KALMAN FILTER BASED PREDICTION SYSTEM FOR WINTERTIME PM$_{10}$ CONCENTRATIONS IN MACAU

K.I. HOI
K.V. YUEN
K.M. MOK*

Department of Civil and Environmental Engineering
University of Macau
Av. Padre Tomás Pereira S.J., Taipa, Macau SAR, China

*to whom all correspondence should be addressed:
e-mail: KMMok@umac.mo

ABSTRACT

In the present study, the Kalman filter algorithm was applied to forecast the wintertime PM$_{10}$ concentrations of Macau. The algorithm was implemented on an AR(2) model and an AREX model, respectively. The AR(2) model is essentially an autoregressive model of order 2, i.e., the daily averaged PM$_{10}$ concentration tomorrow is predicted by a linear combination of the PM$_{10}$ concentrations in the previous two days. The AREX model is built based on the AR(2) model. It is a combination of the autoregressive model and the exogenous inputs such as the wind speed and the wind direction on the day of prediction. Both models were tested by using the PM$_{10}$ concentrations and the meteorological data between November of 2004 and February of 2005. It was found that the mean absolute prediction error percentage of the AR(2) model was 36.36%, with an RMS error of 34.94 μg m^{-3}. The Pearson correlation coefficient between the predictions and the measurements is 0.59. Time-delay problem was associated with the AR(2) model, i.e., the trend of the predicted PM$_{10}$ concentrations generally lagged behind the trend of the measurements. On the other hand, the error percentage of the AREX model was 32.45%, with an RMS error of 27.08 μg m^{-3}. The Pearson correlation coefficient is 0.75. The time-delay problem was improved and the trend of the predictions was in good agreement with the measurements. The AREX model outperformed the AR(2) model since the meteorological conditions could reflect the dispersion condition and the nature of the replenishing air masses on the day of prediction. It was concluded that the Kalman filter was promising in the air quality prediction but caution should be made in the selection of the model classes.

KEYWORDS: Air quality prediction, Kalman Filter, Macau, PM$_{10}$.

INTRODUCTION

Nowadays, the degradation of air quality is widely observed in most urban areas due to rapid population growth and economic development. As more goods and services are produced for human satisfaction, energy is consumed from the fuel combustion processes. At the same time, air pollutants such as particulates, sulfur dioxide, nitrogen dioxide or carbon monoxide are generated and released to the atmosphere. When the air pollutants enter the human body, they may cause adverse health effects such as bronchitis, asthma, and heart disease, etc (Curtis *et al.*, 2006; Dockery *et al.*, 1993). According to the statistics provided by the World Health Organization (WHO), three million people face terminal health problems worldwide by the outdoor air pollution annually from vehicles and industrial emissions. With such relevance to human health, it necessitates the development of the air quality prediction system since accurate air quality forecast can help people to reduce/cancel outdoor activities on the days of high pollutant concentrations. Meanwhile, more flexible and realistic emission control strategies can be implemented if the future air quality is known in advance. Currently, many cities in Canada and the U.S. have regulations in place to curtail industrial and other activities during times of observed poor air quality. In Canada, the Alberta Environment asks industry to cut production during morning hours of poor air quality in Calgary and Edmonton. In the United

States, an ozone action day is called by the state or local air quality agency when ozone levels are forecasted to reach unhealthy levels. During ozone action days, the state or local authority asks the community to take simple voluntary actions to help reducing ground level ozone and hence preventing violation of the National Ambient Air Quality Standards (NAAQS) (Ojha *et al.*, 2002). Therefore, the development of an air quality prediction system plays an important role in fulfilling the aforementioned objectives.

Macau is a coastal city on the West side of the Pearl River Delta Region as shown in Figure 1. It has a geographical area of 28.6 km^2 and a population of 510,000. Table 1 shows the statistics of daily PM_{10} concentrations between 2000 and 2004. It is noted that the annual PM_{10} concentration generally increases during the studied period and the average increase, which is obtained from the slope of the best fit straight line across the data points, is 3.91 $\mu g\ m^{-3}\ yr^{-1}$. However, the maxima and the minima of the daily PM_{10} concentration do not follow the same trend of the annual PM_{10} concentration. Nevertheless, there are also increases in the maxima and minima of the daily averaged PM_{10} concentrations between 2000 and 2004. The average increases are 2.45 $\mu g\ m^{-3}\ yr^{-1}$ and 0.72 $\mu g\ m^{-3}\ yr^{-1}$ for the maxima and the minima, respectively. Therefore, it is concluded that the situation of the PM_{10} pollution in Macau is worsening during the studied period. Figure 2 shows the percentage of days with bad air quality between 2000 and 2004. The air quality is characterized as bad when the daily air quality index is larger than 100.

