

Research Article

Adaptive Modelling of the Daily Behavior of the Boundary Layer Ozone in Macau

K. M. Chao, K. I. Hoi, K. V. Yuen, and K. M. Mok

Department of Civil and Environmental Engineering, University of Macau, Macau, China

Correspondence should be addressed to K. M. Mok, kmmok@umac.mo

Received 5 April 2012; Accepted 21 May 2012

Academic Editors: H. Guo and Z. H. Shon

Copyright © 2012 K. M. Chao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The present study aims to develop an efficient dynamic statistical model to describe the daily behavior of boundary layer ozone in Macau. Four types of Kalman-filter-based models were proposed and applied to model the daily maximum of the 8 hr averaged ozone concentrations within a decade (2000–2009). First, the boundary layer ozone was modelled with the time-varying autoregressive model of order p , TVAR(p), which is a pure time series model hindcasting the ozone concentration by a weighted sum of the ozone histories of the previous p days. Then, it was modelled with the time-varying autoregressive model with linear exogenous input, TVAREX-Lin, which combines the TVAR model and the exogenous input of key meteorological variables in a linear fashion. Next, the nonlinear TVAREX model (TVAREX-NLin) which assumes the nonlinear influence of individual meteorological variable on ozone was adopted. Finally, a semiempirical TVAREX model (TVAREX-O₃) was proposed to address the coastal nature of Macau and the interaction between the input variables. It was found that the proposed TVAREX-O₃ model was the most efficient one among the model candidates in terms of the general modelling performance and the capability of modelling the episode situation.

1. Introduction

The increasing trend of boundary layer ozone concentrations in the Pearl River Delta (PRD) region was observed in recent decades due to increasing anthropogenic emissions of ozone precursors such as oxides of nitrogen (NO_x) and volatile organic compounds (VOCs) [1, 2]. This problem has been receiving greater attention due to increasing evidences discovered for the adverse effects of ozone exposure (respiratory and ocular damage, crop yield reduction, climate change, etc.) [3–5]. Therefore, taking actions towards its proper management is necessary and urgent. Developing a representative model to investigate the variation of the ozone concentrations measured in the existing air quality monitoring network is obviously an initial but essential step.

In the past decade, the statistical models developed to model the boundary layer ozone behavior of the Pearl River Delta region are based on offline techniques such as the artificial neural network and the support vector machine [4, 6, 7]. These techniques assume that the model coefficients

should be time-invariant after model training. However, air quality system changes with time, making the offline models underperform [8]. Therefore, the objective of this study is to develop a representative dynamic statistical model to describe the daily behavior of boundary layer ozone of Macau, a fast developing city in the PRD, for improvement. A dynamic statistical model differs from a static one with its time-varying model coefficients. Through changing the model coefficients at every time step, it is believed that the model can adapt to the change in the actual system. In this study, the extended Kalman filter [9, 10] was adopted to estimate the time-varying model coefficients for the proposed models. This technique has been implemented on similar models used to predict the daily averaged PM₁₀ concentrations of Macau before [8, 11, 12]. Therefore, the present study only focuses on describing the proposed models for the ground-level ozone. Detailed procedures of employing the Kalman filter for the state estimation of the statistical air quality models can be referred to in [11].

2. Data and Methodology

2.1. Study Area and Data. Being Asia's well-known gaming mecca, Macau has been experiencing rapid economic growth in the last decade due to the booming of its gaming and tourism industry. The increasing energy consumption has caused degradation on the ambient air quality as well as the indoor one. In view of the public concern, the Macau Meteorological and Geophysical Bureau has been developing automatic air quality monitoring network spanning over the Macau peninsula (urban core), the Taipa Island (suburban area), and the Coloane Island (rural area) to monitor the variation of key air pollutants including ozone since 1999. In this study, the boundary layer ozone concentrations recorded at the ambient (Taipa) station from January 2000 to December 2009 are studied since this station has the longest dataset of all the monitoring stations. In addition, the boundary layer ozone is a more dominant problem in the Taipa island due to more biogenic emissions of VOCs and the lack of NO_x titration [13–15]. Finally, this station is located at the peak of the Taipa Grande Hill, which has an altitude of 110 m above the sea level. Therefore, the ozone concentrations monitored at this station are expected to represent the general ambient condition.

2.2. Model Candidates. Four adaptive statistical air quality models are proposed here. They can be classified into two major types, namely, the time-varying autoregressive (TVAR) model and the time-varying autoregressive model with exogenous inputs (TVAREX). These model candidates are further described below.

