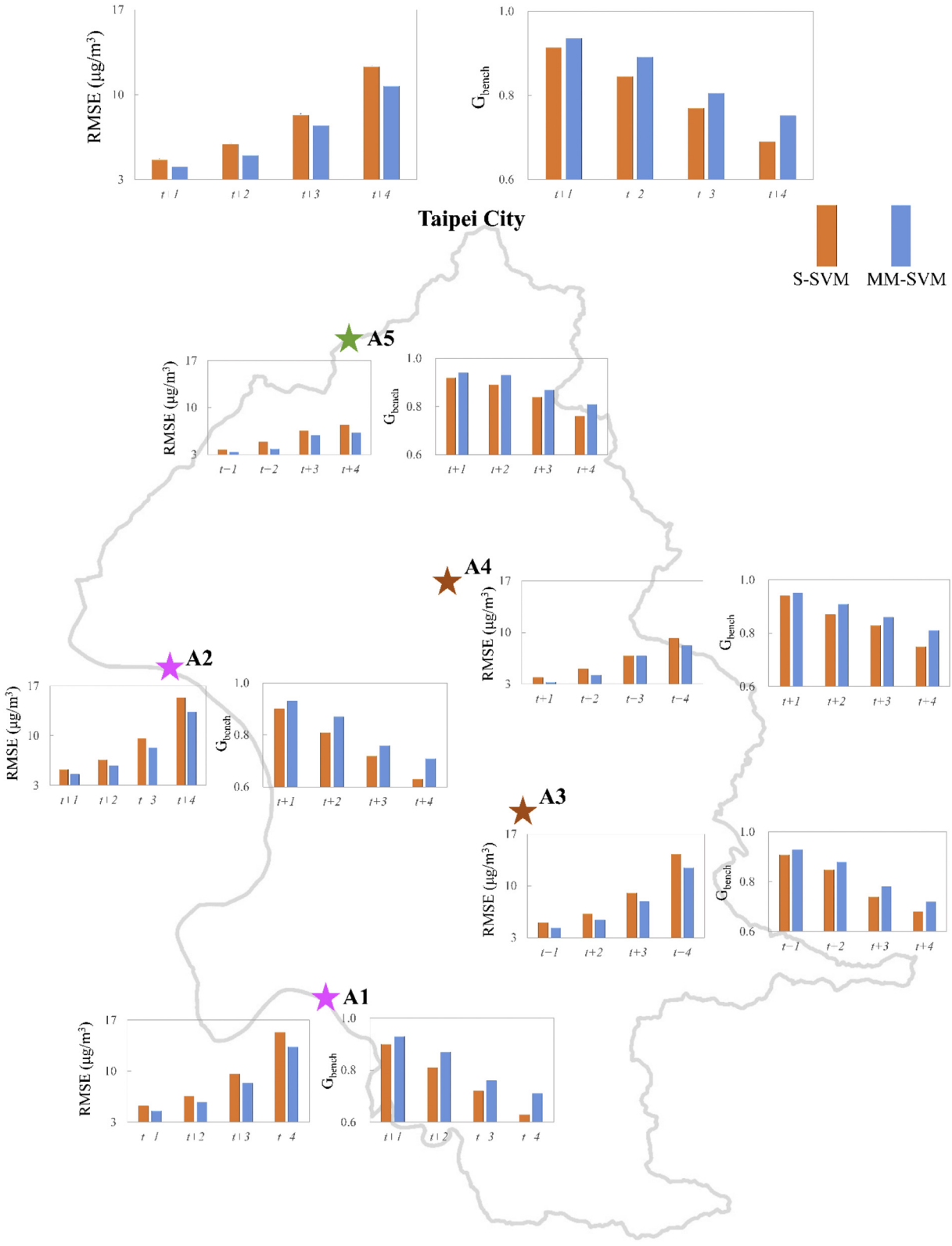


Fig. 4. Model performance in testing stages of the S-SVM and MM-SVM for regional and site-specific multi-step-ahead $PM_{2.5}$ forecasts at horizons from t + 1 to t + 4 at five monitoring stations in Taipei City, respectively. The MM-SVM denotes the hybrid of the multi-output SVM and the multi-task learning.