Figure 1. Geographical Location of Macau in the Pearl River Delta Region (Google Maps, 2007)

Table 1. Statistics of daily averaged PM_{10} concentrations between 2000 and 2004

Year	Mean ($\mu g\ m^{-3}$)	Maxima ($\mu g\ m^{-3}$)	Minima ($\mu g\ m^{-3}$)
2000	48.5	172.1	7.4
2001	57.2	227.0	6.4
2002	50.4	203.5	8.8
2003	59.7	201.5	9.8
2004	66.8	197.1	9.3

It is noted that the number of days with bad air quality generally increases during this period. The peak occurs in 2004 and it corresponds to 8.83% of the whole year. The steep rise in the number of PM_{10} episode days between 2002 and 2004 coincides with the booming of the Macau economy as there is a 45.4% increase in the GDP per capita of Macau from USD15,567 to USD22,634 (Macau Statistics and Census Bureau, 2008). The increase in the economic activities may lead to increased energy consumption and the subsequent increase in the air pollution. Therefore, the problem of air quality deterioration is severe and has raised our concern. To tackle with the current circumstances, an initial step is to develop an air quality prediction system so that citizens can have chances to minimize the time of exposure to the air pollutants on the air quality episode days. It is also the motivation of the present study.
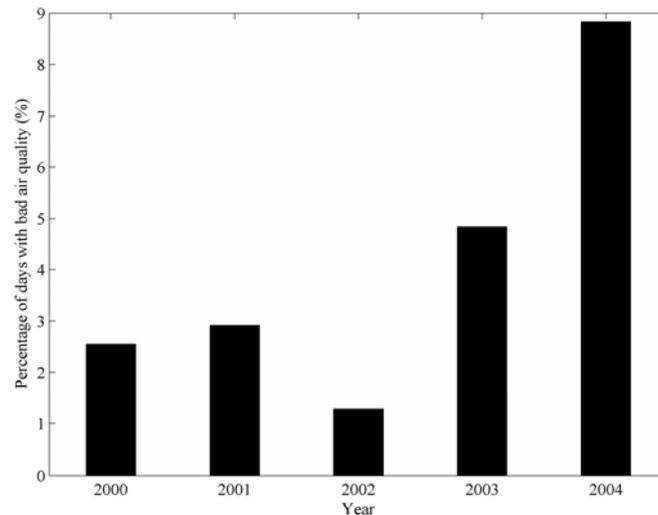
Figure 2. Percentage of Days with Bad Quality between 2000 and 2004

In this study, the Kalman filter algorithm is applied. Kalman filter was originally formulated by Kalman and Bucy for the prediction and filtering of random signals for linear systems (Kalman, 1960; Kalman & Bucy, 1961). Thereafter, the extended Kalman filter was applied to system identification for linear or slightly non-linear dynamic systems, in addition to its state estimation capability, in many disciplines of science and engineering. Kalman filter becomes popular since the algorithm not only provides the response prediction and the estimation of model parameters but also their associated uncertainties. Furthermore, it is operated in an online manner, i.e., the model is updated once a new data point is obtained. Successful application of the Kalman filter in air quality prediction can be found in several literatures (Anh *et al.*, 1998; Ng & Yan, 2004; van der Wal & Jansen, 2000; Zolghadri & Cazaurang, 2006). Therefore, the objective of the present study is to test the applicability of Kalman filter to predict the daily averaged PM_{10} concentrations of Macau. PM_{10} is a collective name used to describe the particles with an aerodynamic size of not greater than 10 μm. It is chosen to be the target species for prediction since it is one of the dominant air pollutants in Macau especially during the winter period. Figure 3 shows the time series of the daily averaged PM_{10} concentrations between 2000 and 2004. It is noted that there is a distinct seasonal pattern, i.e., high PM_{10} concentrations are generally observed in the winter period. In comparison to the other seasons, there are more days that the daily averaged PM_{10} concentrations are larger than 100 μg m^{-3}. The concentration corresponds to an air quality index of 100, which indicates a day of bad air quality.
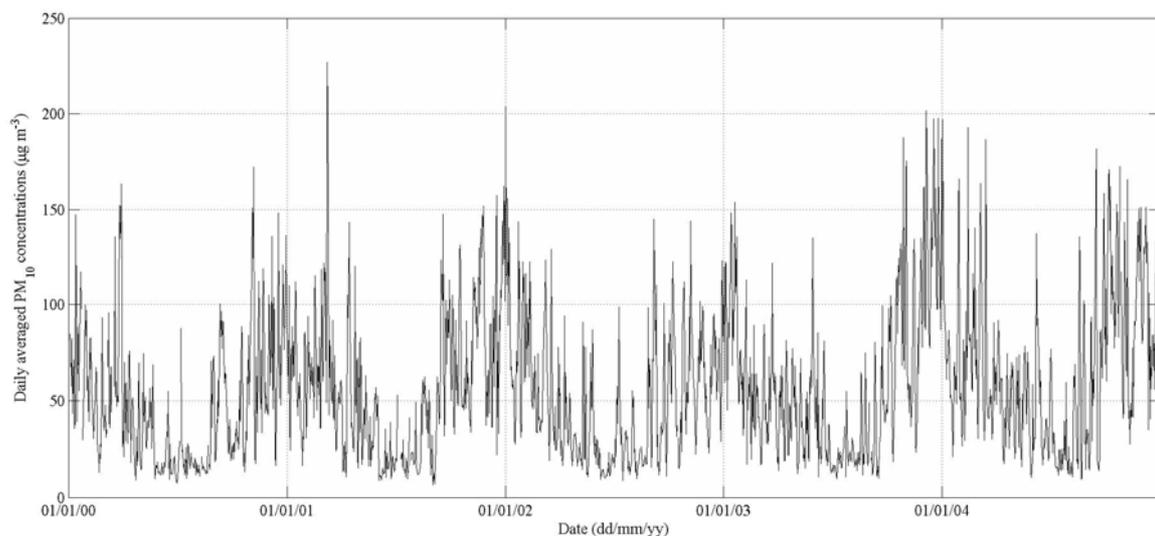
Figure 3. Time Series of Daily Averaged PM_{10} Concentrations between 2000 and 2004

