A statistical model for 1-hour- to 24-hour-ahead prediction of hourly ozone concentrations at ground level in Singapore

X. Liu[1], Y. Hwang[2], K. Yeo[2], J. Hosking[2], A. Barut[2], J. Singh[1] & Y. Amemiya[2]
[1]*IBM Research Collaboratory Singapore, Singapore*
[2]*IBM Thomas J. Watson Research Center, USA*

Abstract

A spatio-temporal statistical model is proposed for 1- to 24-hour-ahead prediction of hourly ozone concentrations. This is a joint work with the National Environmental Agency Singapore, and is Singapore's first predictive model for ozone concentrations. Unlike many existing models which focus on either daily maximum or 8h average daylight ozone concentrations, the present work is concerned with the prediction of hourly ozone concentrations which are usually associated with higher variability. A recently proposed framework for spatio-temporal prediction is used to model ozone concentration data. The macro-scale spatio-temporal variation of ozone concentrations is modeled by a linear function of five carefully constructed predictors, while the micro-scale variation is captured by a mean-zero spatio-temporally correlated random process. We show that this model also provides useful insights about the effects of some complex environmental processes on ozone concentration; this is indeed an attractive feature for any data-driven air quality model.
Keywords: ozone, spatio-temporal statistics, random process.

1 Introduction

Ground-level ozone is one of the major air pollutants regulated by the U.S. Clean Air Act [1] as well as the air quality guidelines of the World Health Organization [2]. Although the upper-atmosphere ozone layer shields us from

WIT Transactions on Ecology and The Environment, Vol 183, © 2014 WIT Press
www.witpress.com, ISSN 1743-3541 (on-line)
doi:10.2495/AIR140321

the harmful ultraviolet rays, ground-level ozone is usually harmful. When ozone is inhaled, it irritates the respiratory system, inflames and damages the lining of the lungs, increase the susceptibility to respiratory infections, etc. When the 8-hour concentration exceeds $240\mu g/m^3$, both healthy adults and asthmatics experience significant reductions in lung function [2, 3]. The Air Quality Index (AQI) is considered good or moderate when the 8-hour mean ozone concentration is not higher than 0.08 parts per million, which is approximately $169\mu g/m^3$. The Singapore National Environmental Agency (NEA) adopts the WHO guidelines and set the target for 8-hour mean ozone concentration to $100\mu g/m^3$.

Ozone is formed in the atmosphere by photochemical reactions in the presence of sunlight and precursor pollutants, such as the oxides of nitrogen (NO_x) and volatile organic compounds (VOCs). It is destroyed by reactions with NO and is deposited to the ground. Several studies have shown that ozone concentration is correlated with various toxic photochemical oxidants arising from similar sources, including the peroxyacyl nitrates, nitric acid and hydrogen peroxide [4].

As a joint work between IBM Research and Singapore NEA, we propose in this paper a novel statistical model to predict the *hourly* average ground-level ozone concentrations 1 to 24 hours in advance. This is Singapore's *first* predictive model for ozone concentrations. Although various methods have been proposed to model ozone concentrations (see [5] for a comprehensive survey), most of the existing methods focus on either the daily maximum or daylight 8-hour average ozone concentration [6–8]. In this research, we focus on the prediction of hourly ozone concentrations, which is more challenging as ozone concentrations at hourly level are usually associated with higher uncertainty. Furthermore, the study area considered in this work, i.e., Singapore, is much smaller than that of many existing studies and is characterized by a dense built-up urban area. Hence, being able to model the local spatio-temporal variability of ozone concentrations becomes more critical.