kinds of air-quality monitoring stations in the testing stage to test the performance of the two models. Finally, to clearly distinguish the forecast ability of S-SVM and MM-SVM models, three $PM_{2.5}$ events at three monitoring stations (traffic station A1 with maximal $PM_{2.5}$ concentration exceeding 250 μg/m^3; general station A3 with maximal $PM_{2.5}$ concentration exceeding 200 μg/m^3; and park station A5 with maximal $PM_{2.5}$ concentration exceeding 100 μg/m^3) are selected to test both models through assessing the goodness-of-fit between observations and forecasts at horizon t + 4, as shown in Fig. 6. It reveals that the MM-SVM model is able to well forecast multi-step-ahead $PM_{2.5}$,

Table 3
Improvement rates of two indicators (G_{bench} and RMSE) in the testing stages of the multi-step-ahead $PM_{2.5}$ forecasting models (the MM-SVM model in comparison with the S-SVM models).

Station name	Horizon	Improvement rate[a] (%)	
		G_{bench}	RMSE
A1	t + 1	2.17	8.16
	t + 2	13.41	16.81
	t + 3	16.00	31.46
	t + 4	22.39	38.43
A2	t + 1	3.19	8.16
	t + 2	11.39	17.61
	t + 3	12.05	30.78
	t + 4	15.28	39.04
A3	t + 1	2.20	10.08
	t + 2	2.35	11.52
	t + 3	2.70	12.11
	t + 4	4.41	13.25
A4	t + 1	1.89	12.78
	t + 2	1.94	12.87
	t + 3	2.70	13.22
	t + 4	4.52	13.22
A5	t + 1	1.99	12.96
	t + 2	2.46	13.07
	t + 3	2.91	13.24
	t + 4	4.62	13.39
Regional	t + 1	2.29	10.43
	t + 2	6.31	14.38
	t + 3	7.27	20.16
	t + 4	10.24	23.47

[a] Improvement rate of $G_{bench} = \frac{(G_{bench}\{MMSVM\} - G_{bench}\{SSVM\})}{G_{bench}\{SSVM\}} \times 100\%$
Improvement rate of $RMSE = \frac{(RMSE\{SSVM\} - RMSE\{MMSVM\})}{RMSE\{SSVM\}} \times 100\%$.

whereas the S-SVM model has significant time-lag phenomena as well as larger gaps between observations and forecasts, where the forecast precision significantly decreases at horizon t + 4. It appears that the MM-SVM model is able to trace the trails of $PM_{2.5}$ events, significantly mitigate time-lag effects, as well as produce much accurate and reliable

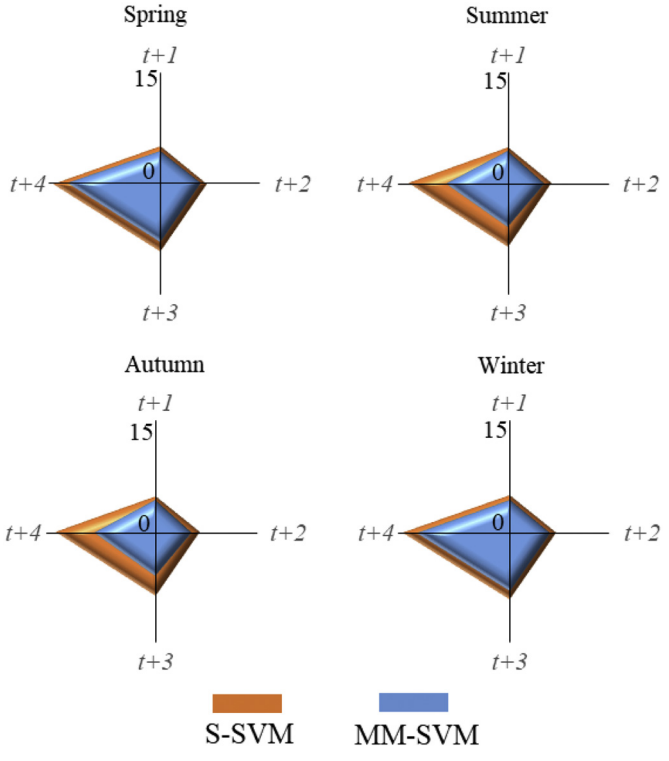


Fig. 5. Seasonal performance (RMSE (μg/m^3)) in the testing stage of the S-SVM and MM-SVM models for regional $PM_{2.5}$ forecasts at horizons from t + 1 to t + 4 in Taipei City. The MM-SVM denotes the hybrid of the multi-output SVM and the multi-task learning.