2.2.1. TVAR(p) Model. The first model candidate TVAR(p) is defined as the time-varying autoregressive model of order p , which has the following form:

$$x_k = \phi_{1,k-1}x_{k-1} + \dots + \phi_{p,k-1}x_{k-p} + f_{k-1}. \quad (1)$$

In it, the daily maximum of the 8 hr averaged ozone concentration on the k th day x_k is predicted by a weighed combination of the ozone past histories of the previous p days [8]. The weights are given by the uncertain time-varying model coefficients $\phi_{i,k-1}$, $i = 1, \dots, p$, which can be estimated by Kalman filter [11]. The last term f_{k-1} denotes the modelling error which represents the unmodeled system dynamics that affect the ozone concentration and is modeled as a Gaussian-independent and identically distributed (i.i.d.) process with zero mean and the process noise variance σ_f^2 .

2.2.2. TVAREX-Lin Model. The second model candidate TVAREX-Lin is defined as the time-varying autoregressive model with linear exogenous inputs:

$$x_k = \phi_{1,k-1}x_{k-1} + \phi_{2,k-1}x'_{k-1} + \phi_{3,k-1}s_k + \phi_{4,k-1}h_k + \phi_{5,k-1}u_k + \phi_{6,k-1}|\theta_k| + \phi_{7,k-1}I_k + f_{k-1}. \quad (2)$$

Unlike the TVAR(p) model, the TVAREX-Lin model assumes that the ground-level ozone does not only evolve according to

its past histories, but is also influenced by other conditions occurring on the day of prediction. These conditions include (i) the photochemical reactions that transform the precursors into ozone, (ii) the nature of the replenishing air masses, and (iii) the dispersion condition. These mechanisms are reflected by using the linear combination of the influencing meteorological variables as the exogenous inputs. In this model, the symbols x_{k-1} and x'_{k-1} denote the daily maximum of the 8 hr averaged ozone concentration on the $(k-1)$ th day and the 8 hr averaged ozone concentration measured before the $(k-1)$ th day, respectively. These two variables are applied to represent the initial ozone condition. The variable s_k is the solar radiation index which is defined as the product of the daily sum of the 8 hr averaged solar radiation and the duration of sunshine on the k th day. It is used to reflect the amount of energy that propels the photochemical reaction for ozone formation. The symbol h_k is the daily average of the 8 hr averaged relative humidity, and this factor can limit the ozone generation since high humidity encourages the chemical reaction of the NO_2 and the NaCl particles brought by sea breeze [16] on the k th day. The parameters u_k and $|\theta_k|$ are the magnitude and the absolute angle of the resultant wind vector which is calculated by the vector sum of the 8 hr averaged wind velocity vectors on the k th day. It is noted that the 8-hr averaged wind velocity vector is calculated by the vector sum of the hourly wind velocity vectors of the previous 8 hours. The magnitude u_k reflects the dispersion condition, and the absolute angle $|\theta_k|$ represents the leading wind direction which influences the source of the replenishing air masses on the k th day. For example, a resultant angle of 0° refers to the prevailing northerly wind, while 180° indicates the prevailing southerly wind. The input I_k is the index of wind reversion on the k th day:

$$I_k = \max_j [\cos(|\theta_{j-1}|) - \cos(|\theta_j|)], \quad (3)$$

where $|\theta_j|$ is the absolute angle of the resultant wind vector at the j th hour. This index is positively correlated with the maximum of the ozone concentration on the day of prediction. It is defined to capture the sudden reversion of wind direction which may recirculate the contaminants and cause ozone enhancement [17]. The symbol $\phi_{i,k-1}$ is the time-varying coefficient for the associated input variable on the $(k-1)$ th day. Finally, the term f_{k-1} represents the modelling error and it is modeled as Gaussian i.i.d. with zero mean and the process noise variance σ_f^2 .

2.2.3. TVAREX-NLin Model. The third model candidate TVAREX-NLin is defined as the time-varying autoregressive model with nonlinear exogenous inputs:

$$x_k = \phi_{1,k-1}x_{k-1}^{q_{1,k-1}} + \phi_{2,k-1}x'_{k-1}^{q_{2,k-1}} + \phi_{3,k-1}s_k^{q_{3,k-1}} + \phi_{4,k-1}h_k^{q_{4,k-1}} + \phi_{5,k-1}u_k^{q_{5,k-1}} + \phi_{6,k-1}|\theta_k|^{q_{6,k-1}} + \phi_{7,k-1}I_k^{q_{7,k-1}} + f_{k-1}. \quad (4)$$

As the formation of ozone is a nonlinear process, it is attempted to investigate the effect of nonlinearity on the modelling performance by using the same set of input variables of the TVAREX-Lin model and modelling the effect of each input variable with the power law. The symbol $q_{i,k-1}$ denotes the time-varying exponent of the input variable on the $(k-1)$ th day.