For the past decades, spatio-temporal statistics have been widely employed in modeling air quality data [8–10]. In these studies, pollutants concentrations are modeled by some random field. Assumptions such as stationary covariance and isotropy (i.e., covariance function only depends on distance) are often empirically made so as to keep the model mathematically tractable. These assumptions might be appropriate when data are aggregated over a relatively large area or long time period. For hourly-level data, however, assumptions like stationary covariance and isotropy rarely hold due to the high variability associated with hourly ozone concentrations and meteorological conditions. In fact, such commonly made assumptions may not even be validated using statistical approaches unless data arise from a large number of air quality monitoring stations [11]. To relax such limitations, this study employs a spatio-temporal prediction framework recently developed by researchers from IBM Watson Research Center [12]. Our study shows that this framework can effectively model ozone concentration and well balances the complexity and the practicality of a spatio-temporal prediction approach. In Section 2, we describe the study area and the data used to construct the model. Section 3 provides the modeling details. Results of model testing are

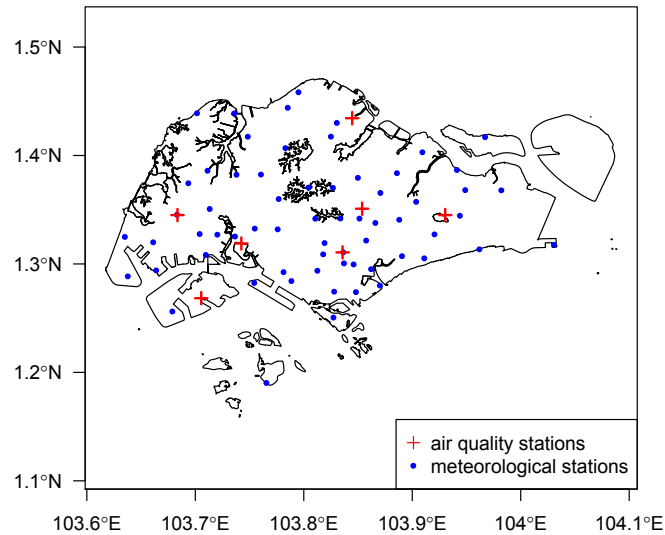

Figure 1: Locations of air quality and meteorological stations in Singapore.

summarized in Section 4. It is also noted that, the model constructed can be interpreted in terms of the basic physics and chemistry of ozone; this is indeed an attractive feature for a data-driven air quality model, and discussions can be found in Section 4.

2 Study area and data

2.1 Study area

As a tropical island country located 137km north of the equator, Singapore has a total land area of 716.1km^2 and a population close to 5.4 million. The population density is $7,669/\text{km}^2$ and ranked the 3rd highest in the world [13]. The motor vehicle population has also reached 974,170 by the end of 2013 [14], and the annual mean temperature typically ranges from $24°\text{C}$ to $32°\text{C}$, creating the ideal conditions for ozone to be generated.

2.2 Data

Hourly-level ozone concentrations ($\mu g/m^3$) measured at seven air quality monitoring stations are available for our study. In this paper, we focus on the data

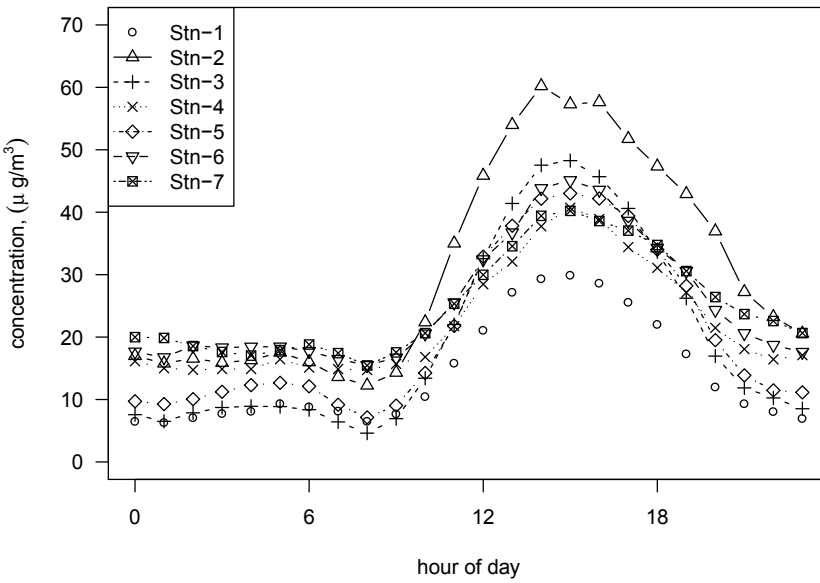

Figure 2: Diurnal variation of ozone concentrations.

collected from 8am 01/06/2013 to 7am 30/08/2013 (Singapore local UTC+08:00 time zone). Locations of stations are shown in Figure 1. Approximately 5.5% of the ozone measurements are missing and the occasions of no measured data are due to system under maintenance. Figure 2 displays the mean ozone concentration at a given hour of a day and for all seven stations. The diurnal ozone concentration is mainly due to the fact ozone can only be generated under sunlight. The spatial variation, on the other hand, is much more dynamic since it depends on factors such as traffic, wind direction and speed, land use, hour of a day and so on.