Mok and Hoi (2005) found that the seasonal behavior was associated with swing of the prevailing wind directions caused by the Asian monsoon climates. The monsoon driven winter north-easterly winds bring upon Macau dry and particle enriched air masses leading to a higher concentration in that period while the summer south-westerly winds transport humid and cleaner air from South China Sea to the region leading to a lower PM_{10} value (Mok & Hoi, 2005). Since the problem of PM_{10} pollution in Macau is more significant during winter, the development of the air quality prediction system in the present study is focused on this season. In this study, the Kalman filter is implemented on two different models, namely an AR(2) model and an AREX model. Both models are tested by using the daily averaged PM_{10} concentrations and the meteorological data provided from Macau Meteorological and Geophysical Bureau. The PM_{10} data were recorded by using the Tapered Element Oscillating Microbalance (TEOM) at the ambient monitoring station of Macau between November of 2004 and February of 2005. Figure 4 shows the surrounding environment of the monitoring station. The station is located at the Taipa Grande Hill which has an altitude of 158.2 m. Therefore, its measurements are representative to indicate the general ambient air quality of Macau. In the following section, the formulation of the model classes and the Kalman filter is briefly described.

Figure 4. General Ambient Monitoring Station and its Surrounding Environment (Macau Meteorological and Geophysical Bureau, 2007)

FORMULATION
AR(2) model
In this section, the Kalman filter is formulated for the autoregressive model of order 2 with time varying coefficients, which is abbreviated as the AR(2) model:

$$x_k = \phi_{1,k-1} x_{k-1} + \phi_{2,k-1} x_{k-2} + f_{k-1} \tag{1}$$

where x_k denotes the daily averaged PM_{10} concentration of the kth day. The input f represents the modeling error, and it is modeled as Gaussian i.i.d. with zero mean and variance σ_f^2. It stands for the neglected factors that influence the PM_{10} concentration on the day of prediction. In addition, it is assumed that the measurement of [PM_{10}], denoted as z_k, is contaminated during the measurement process. The relationship between z_k and the x_k is given as follows:

$$z_k = x_k + n_k \tag{2}$$

where n denotes the measurement noise and it is also modeled as Gaussian i.i.d. with zero mean and variance σ_n^2. The AR(2) model simply means that the PM_{10} concentration of the kth day is a weighted sum of its concentrations of the previous two days and the weights are represented by the time-varying coefficients ϕ_i to be identified. Those unknown coefficients evolve according to the following equation:

$$\phi_{i,k} = \phi_{i,k-1} + w_{i,k-1} \tag{3}$$

where $w_{i,k-1}$ denotes the perturbation to the coefficient $\phi_{i,k-1}$ at the $(k-1)^{th}$ time step and it is one of the components in the random process w_i. The process is modeled as Gaussian i.i.d. with zero mean and variance σ_w^2. Now we define an augmented state vector $\mathbf{Y}_k$ which contains the PM_{10} concentrations of different days and the unknown coefficients to be estimated:

$$\mathbf{Y}_k = \begin{bmatrix} x_{k-1}, x_k, \phi_{1,k}, \phi_{2,k} \end{bmatrix}^T \qquad (4)$$

Also, we define a process noise vector $\mathbf{F}_k$ which contains the process noises f_k and $w_{i,k}$, $i=1,2$, as follows:

$$\mathbf{F}_k = \begin{bmatrix} f_k, w_{1,k}, w_{2,k} \end{bmatrix}^T \qquad (5)$$

Then, one can linearize the state space equation for the AR(2) model:

$$\mathbf{Y}_k = \mathbf{A}_{k-1}\mathbf{Y}_{k-1} + \mathbf{B}\mathbf{F}_{k-1} + \mathbf{G}_{k-1} \qquad (6)$$

where $\mathbf{A}_{k-1}$ and $\mathbf{B}$ are the transition matrices and the vector $\mathbf{G}_{k-1}$ is to compensate the linearization error. It is noted that $\mathbf{B}$ is constant for any time step and the terms $\mathbf{A}_{k-1}$ and $\mathbf{G}_{k-1}$ are evaluated at $\bar{\mathbf{e}}_{\mathbf{Y},k-1}$, which represents the updated state estimate of $\mathbf{Y}_{k-1}$ when the measurements $\mathbf{D}_{k-1}=\{z_1,z_2,\ldots,z_{k-1}\}$ are available.

$$\mathbf{A}_{k-1} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ \phi_{2,k-1} & \phi_{1,k-1} & x_{k-1} & x_{k-2} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}_{\mathbf{Y}_{k-1}=\bar{\mathbf{e}}_{\mathbf{Y},k-1}} \qquad (7)$$

$$\mathbf{B} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad (8)$$

$$\mathbf{G}_{k-1} = \begin{bmatrix} 0 \\ -\phi_{2,k-1}x_{k-2} - \phi_{1,k-1}x_{k-1} \\ 0 \\ 0 \end{bmatrix}_{\mathbf{Y}_{k-1}=\bar{\mathbf{e}}_{\mathbf{Y},k-1}} \qquad (9)$$