2.2.4. TVAREX- O_3 Model. Although the TVAREX-Lin model and the TVAREX-NLin model address the important exogenous inputs which represent the influencing photochemical and physical mechanisms of ozone, their functional forms are solely empirical and may not completely reflect the nature of the study area. In addition, these two models do not account for the interaction between the input variables. Therefore, the semiempirical TVAREX- O_3 model is proposed below to address the coastal characteristics of Macau and the interaction of some input variables:

$$\begin{aligned} x_k = & \left[\phi_{1,k-1}x_{k-1} + \phi_{2,k-1}x'_{k-1} \right. \\ & + \phi_{3,k-1}s_k \exp(-\phi_{4,k-1}h_k - \phi_{5,k-1}|\theta_k| + \phi_{6,k-1}I_k) \\ & \left. \times \exp(-\phi_{7,k-1}u_k) + f_{k-1} \right] \end{aligned} \quad (5)$$

The first two terms are still used to reflect the background ozone conditions, while the third term is used to represent the photochemically formed ozone by the transboundary ozone precursors within the Pearl River Delta region. Since Macau is a coastal city located at the southwest of the PRD, the ozone precursors 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. It is expected that Macau is less influenced by the transboundary air pollution from the cities located to its North when the prevailing wind direction is southerly. Therefore, the absolute angle in the third term is used to represent the effect of the replenishing air masses brought by the prevailing wind on the day of prediction. However, the concentrations of the precursors in the replenishing air masses can be also increased when there is a reversion in the wind direction that causes recirculation and accumulation due to this momentum swing. On the other hand, the amount of water present in the replenishing air masses may affect the concentration of the nitrogen dioxide. Therefore, the variables h_k and I_k also appear in the exponential term to account for these effects on the ozone precursors. Then, the solar radiation index s_k outside the exponential term is adopted to represent the available energy to propel the ozone formation for the precursors present in the incoming air masses due to the photochemical reactions. After the photochemical formation process, the original background concentration together with the photochemically formed one will be dispersed by wind, and the exponential term containing the magnitude of the resultant wind velocity vector is used as a discounting factor on the modelled ground-level ozone concentration.

TABLE 1: General performance of the TVAR(p) models.

p	RMSE ($\mu\text{g m}^{-3}$)	MAPE (%)	R^2	IA
1	35.49	44.34	0.48	0.83
2	35.48	45.54	0.48	0.83
3	35.03	45.68	0.48	0.83
4	34.93	45.67	0.49	0.83
5	34.67	45.89	0.49	0.83
6	34.65	46.02	0.49	0.83
7	34.64	46.04	0.50	0.83

3. Results and Discussion

3.1. General Performance of the Proposed Models. In this section, the proposed models are evaluated based on the measured daily maximum of the 8-hr averaged ozone concentrations recorded at the ambient (Taipa) station between 2001 and 2009. In order to avoid the transient state of Kalman filter, the data in 2000 are not included in the model evaluation. First of all, the general (global) performances of these models are evaluated with some well-known error indices including the root-mean-square error (RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{k=1}^N (z_k - \hat{x}_{k|k-1})^2}, \quad (6)$$

the mean absolute percentage error (MAPE):

$$\text{MAPE} = \frac{1}{N} \sum_{k=1}^N \frac{|z_k - \hat{x}_{k|k-1}|}{z_k} \times 100\%, \quad (7)$$

the coefficient of determination (R^2):

$$R^2 = \left[\frac{\sum_{k=1}^N (z_k - \mu_z)(\hat{x}_{k|k-1} - \mu_x)}{\sqrt{\sum_{k=1}^N (z_k - \mu_z)^2} \sqrt{\sum_{k=1}^N (\hat{x}_{k|k-1} - \mu_x)^2}} \right]^2, \quad (8)$$

and the index of agreement (IA):

$$\text{IA} = 1 - \frac{\sum_{k=1}^N (z_k - \hat{x}_{k|k-1})^2}{\sum_{k=1}^N (|z_k - \mu_z| + |\hat{x}_{k|k-1} - \mu_z|)^2}, \quad (9)$$