Temperature, wind speed and wind direction are also available at meteorological stations. Locations of these stations are shown in Figure 1. Note that, air quality stations and meteorological stations do not share same locations, and therefore, spatial interpolation of meteorological conditions is needed. This is appropriate when the density of meteorological stations is high, just as in our case.

Land use information is also used in our modeling, but we leave the detailed description and processing of these datasets in Section 3.2.

3 The model

3.1 The general framework

Let $\mathrm{oz}_t(\mathbf{s})$ be the ozone concentration at location $\mathbf{s}$ and time t, where the spatial location $\mathbf{s}$ is defined as a vector of its coordinates under SVY21 coordinated

cadastre system. Because the distribution of ozone concentration is right-skewed, we model the square root transformation of $\mathrm{oz}_t(\mathbf{s})$ as follows:

$$Y_t(\mathbf{s}) = \sqrt{\mathrm{oz}_t(\mathbf{s})} = g_t(\mathbf{s}) + Z_t(\mathbf{s}) \tag{1}$$

where $\mathbf{s} \in \{\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_n\}$ and $t = 1, \ldots, m$. In equation (1), $g_t(\mathbf{s})$ is a deterministic function which captures the macro-scale spatio-temporal trend of ozone concentration, and $Z_t(\mathbf{s})$ is a mean-zero spatio-temporal correlated random process which captures the micro-scale spatio-temporal variation. Details of the modeling of $g_t(\mathbf{s})$ and $Z_t(\mathbf{s})$ are provided in sections 3.2 and 3.3, respectively,

3.2 The deterministic spatio-temporal trend

The macro-scale spatio-temporal trend of ozone concentration are influenced by factors such as land use, temperature, wind, etc. Being able to build the right model for $g_t(\mathbf{s})$ largely determines whether the spatio-temporal variation of the ozone concentrations can be accurately predicted. In fact, sophisticated treatment of the random component $Z_t(\mathbf{s})$ might not yield a substantial payoff for prediction accuracy [12].

We model $g_t(\mathbf{s})$ as a linear function of five predictors as follows:

$$\begin{aligned} g_t(\mathbf{s}) = {} & \mathbf{x}_t^{(land)}(\mathbf{s})\beta^{(land)} + \mathbf{x}_t^{(upwind)}(\mathbf{s})\beta^{(upwind)} \\ & + \mathbf{x}_t^{(temp)}(\mathbf{s})\beta^{(temp)} + \mathbf{x}_t^{(ws)}(\mathbf{s})\beta^{(ws)} + \mathbf{x}_t^{(hour)}(\mathbf{s})\beta^{(hour)} \end{aligned} \tag{2}$$

The first predictor $\mathbf{x}_t^{(land)}(\mathbf{s})$ is constructed from the land use of Singapore. Different types of land use to a large extent determine the macro-scale spatial trend of ozone emissions [15]. In the raw dataset of land use, Singapore is divided into 110830 spatial polygons with 32 categories of land use. To model the spatial and temporal patterns of ozone in downtown and suburb, industrial and residential, urban and nature reserve areas, as well as the local reduction of ozone concentrations due to traffic-induced NO-scavenging, we re-group the 32 categories into five main land use types, including residential, road, nature reserve, commercial and industrial.

For any location $\mathbf{s}$ and land use type i ($i = 1, 2, \ldots, 5$), we define the land use index $l^{(i)}(\mathbf{s})$ as the total spatial area of land use of type i within 5km radius of location $\mathbf{s}$. As an illustration, Figure 3a and 3b respectively show the land use index for industrial areas (i.e., $l^{(5)}$) and residential areas (i.e., $l^{(1)}$) for Singapore. Here, each pixel has an area of $0.7\mathrm{km}^2$.