With this state space model, one can perform the prediction and filtering of the PM_{10} concentrations by using the Kalman filter. The essential steps of the Kalman filter are to pre-update and update alternatively for each data point of the dataset $\mathbf{D}_N$. When the measured PM_{10} concentrations up to the $(k-1)^{th}$ day $\mathbf{D}_{k-1}$ are available, the pre-updating procedure is applied to predict the PM_{10} concentration in the next time step by using the conditional probability density function $p(\mathbf{Y}_k|\mathbf{D}_{k-1})$. In the present study, the conditional PDF is assumed to be multi-variate Gaussian. By using Eqn. (6), the pre-updated state vector on the kth day can be estimated from the updated state on the $(k-1)^{th}$ day:

$$\mathbf{e}_{\mathbf{Y},k} \equiv E\begin{bmatrix} \mathbf{Y}_k \mid \mathbf{D}_{k-1} \end{bmatrix} = \begin{bmatrix} x_{k-1} \\ \phi_{2,k-1}x_{k-2} + \phi_{1,k-1}x_{k-1} \\ \phi_{2,k-1} \\ \phi_{1,k-1} \end{bmatrix}_{\mathbf{Y}_{k-1}=\bar{\mathbf{e}}_{\mathbf{Y},k-1}} \qquad (10)$$

In addition, the uncertainty of the air quality prediction and the estimated model parameters are represented by the covariance matrix:

$$\boldsymbol{\Sigma}_{\mathbf{Y},k} \equiv E\begin{bmatrix} (\mathbf{Y}_k - \mathbf{e}_{\mathbf{Y},k})(\mathbf{Y}_k - \mathbf{e}_{\mathbf{Y},k})^T \mid \mathbf{D}_{k-1} \end{bmatrix} = \mathbf{A}_{k-1}\overline{\boldsymbol{\Sigma}}_{\mathbf{Y},k-1}\mathbf{A}_{k-1}^T + \mathbf{B}\boldsymbol{\Sigma}_{\mathbf{F}}\mathbf{B}^T \qquad (11)$$

where $\boldsymbol{\Sigma}_{\mathbf{F}}$ is the covariance matrix of the process noise vector:

$$\boldsymbol{\Sigma}_{\mathbf{F}} \equiv E\left[\mathbf{F}_k \mathbf{F}_k^T\right] = \begin{bmatrix} \sigma_f^2 & 0 & 0 \\ 0 & \sigma_w^2 & 0 \\ 0 & 0 & \sigma_w^2 \end{bmatrix} \tag{12}$$

When the measurement on the kth day is available, the PM_{10} concentration and the model parameters are updated by finding the updated state vector $\mathbf{Y}_k$. This is accomplished by maximizing the conditional PDF $p(\mathbf{Y}_k|\mathbf{D}_k)$:

$$\overline{\mathbf{e}}_{\mathbf{Y},k} \equiv E\left[\mathbf{Y}_k \mid \mathbf{D}_k\right] = \overline{\boldsymbol{\Sigma}}_{\mathbf{Y},k}\left(\boldsymbol{\Sigma}_{\mathbf{Y},k}^{-1}\mathbf{e}_{\mathbf{Y},k} + \sigma_n^{-2}\mathbf{C}^T z_k\right) \tag{13}$$

where $\mathbf{C}$ is the observation matrix which relates z_k and $\mathbf{Y}_k$ in the following form:

$$z_k = \mathbf{C}\mathbf{Y}_k + n_k, \quad \mathbf{C} = \begin{bmatrix} 0 & 1 & 0 & 0 \end{bmatrix} \tag{14}$$

The uncertainty of the state estimation is represented by its covariance matrix:

$$\overline{\boldsymbol{\Sigma}}_{\mathbf{Y},k} \equiv E\left[\left(\mathbf{Y}_k - \overline{\mathbf{e}}_{\mathbf{Y},k}\right)\left(\mathbf{Y}_k - \overline{\mathbf{e}}_{\mathbf{Y},k}\right)^T \mid \mathbf{D}_k\right] = \left(\boldsymbol{\Sigma}_{\mathbf{Y},k}^{-1} + \sigma_n^{-2}\mathbf{C}^T\mathbf{C}\right)^{-1} \tag{15}$$

AREX model

In this section, the Kalman filter is implemented on a time varying autoregressive model with exogenous inputs, which is abbreviated as the AREX model:

$$x_k = \phi_{1,k-1}x_k + \phi_{2,k-1}\exp\left(-\phi_{3,k-1}u_k\right) + \phi_{4,k-1}\left[1 - \exp\left(-\phi_{3,k-1}u_k\right)\right] + \phi_{5,k-1}\theta_k + f_k \tag{16}$$