where N , z_k , $\hat{x}_{k|k-1}$, μ_x , and μ_z represent the number of measured days, the measured daily maximum of the 8-hr averaged ozone concentration on the k th day, the corresponding hindcast by the model on the k th day, the averaged value of the modelled concentration time history and the averaged value of the measured concentration time history, respectively. Generally, an efficient model associates with small values of RMSE and MAPE as well as large values of R^2 and IA which are close to unity. Table 1 shows the general performance of the TVAR(p) models, $p = 1, \dots, 7$. It is noted that the RMSE of the TVAR models decreases slightly by 2.4%, and the value of R^2 increases by 4.2% when p is increased from 1 to 7. However, the value of IA remains constant, and the MAPE increases by 3.8%. The contradiction of the observed trends for different error statistics implies that the prediction performance may

TABLE 2: General performance of the TVAR(1) model and the TVAREX models in different seasons.

Season	Model	RMSE ($\mu\text{g m}^{-3}$)	MAPE (%)	R^2	IA
Spring	TVAR(1)	39.04	46.78	0.37	0.77
	TVAREX-Lin	38.09	41.78	0.46	0.83
	TVAREX-NLin	38.51	43.63	0.40	0.80
	TVAREX-O ₃	25.33	34.87	0.50	0.85
Summer	TVAR(1)	36.79	45.53	0.42	0.79
	TVAREX-Lin	36.28	38.52	0.53	0.86
	TVAREX-NLin	36.81	40.64	0.50	0.81
	TVAREX-O ₃	23.28	30.92	0.65	0.90
Autumn	TVAR(1)	36.61	45.20	0.45	0.81
	TVAREX-Lin	34.56	40.01	0.57	0.87
	TVAREX-NLin	34.64	40.32	0.50	0.83
	TVAREX-O ₃	22.46	30.88	0.71	0.91
Winter	TVAR(1)	32.26	43.05	0.50	0.84
	TVAREX-Lin	30.51	32.67	0.64	0.89
	TVAREX-NLin	31.21	33.21	0.58	0.86
	TVAREX-O ₃	24.32	31.03	0.60	0.90

not be necessarily improved by using a higher-order TVAR model. This finding agrees with the short half-life time of ozone (~ 12 hours) in the atmosphere. Therefore, the most parsimonious TVAR(1) model was chosen as the most plausible model among TVAR(p) model candidates. Table 2 shows the general performance of the TVAR(1) model as well as the TVAREX models in different seasons. First of all, it is observed that TVAR(1) model is the least efficient among four models with the largest RMSE and MAPE as well as smallest values of R^2 and IA throughout the years. It is possibly due to the lack of the information about the meteorological conditions which may highly influence the ozone production. In addition, it is interesting to find that the TVAREX-Lin model generally outperforms the TVAREX-NLin model. It should be noted that both models share the set of input variables. The former is a linear model, and the latter is a nonlinear one with more adjustable time-varying parameters. However, the extra flexibility given to the TVAREX-NLin model may be unnecessary and just make it to be more sensitive to the local fluctuation in the data. Therefore, this flexibility makes the TVAREX-NLin become more difficult to tune with the general behavior of the concentration time history and hence leads to its underperformance. On the contrary, the proposed TVAREX-O₃ model is also a nonlinear model which has just the same number of uncertain parameters as the TVAREX-Lin model. However, it is more efficient compared to the other two empirical TVAREX models. Furthermore, it is noted that the degree of performance improvement of the TVAREX-O₃ model compared to TVAREX-Lin model in Summer and Autumn is larger than that in Spring and Winter as shown in Table 3, meaning that TVAREX-O₃ model is especially efficient in the episode seasons. Therefore, this demonstrates that deciding the relevant input variables before the model construction is important. However, designing a meaningful functional form which reflects the correct role played by each input

TABLE 3: Percentage differences of the error statistics between the TVAREX-O₃ model and the TVAREX-Lin model in different seasons.

Season	RMSE	MAPE	R^2	IA
Spring	33.5%	16.5%	8.7%	2.4%
Summer	35.8%	19.7%	22.6%	4.7%
Autumn	35.0%	22.8%	24.6%	4.6%
Winter	20.3%	5.0%	6.3%	1.1%

variable in the model is also the key leading to the success. Once the global performance of the proposed models has been evaluated and compared, it is also worthy to investigate more details from the local modelling performance of the TVAR models and the TVAREX models. In the following section, the local modelling performance of each model is evaluated based on graphical comparisons of the measured concentrations as well as their hindcasts.