Our exploratory analysis also suggests that different types of land use have different effects on ozone concentrations at different hours of a day. Hence, let $\tau(t)$ be a function that returns the hour of a calendar time t ($\tau(t) \in (1, 2, \ldots, 24)$), we construct the predictor $\mathbf{x}_t^{(land)}(\mathbf{s})$ based on the land use index as follows:

$$\mathbf{x}_t^{(land)}(\mathbf{s}) = \left\{\mathbf{x}_t^{(land_i)}(\mathbf{s})\right\}_{i=1}^{5} \tag{3}$$

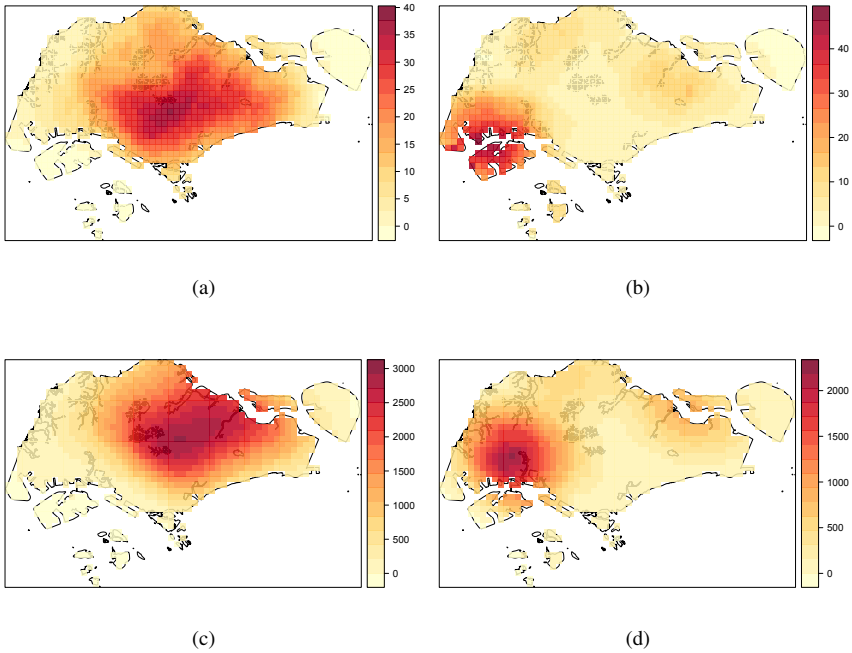

Figure 3: (a) Land use index for industrial areas (km^2); (b) Land use index for residential areas (km^2); (c) Upwind land use index for industrial areas (km^2); (d) Upwind land use index for residential areas (km^2). For (c) and (d), the wind speed equals 8.4km/hour and the wind direction equals 227^o.

Figure 3: (a) Land use index for industrial areas (km^2); (b) Land use index for residential areas (km^2); (c) Upwind land use index for industrial areas (km^2); (d) Upwind land use index for residential areas (km^2). For (c) and (d), the wind speed equals 8.4km/hour and the wind direction equals 227^o.

Our exploratory analysis also suggests that different types of land use have different effects on ozone concentrations at different hours of a day. Hence, let $\tau(t)$ be a function that returns the hour of a calendar time t ($\tau(t) \in (1, 2, \ldots, 24)$), we construct the predictor $\mathbf{x}_t^{(land)}(\mathbf{s})$ based on the land use index as follows:

$$\mathbf{x}_t^{(land)}(\mathbf{s}) = \left\{ x_{t,j}^{(land_i)}(\mathbf{s}) \right\}_{i=1}^{5} \tag{3}$$

and the jth ($j = 1, 2, \ldots, 24$) component of $\mathbf{x}_t^{(land_i)}(\mathbf{s})$ is given by

$$x_{t,j}^{(land_i)}(\mathbf{s}) = \begin{cases} l^{(i)}(\mathbf{s}), & j = \tau(t) \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