multi-step-ahead forecasts. In addition, from the standpoint of forecast accuracy, we have two important findings. The first finding is that the S-SVM model has better performance at the general station A3 and the park station A5 than at the traffic station A1. The second finding is that the MM-SVM model not only can perform as well as the S-SVM model at the general station A3 and the park station A5 but also can significantly enhance the forecast accuracy of $PM_{2.5}$ concentrations at the traffic station A1. That is to say, the influences of primary emission associated with meteorological conditions (e.g., general station A3) on MM-SVM and S-SVM models is insignificant while the influences of secondary emission associated with meteorological conditions (e.g., traffic station A1) on MM-SVM and S-SVM models make a significant difference. The reasons why the MM-SVM model outperforms the S-SVM model are that: (1) the standard S-SVM cannot cope with the multi-output case. In addition, the traditional parameter optimization procedure (e.g., conjugate gradient algorithm) of the SVM individually trains each of the multiple independent S-SVM models for learning each parameter, which could only extract the underlying non-linear correlation among multi-input variables, and thus disregards the potentially non-linear cross relatedness among different outputs; and (2) the MM-SVM model not only could reserve the excellent capability of the S-SVM model but also could be advantageous to learn all outputs simultaneously as well as contribute this correlation to multi-task learning by means of the positive real regularized parameter (λ) of mean weight.

From the standpoint of air-pollutant mechanisms, we further explore the reasons why the S-SVM and MM-SVM models produce different performances at traffic, general and park stations. In Taipei City with fast urban development, regional air quality frequently interacts with intensive human activities, traffic loads and commercial trading. From the perspective of monitoring functions and spatial distribution, the five air quality monitoring stations are typical and representative for regional air quality of Taipei City. As known, a high $PM_{2.5}$ event is usually associated with secondary processes either from regional transportation of aged secondary aerosol or secondary transformation of gaseous pollutants, while a $PM_{2.5}$ event driven by primary or natural processes would be expected to correlate with local weather conditions and primary emissions (Witkowska and Lewandowska, 2016; Berardis and Eleonora, 2017; Zhang et al., 2018). From the perspective of the S-SVM model, it produces better performance (higher G_{bench} values and lower RMSE values, see Fig. 4) at the general stations (A3 & A4) and the park station (A5) than at traffic stations (A1 & A2). From the perspective of the MM-SVM model in comparison with the S-SVM model, the MM-SVM model gains better improvement rates of G_{bench} and RMSE (see Table 3) at traffic stations (A1 & A2) than at the park station (A5) and general stations (A3 & A4). That is to say, the MM-SVM model not only could greatly improve the forecast accuracy of $PM_{2.5}$ concentrations at traffic stations representative of secondary processes by capturing the correlation in regional transportation of aged secondary aerosol but also could perform as well as the S-SVM model at general stations representative of primary processes and the park station representative of natural processes.

The developed MM-SVM model can be applied not only to modelling the heterogeneities in different types of air pollutant-generating mechanisms (e.g. primary mechanism, secondary mechanism and nature situation) but also to mapping the heterogeneities of air pollutant onto different seasons by utilizing mean weight with the positive real regularized parameter for fitting the potentially non-linear interrelationships among $PM_{2.5}$ monitoring stations. In practice, the proposed MM-SVM model for simultaneously making regional $PM_{2.5}$ forecasts could be affected by missing data because the probability of missing a datum would be much higher in a group of monitoring stations than in a single station. This missing data problem in the real application could be solved by filling the missing data fields with the previously observed data and/or model output data, with the data interpolated from nearby monitoring stations, or with the estimates using other more sophisticated methods to increase model reliability. In this