3.2. Local Performance of the Proposed Models. Figure 1(a) shows the scatter plot of daily maximum values of the 8-hr averaged ozone concentrations modelled by the TVAR(1) model against the measurements. A 45° line was also drawn on the same figure for comparison. When a data point falls on the 45° line, it means that the hindcast by the model is exactly equal to the measurement. It is noted that a large number of points are still lying close to the 45° line, meaning that the TVAR(1) is capable to capture the general trend of the time history. For the cases of high predicted ozone concentrations or high measured ozone concentrations, those points are far above or below the 45° line. This is possibly due to the time delay problem, that is, the modelled time history generally lags behind the measured one. This problem may cause large modelling error when

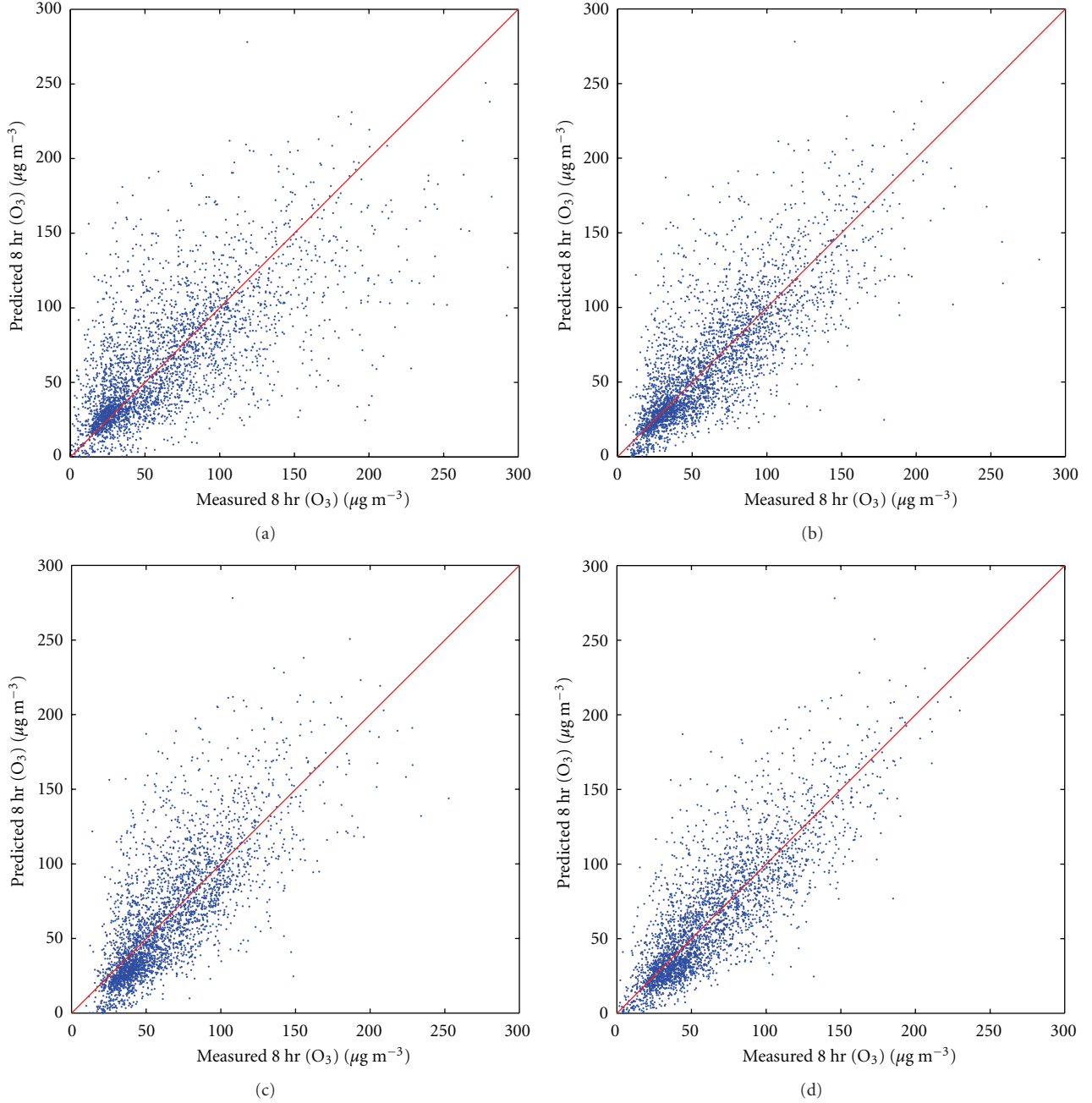


FIGURE 1: Plot of daily maximum of the 8 hr averaged ozone concentrations predicted by (a) TVAR(1), (b) TVAREX-Lin, (c) TVAREX-NLin, and (d) TVAREX- O_3 model against the measurement.

there is significant variation in the daily maximum value, and those points are associated with the onset and retreat of the ozone episode.