The second predictor $\mathbf{x}_t^{(upwind)}(\mathbf{s})$ accounts for the emissions transported downwind. We adopt and extend the idea in [11] and define the upwind land use index $\mathbf{x}^{(upwind)}(\mathbf{s})$. Let $\mathbf{w}_t$ be the island-wide mean wind vector for Singapore at time t, and let v_t denote the wind speed. Then, for any location $\mathbf{s}_0$ and land use type i ($i = 1, 2, \ldots, 5$), the upwind land use index $l^{(i)}(\mathbf{s})$ is defined as the total area of land use of type i within an upwind cone $\mathbb{S}$ defined by

$$\mathbb{S} = \left\{ \mathbf{s} : |\mathbf{s} - \mathbf{s}_0| \leqslant v_t, \frac{(\mathbf{s} - \mathbf{s}_0)\mathbf{w}_t}{v_t} \geqslant |\mathbf{s} - \mathbf{s}_0| \cos(\alpha/2), \frac{(\mathbf{s} - \mathbf{s}_0)\mathbf{w}_t}{v_t} < 0 \right\}$$

For any particular location $\mathbf{s}_0$, $\mathbb{S}$ is an upwind cone with radius proportional to wind speed. The arc of the cone is determined by α and is to capture the variation

of wind directions. Unlike the static land use index, the upwind land use index is dynamic as wind changes its speed and direction. As an illustration, the left and right panels of Figure 3c and Figure 3d respectively show the upwind land use index for industrial areas (i.e., $\tilde{l}^{(5)}$) and for residential areas (i.e., $\tilde{l}^{(1)}$) for Singapore, given that the wind speed equals $8.4\mathrm{km/hour}$ and the wind direction equals 227^o.

The upwind land use index also has different effects on ozone concentrations at different hours of a day. Similarly, we construct the predictor $\mathbf{x}_t^{(upwind)}(\mathbf{s})$ based on the upwind land use index as follows:

$$\mathbf{x}_t^{(upwind)}(\mathbf{s}) = \left\{\mathbf{x}_t^{(upwind_i)}(\mathbf{s})\right\}_{i=1}^{5} \tag{5}$$

and the jth ($j = 1, 2, \ldots, 24$) component of $\mathbf{x}_t^{(upwind_i)}(\mathbf{s})$ is given by

$$x_{t,j}^{(upwind_i)}(\mathbf{s}) = \begin{cases} \tilde{l}^{(i)}(\mathbf{s}), & j = \tau(t) \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

The third predictor $\mathbf{x}_t^{(temp)}$ is obtained from the temperature at time t and location $\mathbf{s}$. Since ozone is only generated during daylight, temperature at night does not significantly affect ozone concentrations. Hence, we construct $\mathbf{x}_t^{(temp)}(\mathbf{s})$ as follows:

$$\mathbf{x}_t^{(temp)}(\mathbf{s}) = \begin{cases} T_t(\mathbf{s}), & \text{if } \tau(t) \in (8, 9, \ldots, 19) \\ 0 & \text{if } \tau(t) \in (1, \ldots, 7) \cup (20, \ldots, 24) \end{cases} \tag{7}$$

where $T_t(\mathbf{s})$ is the temperature at time t and location $\mathbf{s}$.

The fourth predictor $\mathbf{x}_t^{(ws)}$ is constructed from the island-wide mean wind speed v_t at time t. It is mainly used to capture the dilution of ozone concentrations at night due to wind. Hence, we construct $\mathbf{x}_t^{(ws)}(\mathbf{s})$ as follows:

$$\mathbf{x}_t^{(ws)}(\mathbf{s}) = \begin{cases} \mathbf{v}_t, & \text{if } \tau(t) \in (1, 2, \ldots, 7) \cup (20, \ldots, 24) \\ 0 & \text{if } \tau(t) \in (8, \ldots, 19) \end{cases} \tag{8}$$

The last predictor $\mathbf{x}_t^{(hour)}(\mathbf{s})$, which is given by equation (9), accounts for the hourly effect of ozone concentration of a day.

$$x_{t,j}^{(hour)}(\mathbf{s}) = \begin{cases} 1, & j = \tau(t) \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

where $x_{t,j}^{(hour)}(\mathbf{s})$ is the jth component of $\mathbf{x}_t^{(hour)}(\mathbf{s})$, and $j = 1, 2, \ldots, 24$.