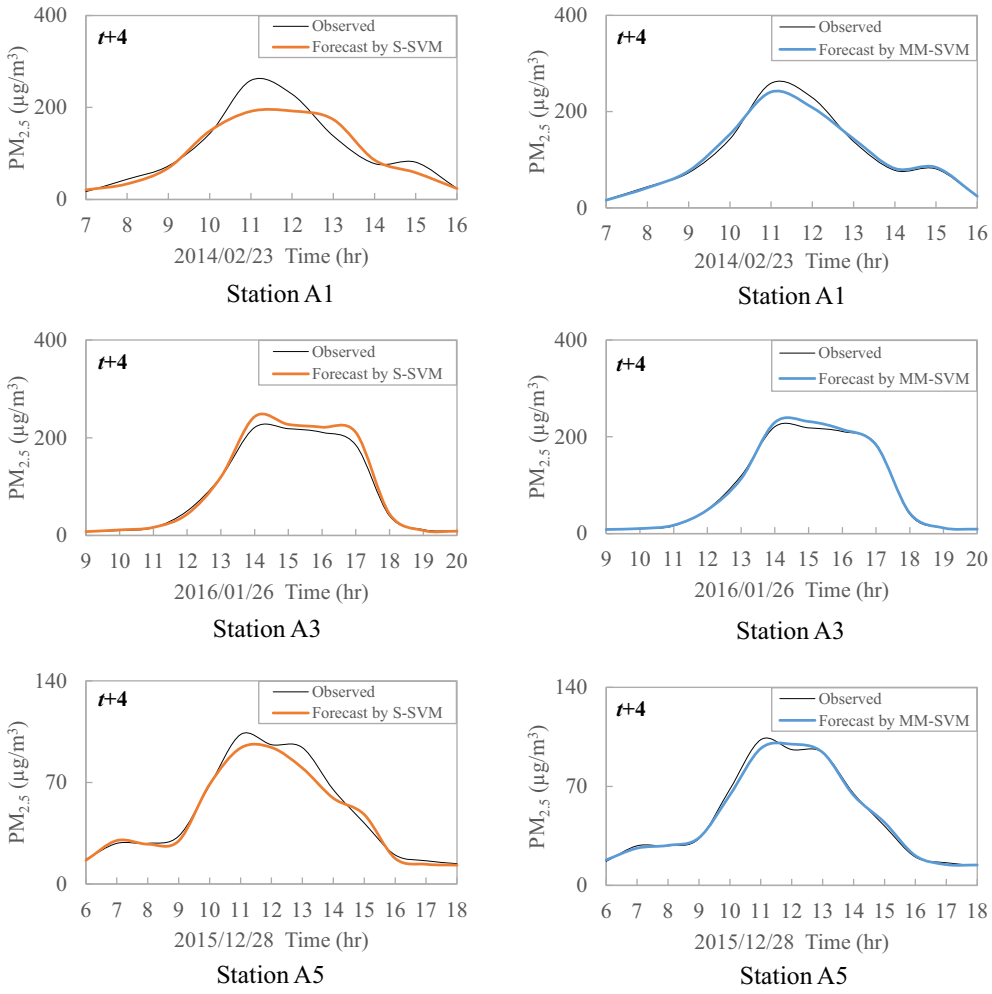


Fig. 6. Multi-step-ahead $PM_{2.5}$ forecast results of S-SVM and MM-SVM models in the testing stages at horizon t + 4 at the traffic Station A1 (Yonghe), the general Station A3 (Songshan) and the park station A5 (Yangming), respectively. The MM-SVM denotes the hybrid of the multi-output SVM and the multi-task learning. The test event with maximal $PM_{2.5}$ concentration exceeding 250 μg/m^3 occurred at the traffic Station A1. The test event with maximal $PM_{2.5}$ concentration exceeding 200 μg/m^3 occurred at the general Station A3. The test event with maximal $PM_{2.5}$ concentration exceeding 100 μg/m^3 occurred at the park Station A5.

study, we focus on exploring the proposed MM-SVM model for effectively increasing the accuracy of regional multi-step-ahead forecasts through tackling error accumulation and propagation that is commonly encountered in regional forecasting. The proposed model is a static one, and the datasets used for model training and testing are historical data without any missing data. Additionally, the accuracy, robustness and uncertainty of the investigative models are considered as three main research interests in the domain of air-quality forecasting. In our future study, we would pay attention to addressing the robustness and uncertainty of multi-step-ahead air-quality forecast models.