Figures 1(b)–1(d) show the scatter plots of the modelled daily maximum of the 8-hr averaged ozone concentrations versus their measurements for TVAREX-Lin, TVAREX-NLin, and TVAREX- O_3 models, respectively. It is obvious to see that the points in these figures are lying closer to the 45° line compared to the points in Figure 1(a); hence signifying the improvement of the time delay problem. By a detailed comparison of these three figures, it is noted that a portion

of the points in Figures 1(b) and 1(c) are lying far above the 45° line compared to the same region in Figure 1(d).

This indicates that both the TVAREX-Lin and the TVAREX-NLin models may produce more false alarms compared to the TVAREX- O_3 model. In addition, it is noted that the points of TVAREX- O_3 model are more concentrated around the 45° line compared to those of the TVAREX-Lin and TVAREX-NLin models even under the conditions of high ozone concentrations. Therefore, the TVAREX- O_3 model has the most efficient local performance, and this further supports the observations in the previous subsection.

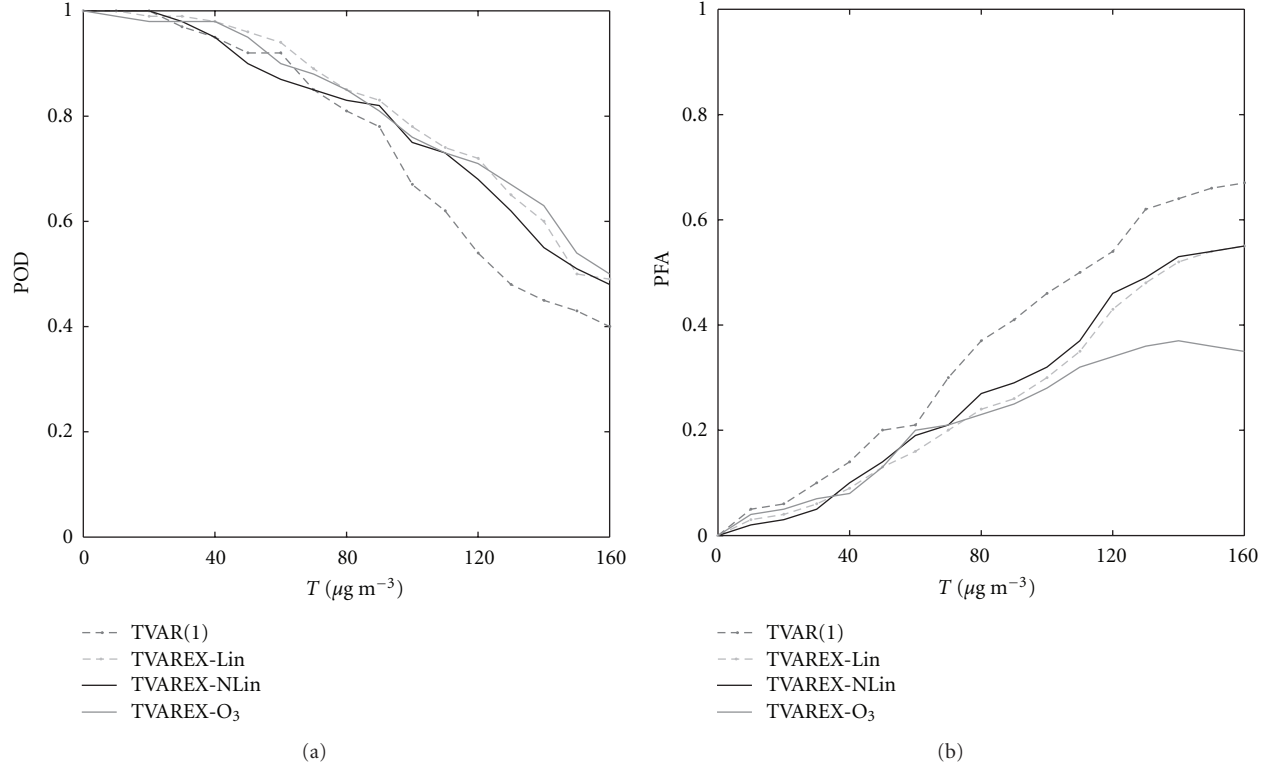


FIGURE 2: Variation of (a) PODs and (b) PFAs for the TVAR(1) model and the TVAREX models.