3.3 The spatio-temporal random process

Following the idea of [12], the spatio-temporal process in equation (1), $Z_t(s)$, is modeled as an autoregressive (AR) model of order L

$$Z_t(\mathbf{s}) = \sum_{l=1}^{L} \gamma_l Z_{t-l}(\mathbf{s}) + \epsilon(\mathbf{s}) \tag{10}$$

with the AR residuals, $\epsilon = (\epsilon(\mathbf{s}_1), \epsilon(\mathbf{s}_2), \ldots, \epsilon(\mathbf{s}_n))$, being a mean-zero, second-order stationary and isotropic stochastic random field, i.e., $E(\epsilon) = \mathbf{0}$ and $Var(\epsilon) = \Sigma = (\Sigma)_{i,j=1,2,\ldots,n}$. Here, an Exponential spatial covariance function is adopted

$$\Sigma_{i,j} = Cov(\epsilon(\mathbf{s}_i), \epsilon(\mathbf{s}_j)) = \begin{cases} \sigma^2 \exp(-\theta h_{i,j}), & h_{i,j} > 0 \\ \sigma^2 + \kappa^2, & h_{i,j} = 0 \end{cases} \tag{11}$$

where $h_{i,j} = \|\mathbf{s}_i - \mathbf{s}_j\|$ is the distance between site $\mathbf{s}_i$ and $\mathbf{s}_j$ for $i, j = 1, 2, \ldots, n$, κ is known as the nugget effect in spatial statistics, and θ is an unknown parameter.

Hence, the spatio-temporal process $\mathbf{Z} = (Z_{t-L+1}(\mathbf{s}), \ldots, Z_t(\mathbf{s}))$ has a separable space-time covariance function such that

$$Var(\mathbf{Z}) = \Gamma \bigotimes \Sigma \tag{12}$$

where Γ is the AR(L) covariance matrix.

4 Model evaluation and discussions

The model is evaluated using the data collected from 8am 01-June-2013 to 7am 30-Aug-2013, including the observed hourly ozone concentrations, observed hourly weather data, Numerical Weather Prediction (NWP) data and land use data. Starting from 7am 01-July-2013, predictions of hourly ozone concentrations for the next 24 hours (i.e., from 8am to 7am next day) are generated and compared to the observed ozone concentrations. The model is run on daily basis till 7am 29-Aug-2013, and model parameters are dynamically re-estimated using the data in the most recent 168 hours.

With reference to the performance metrics recommended by USEPA [5, 16] and UK DEFRA [17] for evaluation of air quality models, the following metrics are used: Mean Normalized Gross Error (MNGE), Mean Gross Error (MGE), Root-Mean-Square Error (RMSE), Normalized Mean Bias (NMB), Mean Normalized Bias (MNB), Normalized Mean Error (NME) and Mean Bias (MB). Table 1 shows the values for above performance metrics for our model. As an illustration, Figure 4 shows both the predicted and observed hourly pollutant concentrations in August.

Results presented in Table 1 show that the accuracy of the proposed model is comparable to that of many existing models, such as CMAQ, reported in the