5. Conclusion

This research is motivated by the wide concern raised in recent years, namely the increase of frequency and intensity of extreme air quality events caused by climate change as well as human activities, along with population boosting and urban development. Accurate and reliable multi-step-ahead air quality forecasts are very crucial and beneficial to mitigate health risks caused by outdoor activities. In this study, a multi-output SVM incorporated with a multi-task learning algorithm (MM-SVM) was explored for the first time to make regional multi-step-ahead $PM_{2.5}$ forecasts. Its capability of efficient learning and accurate forecasting was tested and verified at five air quality monitoring stations in Taipei City. The single-output SVM constructed independently for individual air quality monitoring station was implemented for comparative analysis.

The results of regional $PM_{2.5}$ forecasts demonstrated that the proposed MM-SVM model performed more prominently than the S-SVM model in multi-step-ahead forecasting for all the cases, in terms of RMSE and G_{bench}. It showed that the MM-SVM model that adequately takes underlying non-linear spatial relationships among five $PM_{2.5}$ monitoring stations could effectively adjust synaptic weights and thus the model could provide reliable and accurate regional multi-step-ahead $PM_{2.5}$ forecasts. When assessing the regional $PM_{2.5}$ forecast models established for Taipei City, the proposed MM-SVM model could significantly mitigate time-lag phenomena. The S-SVM model, however, failed to achieve satisfactory forecast results in both training and testing stages, which implied the SVM demanded for more sophisticated techniques, such as the multi-task learning algorithm under a multi-output architecture, to enhance model stability and generalizability in spatiotemporal scales. The developed MM-SVM model can effectively model the heterogeneities in different types of air pollutant-generating mechanisms (e.g., primary mechanism, secondary mechanism and nature situation) as well as can map the heterogeneities of air pollutant in different seasons to fitting the potentially non-linear inter-relationships among different air quality monitoring stations. We demonstrate the developed MM-SVM model can be topical and timely for regional air quality forecasting and early warning to mitigate human health risks, coinciding with the Green Urbanization

Development Pathway promoted by governmental administrations and stakeholders to adapt to fast-changing conditions by considering the relationships and feedbacks between natural and human-induced environmental changes.

Acknowledgement

This study is financially supported by the Ministry of Science and Technology, Taiwan, ROC (MOST: 106-3114-M-002-001-A and 106-2811-B-002-087-) and the China Postdoctoral Science Foundation (No. 2017M620336). The datasets provided by the Environmental Protection Administration, Taiwan, ROC, are acknowledged. The authors would like to thank the Editors and anonymous Reviewers for their valuable and constructive comments related to this manuscript.