where u_k and θ_k denote the wind speed and wind direction at the kth day, respectively. The model is different from the previous one since the prediction is not solely made based on its past history. On the contrary, some physical quantities that can influence the PM_{10} concentration on the day of prediction are also taken into account. Those physical quantities are reflected from the second, third and fourth term in Eqn. (16). The second term represents the contribution of PM_{10} concentration from the local sources. The wind speed inside this term indicates the dispersion condition. Usually low wind speed is associated with poor dispersion condition. When the wind speed becomes low, the atmosphere becomes stagnant. The pollutants are trapped inside the boundary layer and cannot be transported away readily from local to downwind areas through advection. Therefore, the second term means that the contribution from the local sources decreases exponentially with the wind speed. The effect of wind speed is assumed to lead to have an exponential decay of the concentration since the mass flux of the air pollutant being transported away is enhanced with increasing wind speed. At the same time, the increasing wind speed is associated with increased turbulence intensity, which in turns enhances the vertical mixing of the atmosphere. Therefore, a nonlinear relationship between the concentration and the wind speed is assumed here. In fact, a similar power law relationship between the pollutant concentrations and the wind speed was adopted in previous studies (Shi & Harrison, 1995; Goyal *et al.*, 2006). The coefficient ϕ_2 denotes the local source strength and the coefficient ϕ_3 controls the decaying speed. On the other hand, the third term and the fourth term in Eqn. (16) represent the contribution of PM_{10} concentration from the regional sources. The third term means that the regional contribution increases with the wind speed until it reaches certain limit. After the wind speed passes that limit, the regional contribution will be insensitive. The coefficient ϕ_4 denotes the strength of the regional sources. The fourth term represents the effect of wind direction on the PM_{10} concentration. The wind direction is a crucial factor in determining the sources of air pollution being brought to Macau. In fact, previous studies have shown that the change of wind direction can affect the pollutant concentrations of the Pearl River Delta region (Lee & Hills, 2003; Mok & Hoi, 2003; Chang *et al.*, 2007). In addition, this variable was also included as one of the input variables in the air quality prediction models of previous studies applied to other cities (Shi & Harrison, 1995; Goyal *et al.*, 2006). The wind direction is defined here as the direction from which the wind is blowing measured clockwise from the Geographic True North. For example, 0° represents the North direction whereas 90° is the East direction. The input variable θ used in this study is the

number of wind sector, which represents a particular range of wind direction. The number ranges from 0 and 8 and each number contains a 40° range of wind direction. For example, the number '0' represents the range between -20° and 20°. Therefore, the entire range of wind directions (-180° to 180°) can be classified into 9 sectors. The number of wind sector is used here to indicate the type of replenishing air masses being transported to Macau on the day of prediction. As the number of wind sector approaches from 0 to 8, the direction of surface wind blowing to Macau changes gradually from northerly to southerly. As Macau is located at the southwest of the Pearl River Delta Region which is composed of several fast developing cities such as Guangzhou and Shenzhen in the recent decades, the air pollution problem in this region is severe. Particulates can be easily transported to Macau from its upwind cities nearby when the prevailing winds are blowing from the northerly directions. On the contrary, Macau is facing the South China Sea located to its South direction. When the winds are blowing from the South China Sea, Macau is less influenced by the transboundary air pollution from the cities located to its North. Therefore, the fourth term can be used to counterbalance the effect of the second term and the third term when the winds are blowing from the sea. Finally, the term f_{k-1} represents the modeling error and it is modeled as Gaussian i.i.d. with zero mean and variance σ_f^2.

Now we define an augmented state vector $\mathbf{Y}_k$ which contains the PM$_{10}$ concentration yesterday and the unknown coefficients to be estimated:

$$\mathbf{Y}_k = \left[x_k, \phi_{1,k}, \phi_{2,k}, \phi_{3,k}, \phi_{4,k}, \phi_{5,k} \right]^T \tag{17}$$

Also, we define a process noise vector $\mathbf{F}_k$ which contains the process noises f_k and $w_{i,k}$ as follows:

$$\mathbf{F}_k = \left[f_k, w_{1,k}, w_{2,k}, w_{3,k}, w_{4,k}, w_{5,k} \right]^T \tag{18}$$

The covariance matrix of the process noise vector is shown below:

$$\Sigma_{\mathbf{F}} = \begin{bmatrix} \sigma_f^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_w^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_w^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_w^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_w^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_w^2 \end{bmatrix} \tag{19}$$

Then, one can linearize the state space equation for the AREX model. The transition matrices $\mathbf{A}_{k-1}$, $\mathbf{B}$ and the compensation vector $\mathbf{G}_{k-1}$ are given below:

$$\mathbf{A}_{k-1} = \begin{bmatrix} \phi_{1,k-1} & x_{k-1} & e^{-\phi_{3,k-1}u_k} & u_k\left(\phi_{4,k-1} - \phi_{2,k-1}\right)e^{-\phi_{3,k-1}u_k} & 1 - e^{-\phi_{3,k-1}u_k} & \theta_k \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}_{\mathbf{Y}_{k-1} = \bar{\mathbf{e}}_{\mathbf{Y},k-1}} \tag{20}$$

$$\mathbf{B} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{21}$$

$$\mathbf{G}_{k-1} = \left[\begin{array}{c} -\phi_{1,k-1}x_{k-1} + u_k\left(\phi_{2,k-1} - \phi_{4,k-1}\right)\phi_{3,k-1}e^{-\phi_{3,k-1}u_k} \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{array}\right]_{\mathbf{Y}_{k-1}=\overline{\mathbf{e}}_{\mathbf{Y},k-1}} \qquad (22)$$

Then the Kalman filter is operated by applying the pre-updating procedure and the updating procedure sequentially on each measurement z_k. By using Eqn. (6), the pre-updated state vector on the kth day can be estimated from the updated state on the $(k-1)^{th}$ day:

$$\mathbf{e}_{\mathbf{Y},k} = \left[\begin{array}{c} \phi_{1,k-1}x_k + \phi_{2,k-1}\exp\left(-\phi_{3,k-1}u_k\right) + \phi_{4,k-1}\left[1 - \exp\left(-\phi_{3,k-1}u_k\right)\right] + \phi_{5,k-1}\theta_k \\ \phi_{1,k-1} \\ \phi_{2,k-1} \\ \phi_{3,k-1} \\ \phi_{4,k-1} \\ \phi_{5,k-1} \end{array}\right]_{\mathbf{Y}_{k-1}=\overline{\mathbf{e}}_{\mathbf{Y},k-1}} \qquad (23)$$