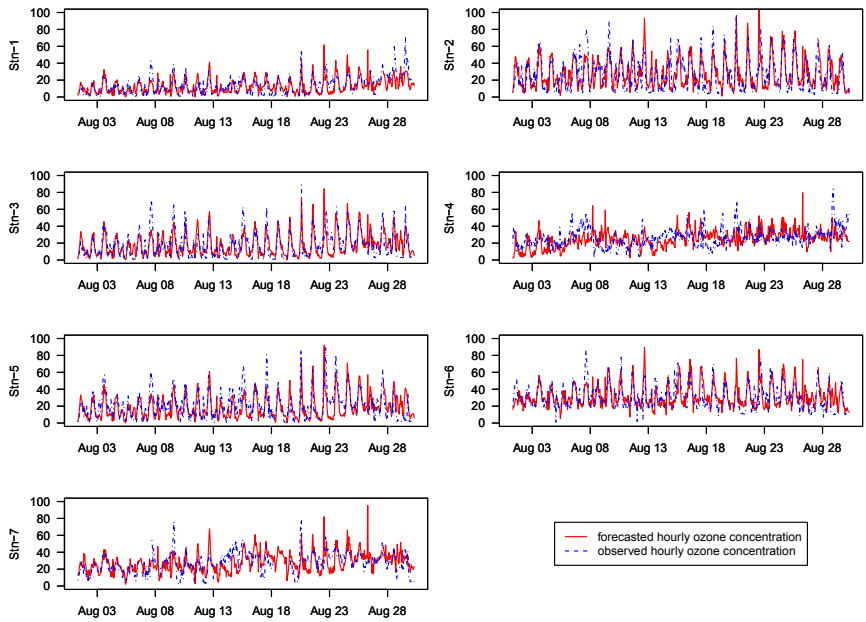

Figure 4: Time series plot of the forecasted and observed hourly ozone concentrations ($\mu g/m^3$).

Table 1: A summary of model performance.

	MGE ($\mu g/m^3$)	MNGE (%)	RMSE	NMB
Hourly	9.05	35.61	12.12	−0.01
Hourly (daylight)	9.49	28.17	12.80	−0.02
8-hour average	6.74	19.76	8.82	−0.02
	MNB (%)	NME (%)	MB ($\mu g/m^3$)	
Hourly	3.03	39.00	−0.23	
Hourly (daylight)	0.42	32.64	−0.52	
8-hour average	2.57	21.87	−0.56	

literature [7, 18–20]. However, being a statistical model, the proposed model is much more efficient in terms of computation time.

It is worth noting that the model constructed in this paper also provides useful insights about environmental processes responsible for ozone concentration. For

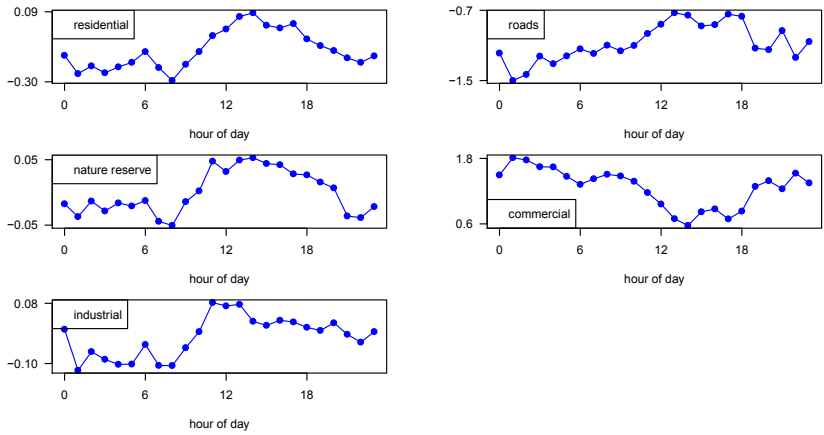

Figure 5: Effects of different land use on ozone concentrations.

illustration purposes, Figure 5 shows the estimated effects of different land use on ozone concentrations based on the data from Aug 8 to Aug 15. We see that:

- Higher residential land use and road density generally result in lower ozone concentrations. As the exhaust fumes of vehicles are the main anthropogenic source of oxides of nitrogen (NO), an increase in NO concentrations cause a decrease in ozone concentrations (i.e., NO-scavenging [11]). In fact, it is possible to see that the negative effect of residential area on ozone concentrations becomes stronger at 7am and 8am during the morning rush hour. In addition, because ozone cannot be formed without solar radiation, the negative effect of road density on ozone concentrations also becomes stronger at night.
- The effects of nature reserve land use type and industrial land use type have similar effects on ozone concentration over a day. Both land use types have negative effects on ozone concentrations during night, but positive effect on ozone concentrations during daylight. This is primarily because ozone is destroyed at night by NO at night, especially in industrial areas.
- The commercial land use type have positive effects on ozone concentrations. This implies that the ozone concentration is higher in downtown Singapore where most of the land is for commercial use. Because of the high NO concentration during daylight, the effect of commercial land use on ozone concentrations becomes weaker during daylight.