3.3. Episode Capturing Capability of the Proposed Models. As indicated in the previous graphical comparison of the local modelling performance, the time-varying model candidates can have significantly different capabilities of capturing the pollution episodes. In this section, it is aimed to further investigate this diversity more quantitatively by using two more numerical indicators, namely, the probability of detection (POD) and the probability of false alarm (PFA). The probability of detection (POD) is defined as the probability of issuing a successful warning for a given episode day:

$$\text{POD} = P(\hat{x}_{k|k-1} \geq T \mid z_k \geq T), \quad (10)$$

where T is the threshold concentration of the ozone episode. High POD value usually indicates that the model is capable to warn the public for the bad air quality. However, it should be kept in mind that overestimated predictions of the model may also associate with a high POD. Therefore, the probability of false alarm (PFA) is commonly used together with POD to check the episode capturing capability of a model. The PFA is defined as the probability of having a nonepisode day given that a warning has been issued:

$$\text{PFA} = P(z_k < T \mid \hat{x}_{k|k-1} \geq T). \quad (11)$$

For example, a perfect model should have a POD of 1 and a PFA of 0. Figure 2(a) shows the PODs of the TVAR(1) model and the TVAREX models for different thresholds ranging from 0 to $160 \mu\text{g m}^{-3}$, respectively. It is noted that the PODs of the TVAR(1) model are significantly lower than those of

the TVAREX models when the threshold T is larger than $80 \mu\text{g m}^{-3}$. This implies that the onset of ozone episodes in Macau is closely linked to the meteorological conditions on the day of prediction and may explain why the TVAR(1) model underperforms in this situation. For the TVAREX models, it is noted that the PODs of the TVAREX-O₃ model are slightly higher than those of the TVAREX-Lin and TVAREX-NLin model when the threshold T is larger than $120 \mu\text{g m}^{-3}$. This finding reinforces similar observation made in the graphical comparison of the local performances between the TVAR(1) model and the TVAREX models. Figure 2(b) shows the PFAs of the TVAR(1) model and the TVAREX models against different threshold concentrations. First of all, it is noted that the TVAR(1) model produces more frequent false alarm compared to the TVAREX models. Since the TVAR(1) model predicts the daily maximum value of the 8-hr averaged ozone concentration tomorrow based on the filtered estimate of today only, there is a high probability of predicting it is an episode day tomorrow given that today is an ozone episode day. However, the episode may retreat tomorrow due to the swing of prevailing wind direction from north to south, or the decrease in the solar radiation. Therefore, false alarm will be produced by the TVAR(1) model under this situation. In addition to the PODs, it is noted that the PFAs of the TVAREX-O₃ model are significantly less than those of the TVAREX-Lin and TVAREX-NLin models when the threshold concentration is larger than $100 \mu\text{g m}^{-3}$. This confirms that the improvement of the episode capturing capability in the TVAREX-O₃ model

is not done by sacrificing the false alarm rate. Finally, it should be concluded that the TVAREX-O₃ model is the most efficient model to capture the ozone pollution episodes in this study.

4. Conclusion

Four types of time-varying statistical models, namely, the TVAR(p) model, the TVAREX-Lin model, the TVAREX-NLin model, and the TVAREX-O₃ model were proposed to model the daily maximum of the 8-hr averaged ozone concentrations observed at the ambient (Taipa) station of Macau within a decade (2000–2009). Throughout comparing the general performance of the TVAR(p) model with the order ranging from 1 to 7, it was concluded that the daily behavior of the ground level ozone system in Macau has the Markovian property [18], and this coincides with its short half-life time. Further comparison of the TVAR(1) model and the TVAREX models in terms of the general modelling performance, the local modelling performance and the episode capturing capability indicates that the TVAR(1) model underperforms. Therefore, it was concluded that the daily fluctuation of the ozone time history in Macau is dominated by the meteorological conditions on the day of prediction. In addition, it was found that the TVAREX-Lin model outperforms the TVAREX-NLin model, and this might be due to the unnecessary model parameters of the TVAREX-NLin model, which makes the model itself become more difficult to tune with the general behavior of the data. Therefore, a complex model is not necessarily better than a simple one. Finally, it was found that the TVAREX-O₃ model, which considers the coastal nature of Macau and the interaction of the input variables in its functional form, outperforms the other two empirical TVAREX models. Therefore, it was concluded that selection of the relevant input variables during the model construction is important. However, designing a meaningful functional form which reflects the correct role played by each input variable in the model is also the key leading to the success.