The covariance matrix of the pre-updated state vector is calculated by using Eqn. (11). The updated state vector and the corresponding covariance matrix can be found by applying Eqn. (13) and Eqn. (15). However, caution should be made in applying Eqn. (15) since the observation matrix **C** is different from the previous one and the revised observation matrix is shown below:

$$\mathbf{C} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \qquad (24)$$

In the following section, both models are tested by using the daily averaged PM_{10} concentrations and the meteorological data provided from Macau Meteorological and Geophysical Bureau. The supplied data were recorded at the ambient monitoring station of Macau between November of 2004 and February of 2005. Figure 3 shows the surrounding environment of the monitoring station. The station is located at the Taipa Grande Hill which has an altitude of 158.2m. Therefore, its measurements are representative to indicate the general ambient air quality of Macau.

RESULTS
The prediction result obtained from the AR(2) model is shown in Figure 5a. The solid line shows the measured daily averaged PM_{10} concentration, whereas the dotted line shows the result of prediction. It is noted that there is a time-delay problem associated with the AR(2) model, i.e., the trend of the predictions generally lags behind the trend of the measurements. The problem appears since the prediction is made based on its own past history. Those influencing factors such as the dispersion condition on the day of prediction and the nature of replenishing air masses which can be continental or oceanic are treated as the process noises. Therefore, the process noise is large relative to the RMS of the signal and the large noise-to-signal ratio causes the predicted signal to be delayed. In order to quantify the error of prediction, three performance measures, namely the mean absolute percentage error (MAPE), the root mean squared error (RMSE) and the Pearson correlation coefficient r are introduced as follows:

$$MAPE = \frac{1}{N}\sum_{k=1}^{N}\frac{|z_k - x_k|}{z_k} \times 100\% \qquad (25)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{k=1}^{N}(z_k - x_k)^2}$$

(26)

$$r = \frac{\sum_{k=1}^{N}(z_k - \mu_z)(x_k - \mu_x)}{\sqrt{\sum_{k=1}^{N}(z_k - \mu_z)^2}\sqrt{\sum_{k=1}^{N}(x_k - \mu_x)^2}}$$

(27)

where z_k, x_k, and N are the measurement, the prediction, and the number of samples in the dataset, respectively. The symbols μ_z and μ_x denote the average values of the measurements and predictions, respectively. Table 2 shows the results on the model validation of the AR(2) model and the AREX model. The MAPE, the RMSE and the r value for the AR(2) model are 36.36%, 34.94 μg m^{-3}, and 0.59, respectively. Figure 5b and 5c show the corresponding contributions to the predicted PM$_{10}$ concentrations from the first term and the second term in Eqn. (1).

Table 2. Results on the model validation of the AR(2) model and the AREX model

Model class	MAPE (%)	RMSE (μg m^{-3})	r
AR(2)	36.36	34.94	0.59
AREX	32.45	27.08	0.75
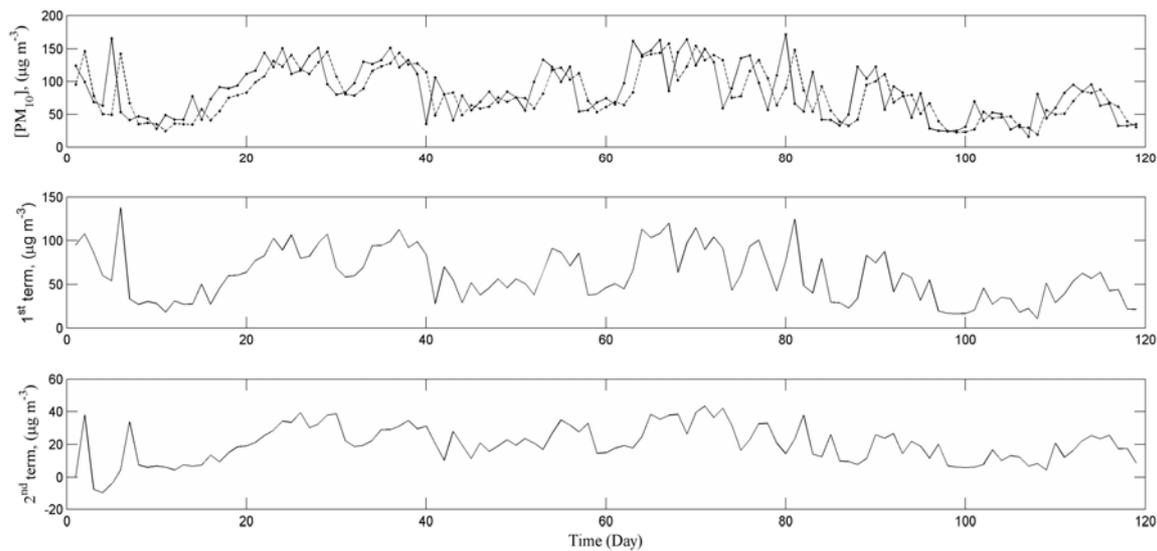
Figure 5. (a) Time History of Measured Daily PM$_{10}$ Concentrations (solid line) and its Prediction (dotted line) from AR(2) Model; (b-c) Time Histories of Contributions to the Predicted PM$_{10}$ Concentrations from the 1st Term $\phi_{1,k-1}x_{k-1}$ and the 2nd Term $\phi_{2,k-1}x_{k-2}$.