5 Conclusions

A statistical model has been proposed for predicting hourly ozone concentrations, and this is the first predictive model for ozone concentrations for Singapore. Both the mathematical formulation and testing results were presented in this paper. It

has been shown that the accuracy of the proposed model is comparable to that of many existing models. Our next step is to test the proposed model on a continual daily basis and explore the modeling of other pollutants under a similar framework.

References

[1] US Environmental Protection Agency, *Clean air act – air pollution prevention and control*, 2012.

[2] World Health Organization, *WHO air quality guidelines for particulate matter, ozone, nitrogen dioxide and sulfur dioxide – global update*, **WHO/SDE/PHE/OEH/06.02**, 2005.

[3] US Environmental Protection Agency, *Air quality index - a guide to air quality and your health*, **EPA-454/K-03-002**, 2003.

[4] Han, S., Bian, H., Feng, Y., Liu, A., Li, X., Zeng, F. & Zhang, X., Analysis of the relationship between O_3, NO and NO_2 in Tianjin, China. *Aerosol and Air Quality Research*, **11**, pp. 128–139, 2011.

[5] US Environmental Protection Agency, *Guidelines for developing an air quality (ozone and $PM_{2.5}$) forecasting program*, **EPA-456/R-03-002**, 2003.

[6] Davis, J.M. & Speckman, P., A model for predicting maximum and 8h average ozone in Houston. *Atmospheric Environment*, **33**, pp. 2487–2500, 1999.

[7] Wang, W., Lu, W., Wang, X. & Leung, A., Prediction of maximum daily ozone level using combined neural network and statistical characteristics. *Environment International*, **29**, pp. 555–562, 2003.

[8] Bogaert, P., Chistakos, G., Jerrett, M. & Yu, H.L., Spatiotemporal modelling of ozone distribution in the state of California. *Atmospheric Environment*, **43**, pp. 2471–2480, 2009.

[9] Carroll, R.J., Chen, R., George, E.I., Li, T.H., Newton, H.J., Schmiediche, H. & Wang, N., Ozone exposure and population density in Harris county, Texas. *Journal of the American Statistical Association*, **92**, pp. 392–404, 1997.

[10] Chistakos, G. & Vyas, V.M., A composite space/time approach to studying ozone distribution over eastern United States. *Atmospheric Environment*, **16**, pp. 2845–2857, 1998.

[11] Abraham, J.S. & Comrie, A.C., Real-time ozone mapping using a regression-interpolation hybrid approach applied to Tucson Arizona. *Journal of the Air and Waste Management Association*, **45**, pp. 914–925, 2004.

[12] Jiang, H.J., Schorgendorfer, A., Hwang, Y. & Amemiya, Y., A practical approach to spatio-temporal analysis. *Statistica Sinica*, **submitted**, 2014.

[13] Singapore Department of Statistics, *Population in brief*, 2013.

[14] Singapore Department of Statistics, *Monthly digest of statistics Singapore*, 2014.

[15] Xu, Y., Vizuete, W. & Serre, M., Characterization of air quality ozone model performance using land use regression model: An application in exposure assessment for epidemiology studies. *the 11th Annual CMAS Conference*, Chapel Hill, North Carolina, 2012.

[16] Air quality modeling technical support document. *US Environmental Protection Agency*, **EPA-454/R-11-009**, 2011.
[17] UK Department for Environment Food and Rural Affairs, *Evaluating the performance of air quality models*, 2010.
[18] Arasa, R., Soler, M.R., Olid, M. & Merino, M., A performance evaluation of MM5/MNEQA/CMAQ air quality modelling system to forecast ozone concentrations in Catalonia. *Journal of Mediterranean Meteorology and Climatology*, **7**, pp. 11–23, 2010.
[19] Eder, B., Kang, D., Mathur, R., Yu, S. & Schere, K., An operational evaluation of the ETA-CMAQ air quality forecast model. *Atmospheric Environment*, **40**, pp. 4894–4905, 2006.
[20] Pires, J.C.M., Alvim-Ferraz, M.C.M., Pereira, M.C. & Martins, F.G., Comparison of several linear statistical models to predict tropospheric ozone concentrations. *Journal of Statistical Computation and Simulation*, **82**, pp. 183–192, 2012.