EnvironmentAsia
The international journal published by the Thai Society of Higher Education Institutes on Environment

Available online at www.tshe.org/ea/index.html
EnvironmentAsia 10(2) (2017) 105-117
DOI

Tropical Ground-level Ozone Modeling in Urban Areas of Thailand

Jirateep Theapiriyakit, Sawarose Suwannakoot and Sitthichok Puangthongthub

Department of Environmental Science, Faculty of Science, Chulalongkorn University, Thailand

Abstract

This work applied multiple linear regression to model day time ozone (O_3day) and daily maximum ozone (O_3max) during 7:00-18:00 h by using its lagged day time ozone (O_3lagday) and lagged daily maximum ozone (O_3lagmax) in two urban areas (Bangkok and Samutprakarn) having different ozone precursor sources and urban topography, and in two seasons (wet and dry) having different meteorological variation by using a SAS® 9.2 software to analyze 16-year (1997-2012) data including 14,247,085 actual hourly measurements of O_3, CO, NO_2, and SO_2 and meteorological variables such as temperature (T), wind speed (WS), relative humidity (RH) and solar radiation (SR). The results showed that a daily O_3 average in Samutprakarn was 19.77 ± 11.30 ppb, higher than that in Bangkok at an average of 14.06 ± 8.74 ppb. In dry season, a daily O_3 average (17.71 ± 10.08 ppb) was higher and more fluctuating than that in wet season (10.84 ± 6.24 ppb). O_3day and O_3max metrics showed the strongest correlation with O_3lagday (*r* at *0.77* and *0.68* respectively) and O_3lagmax (*r* at *0.68* and *0.66* respectively) following by RH, SR, CO, and NO_2. The models with log-transformed O_3 outcomes (lnO_3day and lnO_3max) and a lnO_3lagday predictor provided better model R^2 values. The lnO_3day model had R^2 ranging from *0.644 - 0.692* and was commonly predicted by lnO_3lagday, CO7-18 and RHmin7-18. The lnO_3max model had R^2 ranging from *0.551 - 0.661* and was commonly predicted by lnO_3lagday, COmin7-18 and RHmin7-18. The validation R^2 values between observed O_3 and predicted O_3 using testing data ranged from *0.370* to *0.659.* The predicted values trended to follow lagged O_3 that was a dominant predictor. Model fitting could be improved in future if total VOCs data were available.

Keywords: ground-level ozone; ozone regression modeling; wet and dry seasons; meteorological effects

1. Introduction

Ground-level ozone (O_3) is the principal index substance of photochemical smog. It has been recognized as one of the principal pollutants that degrade air quality. It is therefore a strong oxidizing agent. High ozone level has a direct effect on human, vegetation and materials. Effects on human are headache, eyes irritation, upper respiratory system irritation and lung tissue damage as well as premature death (Wilson *et al*., 2017). Prediction and control of ozone concentrations can therefore manage and minimize effects of tropospheric ozone on human health and ecosystems (Nghiem and Oanh, 2004).

Ozone is produced when primary pollutant oxide of nitrogen (NO_X) and volatile organic compounds (VOCs) (often called non-methane hydrocarbons, NMHC) interact under the action of sunlight. NO_X and VOCs are referred to as ozone precursors (Abdul-Wahab *et al*., 2005; Singla *et al*., 2012). Both NO_X and VOCs are emitted from fuel combustion. As ozone is a secondary photochemical pollutant and it is not emitted directly into the air, this is the main reason why ozone is difficult to be controlled and modeled (Al-Alawi *et al*., 2008; Rajab *et al*., 2013; Sousa *et al*., 2007). Ozone formation is strongly related to meteorological parameters such as temperature, solar radiation, and wind speed (Barrero *et al*., 2006; Brönnimann and Neu, 1997; Rubio and Eduardo, 2014; Singla *et al*., 2012). Therefore, understanding how ozone variation can be predicted by its precursors and meteorological factors is helpful especially where urban source activities and season fluctuation are significantly different.