References

Apte, J.S., Marshall, J.D., Cohen, A.J., Brauer, M., 2015. Addressing global mortality from ambient $PM_{2.5}$. Environ. Sci. Technol. 49 (13), 8057–8066.
Bai, Y., Li, Y., Wang, X., Xie, J., Li, C., 2016. Air pollutants concentrations forecasting using back propagation neural network based on wavelet decomposition with meteorological conditions. Atmos. Pollut. Res. 7 (3), 557–566.
Baxter, J., 1997. A Bayesian/information theoretic model of learning to learn via multipe task sampling. Mach. Learn. 28 (1), 7–39.
Berardis, D., Eleonora, M., 2017. Analysis of major pollutants and physico-chemical characteristics of $PM_{2.5}$ at an urban site in Rome. Sci. Total Environ. 617, 1457–1468.
Chandra, R., Ong, Y.S., Goh, C.K., 2017. Co-evolutionary multi-task learning with predictive recurrence for multi-step chaotic time series prediction. Neurocomputing 243, 21–34.
Chang, L.C., Chen, P.A., Chang, F.J., 2012. Reinforced two-step-ahead weight adjustment technique for online training of recurrent neural networks. IEEE Trans. Neural Netw. Learn. Syst. 23 (8), 1269–1278.
Chen, P.A., Chang, L.C., Chang, F.J., 2013. Reinforced recurrent neural networks for multi-step-ahead flood forecasts. J. Hydrol. 497, 71–79.
Cobourn, W.G., 2010. An enhanced $PM_{2.5}$ air quality forecast model based on nonlinear regression and back-trajectory concentrations. Atmos. Environ. 44 (25), 3015–3023.
Coelho, M.C., Fontes, T., Bandeira, J.M., Pereira, S.R., Tchepel, O., Dias, D., et al., 2014. Assessment of potential improvements on regional air quality modelling related with implementation of a detailed methodology for traffic emission estimation. Sci. Total Environ. 470 (2), 127–137.
Di, Q., Kloog, I., Koutrakis, P., Lyapustin, A., Wang, Y., Schwartz, J., 2016. Assessing $PM_{2.5}$ exposures with high spatiotemporal resolution across the continental United States. Environ. Sci. Technol. 50 (9), 4712–4721.
Evgeniou, T., Pontil, M., 2004. Regularized multi–task learning. Knowledge Discovery and Data Mining (KDD)ACM, pp. 109–117.
Gong, B., Ordieres, M.J., 2016. Prediction of daily maximum ozone threshold exceedances by preprocessing and ensemble artificial intelligence techniques: case study of Hong Kong. Environ. Model. Softw. 84, 290–303.
Hrust, L., Klaic, Z.B., Krizan, J., Antonic, O., Hercog, P., 2009. Neural network forecasting of air pollutants hourly concentrations using optimised temporal averages of meteorological variables and pollutant concentrations. Atmos. Environ. 43, 5588–5596.
Huang, Y., Shen, H., Chen, H., Wang, R., Zhang, Y., Su, S., et al., 2014. Quantification of global primary emissions of $PM_{2.5}$, PM_{10}, and TSP from combustion and industrial process sources. Environ. Sci. Technol. 48 (23), 13834–13843.
Jiang, P., Dong, Q., Li, P., 2017. A novel hybrid strategy for $PM_{2.5}$ concentration analysis and prediction. J. Environ. Manag. 196, 443–457.
Li, C., Zhu, Z., 2018. Research and application of a novel hybrid air quality early-warning system: a case study in China. Sci. Total Environ. 626, 1421–1438.
Li, X., Peng, L., Hu, Y., Shao, J., Chi, T., 2016. Deep learning architecture for air quality predictions. Environ. Sci. Pollut. Res. 23 (22), 22408–22417.
Li, N., Chen, J.P., Tsai, C.I., et al., 2017. Potential impacts of electric vehicles on air quality in Taiwan. Sci. Total Environ. 566, 919–928.
Liao, Z., Gao, M., Sun, J., Fan, S., 2017. The impact of synoptic circulation on air quality and pollution-related human health in the Yangtze River Delta region. Sci. Total Environ. 607, 838–846.
Liu, A.A., Xu, N., Su, Y.T., et al., 2015. Single/multi-view human action recognition via regularized multi-task learning. Neurocomputing 151 (10), 544–553.
Lv, B., Cobourn, W.G., Bai, Y., 2016. Development of nonlinear empirical models to forecast daily $PM_{2.5}$ and ozone levels in three large Chinese cities. Atmos. Environ. 147, 209–223.
Maidment, D.R., 1993. Handbook of Hydrology. McGraw-Hill, New York.
Méheust, D., Gangneux, J.P., Reponen, T., Wymer, L., Vesper, S., Le Cann, P., 2012. Correlation between environmental relative moldiness index (ERMI) values in French dwellings and other measures of fungal contamination. Sci. Total Environ. 438, 319–324.
Nguyen, V.A., Starzyk, J.A., Goh, W.B., Jachyra, D., 2012. Neural network structure for spatio-temporal long-term memory. IEEE Trans. Neural Netw. Learn. Syst. 23 (6), 971–983.