Acknowledgments

This work is supported by the Macau Science and Technology Development Fund (025/2011/A) and the University of Macau (UL019/08-Y4/CEE/MKM01/FST). The Macau Meteorological and Geophysical Bureau is thanked for supplying the data of the ambient station.

References

- [1] S. Solberg, R. Bergström, J. Langner, T. Laurila, and A. Lindskog, "Changes in Nordic surface ozone episodes due to European emission reductions in the 1990s," *Atmospheric Environment*, vol. 39, no. 1, pp. 179–192, 2005.
- [2] S. I. V. Sousa, F. G. Martins, M. C. M. Alvim-Ferraz, and M. C. Pereira, "Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations," *Environmental Modelling and Software*, vol. 22, no. 1, pp. 97–103, 2007.
- [3] S. Syri, M. Amann, W. Schöpp, and C. Heyes, "Estimating long-term population exposure to ozone in urban areas of Europe," *Environmental Pollution*, vol. 113, no. 1, pp. 59–69, 2001.
- [4] W. Wang, W. Lu, X. Wang, and A. Y. T. Leung, "Prediction of maximum daily ozone level using combined neural network and statistical characteristics," *Environment International*, vol. 29, no. 5, pp. 555–562, 2003.
- [5] O. Pastor-Bárcenas, E. Soria-Olivas, J. D. Martín-Guerrero, G. Camps-Valls, J. L. Carrasco-Rodríguez, and S. Del Valle-Tascón, "Unbiased sensitivity analysis and pruning techniques in neural networks for surface ozone modelling," *Ecological Modelling*, vol. 182, no. 2, pp. 149–158, 2005.
- [6] M. Cai, Y. Yin, and M. Xie, "Prediction of hourly air pollutant concentrations near urban arterials using artificial neural network approach," *Transportation Research Part D*, vol. 14, no. 1, pp. 32–41, 2009.
- [7] W. F. Ip, C. M. Vong, J. Y. Yang, and P. K. Wong, "Forecasting daily ambient air pollution based on least squares support vector machines," in *Proceedings of the IEEE International Conference on Information and Automation (ICIA '10)*, pp. 571–575, Haerbin, Heilongjiang, China, June 2010.
- [8] K. I. Hoi, K. V. Yuen, and K. M. Mok, "Prediction of daily averaged PM₁₀ concentrations by statistical time-varying model," *Atmospheric Environment*, vol. 43, no. 16, pp. 2579–2581, 2009.
- [9] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Transactions of the ASME, Journal of Basic Engineering*, vol. 82, pp. 35–45, 1960.
- [10] R. E. Kalman and R. S. Bucy, "New results in linear filtering and prediction theory," *Transactions of the ASME, Journal of Basic Engineering*, vol. 83, pp. 95–107, 1961.
- [11] K. I. Hoi, K. V. Yuen, and K. M. Mok, "Optimizing the performance of kalman filter based statistical time-varying air quality models," *Global Nest Journal*, vol. 12, no. 1, pp. 27–39, 2010.
- [12] K. V. Yuen, *Bayesian Methods for Structural Dynamics and Civil Engineering*, John Wiley and Sons, 2010.
- [13] M. Q. Pan, "Analysis of air quality situation in Macau," Final Year Project Report of B.Sc. in Civil Engineering, University of Macau, Macau, 2009.
- [14] C. Dueñas, M. C. Fernández, S. Cañete, J. Carretero, and E. Liger, "Analyses of ozone in urban and rural sites in Málaga (Spain)," *Chemosphere*, vol. 56, no. 6, pp. 631–639, 2004.
- [15] S. Sillman, "The relation between ozone, NO_x and hydrocarbons in urban and polluted rural environments," *Atmospheric Environment*, vol. 33, no. 12, pp. 1821–1845, 1999.
- [16] C. Dueñas, M. C. Fernández, S. Caete, J. Carretero, and E. Liger, "Assessment of ozone variations and meteorological effects in an urban area in the Mediterranean Coast," *Science of the Total Environment*, vol. 299, no. 1–3, pp. 97–113, 2002.
- [17] T. Wang, Y. Y. Wu, T. F. Cheung, and K. S. Lam, "A study of surface ozone and the relation to complex wind flow in Hong Kong," *Atmospheric Environment*, vol. 35, no. 18, pp. 3203–3215, 2001.
- [18] D. S. Wilks, *Statistical Methods in the Atmospheric Sciences*, Academic Press, 3rd edition, 2011.