In Thailand, it has been reported that ground-level ozone trends to exceed ozone standards of 1-hour and 8-hour averages (100 ppb and 70 ppb respectively), especially in Samutprakarn following by Bangkok and its other vicinity provinces. In 2014, the highest 1-hour and 8-hour averages in Samutprakarn were reported at 233 ppb and 173 ppb respectively. These areas have made the national highest 1-hour ozone average exceeding the standard for the past 2 decades (Thai Pollution Control Department, 2015). In Thailand, Bangkok is the most populated urban area with the greatest economic growth having a lot of vehicles and traffic congestion with high ozone precursors of VOCs and NO_X emission causing atmospheric smog for less sun light intensity while Samutprakarn, a suburban area with different city topography with less atmospheric smog episodes has less severe traffic problem but has many more industrial plants emitting different species of VOCs. These differences in the ozone-precursor source activities and the city topography in two areas can affect ozone variation in different ways.

In order to model ozone levels, a multiple linear regression analysis (MLR) can be used to fit co-pollutants and meteorological predictors. However, the MLR approach can face a problem of multicollinearity as independent variables are well correlated to each other in a yielded regression equation. This can affect a variance of regression coefficients, if a modeler did not take this ambiguity into account. All of the above makes us interested in modeling day time ozone and daily maximum ozone in two specific areas (Bangkok and Samutprakarn) and in two seasons. We aimed to address correlation between ground-level ozone and its associated factors (ozone's co-pollutants and meteorological variables) and to develop ozone regression models.

2. Materials and Methods

2.1 Data acquisition

Hourly average measurements of ambient air pollutants and meteorological parameters from 1997 to 2012 were acquired from an 18-automatic monitoring station network operated by the Thai Pollution Control Department (PCD) in Bangkok city and Samutprakarn province. Locations of the monitoring station network are shown in Fig. 1. The monitoring network in Bangkok comprised of 13 automatic stations located in 10 districts and the Samutprakarn monitoring network comprised of 5 stations located in 3 districts. Data sets included 8 parameters that were 4 gaseous pollutants (O_3, CO, NO_2, and SO_2), and 4 meteorological parameters (temperature (T), wind speed (WS), relative (RH) and solar radiation (SR)). In urban city setting areas of Bangkok and Samutprakarn, wind may not dilute daily air pollutants well due to surface boundary obstacles or seasonal atmospheric condition factors such as low wind speed, termperature inversion, or no precipitaiton resulting in stagnant air. Thus, a generated independent variable, the previous day O_3, i.e., lagged daytime ozone (O_3lagday) or lagged daily O_3 maximum (O_3lagmax) is useful and can be applied for predicting its next day O_3. The lagged ozone has also been showed as one of the helpful predictors in another study. (Barrero *et al.*, 2006).

2.2 Data preparation

For all 8 acquired parameters from 18 stations in 16 years, we obtained 14,247,085 actual hourly average measurements out of 20,196,864 possible measurements as some were missing measurements. We then arranged hourly average measurements using a criterion of having at least 18 hourly measurements a day and yielded 616,840 daily observations of the 8 parameters. The yielded daily data were divided into 2 data sets for 2 purposes: 1) for fitting regression models using data from years 1997 to 2008; and 2) for testing model data set using data from years 2009 to 2012. Two O_3 metrics were modeled in this study, daily daytime ozone (O_3day) and daily maximum ozone (O_3max), in 5 sub analyses: 1) ALL including data from all 18 stations; 2) BKK including data from 13 stations in Bangkok; 3) SPK including 5 stations in Samutprakarn; 4) WET including data from all 18 stations only recorded in wet season between mid-May and mid-October; and 5) DRY including data from 18 stations only recorded in dry season between mid-October and mid-May. For an amount of valid observations available for modeling O_3day metric with its complete co-pollutant and meteorological predictors in ALL sub analysis, there were 27,453 station-day observations ready for model fitting, and 26,199 observations for BKK, 1,254 observations for SPK, 11,951 observations for WET and 15,502 observations for DRY. For O_3max metric analysis, there were 27,464 (ALL), 26,210 (BKK), 1,254 (SPK), 11,953 (WET) and 15,511 (DRY) observations.