It is noted that the contribution from the first term is significantly larger than the contribution from the second term. Figure 6a shows the prediction result obtained from the AREX model. The solid line shows the measured daily average of PM$_{10}$ concentration, whereas the dotted line shows the prediction result. It is noted that the time-delay problem observed previously in the AR(2) model has been improved. The MAPE, the RMSE, and the r value for the AREX model are 32.45%, 27.08 μg m^{-3}, and 0.75, respectively. The percentage decreases of MAPE and RMSE are 10.75% and 22.50%, respectively. The percentage increase of the r value is 27.12%. The improvement in the prediction accuracy and the problem of time-delay can be explained by investigating the contributions to the predicted PM$_{10}$ concentrations from different terms in Eqn. (16) as shown in Figures 6b-6e. It is noted that the predicted PM$_{10}$ concentration does not only rely on the contribution from its past concentration since the contribution from the first term as shown in Figure 1b is reduced. On the contrary, the second term and the fourth

term of the equation have significant contribution to the prediction. The second term has positive contribution to the predicted PM_{10} concentration since the wind speed can somehow reflect the dispersion condition on the day of prediction. When the wind speed becomes low, the atmosphere becomes stagnant. The pollutants are trapped inside the boundary layer and cannot be transported away readily from local to downwind areas through advection. Therefore, relatively high contributions are generally associated with low wind speed. On the other hand, the fourth term has negative contribution to the predicted concentration since the number of wind sector can reflect the nature of replenishing air masses on the day of prediction. As the number of wind sector approaches from 0 to 8, the source of replenishing air changes from the continent of China to the South China Sea. Therefore, negative contributions are generally associated with large numbers of wind sectors. However, it is surprising that the third term has little contribution to the predicted PM_{10} concentration. The third term represents the transport of PM_{10} from regional sources. It is supposed that large positive contribution to the prediction should be associated with high
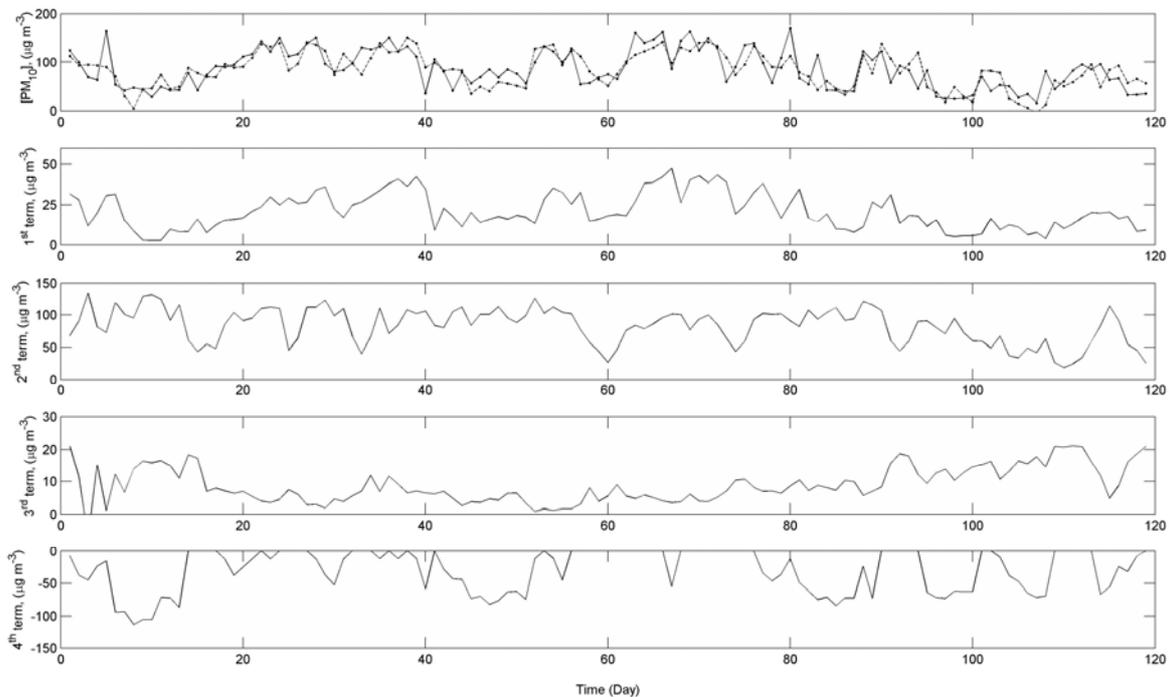
Figure 6. (a) Time History of Measured Daily PM_{10} concentrations (solid line) and its Prediction (dotted line) from AREX Model; (b-e) Time Histories of Contributions to the Predicted PM_{10} Concentrations from the 1st Term to the 4th Term.

wind speed and northerly wind direction. However, the result does not fit with the assumption. Therefore, the regional sources causing high PM_{10} concentrations of Macau during the test period may be near Macau and even low wind speed is enough to transport the pollutants to Macau when the winds are blowing from the continent. More fundamental understanding of the phenomenon causing high PM_{10} concentration is necessary in order to modify the model class and achieve higher prediction accuracy. Concluding from those findings, the prediction accuracy and the time-delay problem are improved since the information of wind speed and wind direction make the model class more representative hence reducing the process noise of the system.