Nieto, P.J.G., Lasheras, F.S., García-Gonzalo, E., Juez, F.J.D.C., 2018. PM_{10} concentration forecasting in the metropolitan area of Oviedo (Northern Spain) using models based on SVM, MLP, VARMA and ARIMA: a case study. Sci. Total Environ. 621, 753–761.
Oprea, M., Mihalache, S.F., Popescu, M., 2016. A comparative study of computational intelligence techniques applied to $PM_{2.5}$ air pollution forecasting. International Conference on Computers Communications and Control. IEEE, pp. 103–108.
Ping, W., Yong, L., Qin, Z., Zhang, G., 2015. A novel hybrid forecasting model for PM_{10}, and SO_2, daily concentrations. Sci. Total Environ. 505, 1202–1212.
Prasad, K., Gorai, A.K., Goyal, P., 2016. Development of ANFIS models for air quality forecasting and input optimization for reducing the computational cost and time. Atmos. Environ. 128 (1), 246–262.
Shireen, T., Shao, C., Wang, H., Li, J., Zhang, X., Li, M., 2018. Iterative multi-task learning for time-series modeling of solar panel PV outputs. Appl. Energy 212, 654–662.
Tong, L., Lau, A.K.H., Kai, S., Fung, J.C.H., 2018. Time series forecasting of air quality based on regional numerical modeling in Hong Kong. J. Geophys. Res. Atmos. 123 (8), 4175–4196.
Voukantsis, D., Karatzas, K., Kukkonen, J., Räsänen, T., Karppinen, A., Kolehmainen, M., 2011. Intercomparison of air quality data using principal component analysis, and forecasting of PM_{10} and $PM_{2.5}$ concentrations using artificial neural networks, in Thessaloniki and Helsinki. Sci. Total Environ. 409 (7), 1266–1276.
Wang, Y., Sun, M., Yang, X., Yuan, X., 2016. Public awareness and willingness to pay for tackling smog pollution in China: a case study. J. Clean. Prod. 112, 1627–1634.
Witkowska, A., Lewandowska, A.U., 2016. Water soluble organic carbon in aerosols (PM_1, $PM_{2.5}$, PM_{10}) and various precipitation forms (rain, snow, mixed) over the southern Baltic Sea station. Sci. Total Environ. 573, 337–346.
Xu, Y., Yang, W., Wang, J., 2016. Air quality early-warning system for cities in China. Atmos. Environ. 148, 239–257.
Xu, Y.L., Li, X.X., Chen, D.R., et al., 2018. Learning rates of regularized regression with multiple Gaussian kernels for multi-task learning. IEEE Trans. Neural Netw. Learn. Syst. 99, 1–11.
Yamane, I., Sasaki, H., Sugiyama, M., 2015. Regularized multitask learning for multidimensional log-density gradient estimation. Neural Comput. 28 (7), 1–23.
Yang, G., Huang, J., Li, X., 2018. Mining sequential patterns of $PM_{2.5}$ pollution in three zones in China. J. Clean. Prod. 170, 388–398.
Yeganeh, B., Hewson, M.G., Clifford, S., Tavassoli, A., Knibbs, L.D., Morawska, L., 2018. Estimating the spatiotemporal variation of NO_2 concentration using an adaptive neuro-fuzzy inference system. Environ. Model. Softw. 100, 222–235.
Yu, H., Stuart, A.L., 2017. Impacts of compact growth and electric vehicles on future air quality and urban exposures may be mixed. Sci. Total Environ. 576, 148–158.
Zhai, B., Chen, J., 2018. Development of a stacked ensemble model for forecasting and analyzing daily average $PM_{2.5}$ concentrations in Beijing, China. Sci. Total Environ. 635, 644–658.
Zhai, L., Li, S., Zou, B., Sang, H., Fang, X., Xu, S., 2018. An improved geographically weighted regression model for $PM_{2.5}$ concentration estimation in large areas. Atmos. Environ. 181, 145–154.
Zhang, C., Tao, D., Hu, T., Li, X., 2014. Task-group relatedness and generalization bounds for regularized multi-task learning. Comput. Sci. 114, 177–183.
Zhang, Y., Lang, J., Cheng, S., Li, S., Zhou, Y., Chen, D., Wang, H., 2018. Chemical composition and sources of PM_1 and $PM_{2.5}$ in Beijing in autumn. Sci. Total Environ. 630, 72–82.
Zhao, L., Sun, Q., Ye, J., Chen, F., Lu, C.T., Ramakrishnan, N., 2017. Feature constrained multi-task learning models for spatiotemporal event forecasting. IEEE Trans. Knowl. Data Eng. 29 (5), 1059–1072.
Zhu, X., Suk, H.I., Lee, S.W., Shen, D., 2016. Subspace regularized sparse multi-task learning for multi-class neurodegenerative disease identification. IEEE Trans. Biomed. Eng. 63 (3), 607–618.
Zhu, S., Lian, X., Wei, L., Che, J., Shen, X., Yang, L., Li, J., 2018. $PM_{2.5}$ forecasting using SVR with PSOGSA algorithm based on CEEMD, GRNN and GCA considering meteorological factors. Atmos. Environ. 183, 20–32.