CONCLUSIONS

The extended Kalman filter was applied to forecast the daily averaged PM_{10} concentrations based on an AR(2) model and an AREX model. Both models were tested with the measured PM_{10} concentrations and meteorological data between November of 2004 and February of 2005. It was found that the AREX model was superior to the AR(2) model based on the judgment of prediction accuracy and the agreement in the trends of measured PM_{10} concentrations and predicted PM_{10} concentrations. It was concluded that knowing the past

history of air quality was not sufficient enough for prediction since the meteorological conditions also played an important role in the accumulation or removal of PM_{10} on the day of prediction. However, the AREX model proposed in the present study is still preliminary. In the future study, the model can be improved by further investigating the meteorological factors which lead to the accumulation and removal of PM_{10} concentrations in Macau (Chang *et al.*, 2007). In addition to the model class modification, further improvement can be achieved by adjusting the process noise and measurement noise parameters which are demonstrated to affect the performance of Kalman filter (Yuen *et al.*, 2007). Finally, it was concluded that the Kalman filter was promising in the air quality prediction. However, caution should be made in the selection of the model class.

ACKNOWLEDGEMENT

RERERENCES

Anh V.V., Azzi M., Duc H., Johnson G.M. and Tieng Q. (1998), A reactive state-space model for prediction of urban air pollution, *Environmental Modelling and Software*, **13**, 239-246.

Curtis L., Rea W., Smith-Willis P., Fenyves E. and Pan Y. (2006), Adverse health effects of outdoor air pollutants, *Environmental International*, **32**, 815-830.

Chang S.W., Mok K.M. and Yuen K.V. (2007), Association of PM_{10} pollution episodes with the meteorological conditions in Macau, *Proceedings of the 10th International Conference on Environmental Science and Technology*, Kos Island, Greece, 5–7 September 2007, pp. 90-95.

Dockery D.W., Pope CAIII, Xu X., Spengler J.D., Ware J.H., Fay M.E., Ferris B.G. and Speizer F.E. (1993), An association between air pollution and mortality in six U.S. cities, *The New England Journal of Medicine*, **329**, 1753-1759.

Google maps (2007), Geographical location of Macau in the Pearl River Delta region, Retrieved October 1, 2007, from http://maps.google.com/.

Goyal P., Chan T. and Jaiswal N. (2006), Statistical models for the prediction of respirable suspended particulate matter in urban cities, *Atmospheric Environment*, **40**, 2068-2077.

Kalman R.E. (1960), A new approach to linear filtering and prediction problems, *Transactions of ASME, Journal of Basic Engineering*, **82**, 35-45.

Kalman R.E. and Bucy R.S. (1961), New results in linear filtering and prediction theory, *Transactions of ASME, Journal of Basic Engineering*, **83**, 95-107.

Lee Y. C. and Hills P.R. (2003), Cool season pollution episodes in Hong Kong, 1996-2002, *Atmospheric Environment*, **37**, 2927-2939.

Mok K. M. and Hoi K.I. (2003), Association of PM_{10} Concentrations and Wind Directions in Macau, *Proceedings of the International Conference on Pollution in Metropolitan and Urban Environment - POLMET 2003 (POLMET CD Rom)*, Hong Kong, Nov. 3-5, 2003, Session 6A, 6 pages.

Mok K.M. and Hoi K.I. (2005), Effect of meteorological conditions on PM_{10} concentrations – A study in Macau, *Environmental Monitoring and Assessment*, **102**, 201-223.

Macau Meteorological and Geophysical Bureau (2007), General ambient monitoring station and its surrounding environment. Retrieved October 1, 2007, from *http://www.smg.gov.mo/.*

Macau Statistics and Census Bureau (2008), GDP per capita of Macau between 2002 and 2004. Retrieved June 18, 2008, from *http://www.dsec.gov.mo/.*

Ng C.N. and Yan T.L. (2004), Recursive estimation of model parameters with sharp discontinuity in non-stationary air quality data, *Environmental Modelling and Software*, **19**, 19-25.

Ojha S., Coutinho J. and Kumar A. (2002), Developing system to forecast ozone and particulate matter levels, *Environmental Progress*, **21**, J7-J12.

Shi J.P. and Harrison R.M. (1995), Regression modeling of hourly NO_x and NO_2 concentrations in urban air in London, *Atmospheric Environment*, **31**, 4081-4094.

van der Wal J.T. and Jansen L.H.J.M. (2000), Analysis of spatial and temporal variations of PM_{10} concentrations in the Netherlands using Kalman filtering, *Atmospheric Environment*, **34**, 3675-3687.

Yuen K.V., Hoi K.I. and Mok K.M. (2007), Selection of noise parameters for Kalman filter, *Journal of Earthquake Engineering and Engineering Vibration*, **6**, 49-56.

Zolghadri A. and Cazaurang F. (2006), Adaptive nonlinear state-space modelling for the prediction of daily mean PM_{10} concentrations, *Environmental Modelling and Software*, **21**, 885-894.