2.3 Correlation analysis

Correlation coefficient (r) was estimated as a measure of the strength and the direction of a linear relationship between two parameters, for example a pair of each O_3 metric and its predictor or a pair of a predictor and other each predictor. The population value is given by p (rho) (Field, 2009; Kuzma and Bohnenblust, 2001; O'Rourke *et al.*, 2005). The sample correlation coefficient between 2 variables is defined in equation (1)

$$r = \sum(x_i - \bar{x})(y_i - \bar{y}) / \sqrt{(x_i - \bar{x})^2}\sqrt{(y_i - \bar{y})^2} \quad (1)$$

where x_i is the x value for observation i,
$\bar{x}$ is the mean x value,
y_i is the y value for observation i, and
$\bar{y}$ is the mean y value

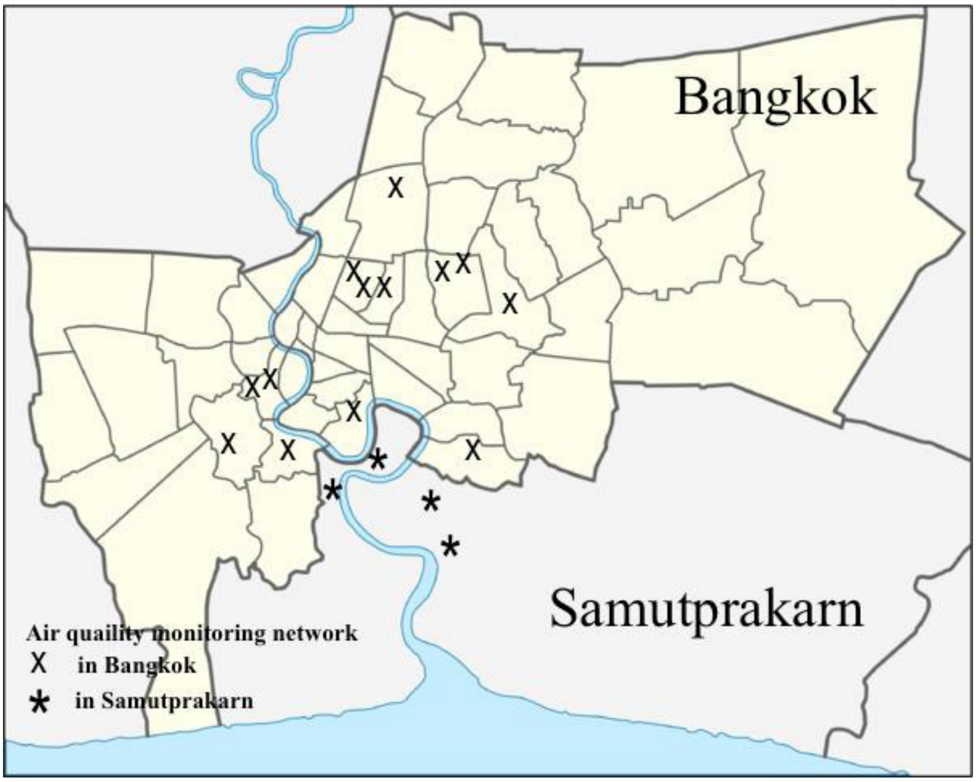


Figure 1. Location of an automatic ambient air quality monitoring station network of PCD in Bangkok and Samutprakarn

2.4 Regression modeling analysis

MLR analysis is one of the most widely used method to identify relationship between predictors and an outcome (Barrero *et al*., 2006). In this study, MLR was used to develop 10 prediction models to predict 2 ozone metrics with co-pollutants and meteorological parameters in 5 sub analyses using a forward stepwise method for variable introduction by the SAS® statistical software version 9.2 to determine a model coefficient of determination (R^2) and standardized regression coefficients (β) (O'Rourke *et al*., 2005). The regression model has a general form in equation (2)

$$O_3 = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_i x_i \quad (2)$$

where β_i is a regression coefficient of a predictor i (X_i), and
X_i is an independent predictor i.

However, MLR has a restriction when the predictors are highly correlated with each other in a regression equation, which is called "multicollinearity". Multicollinearity identifying that the predictors are highly intercorrelated so that little changes in the data values may cause large changes in the prediction model. In other words, a slight change in predictors results in a large change in regression coefficients (Field, 2009; Mustafa and Mohammed, 2012). Thus, it makes difficulty in identifying what major predictors are primary factors controlling ozone variation. However, we can detect how much the variance of each coefficient is inflated by multicollinearity with Variance Inflation Factors (VIF). VIF can be calculated to measure the variance magnitude of each coefficient that is inflated by multicollinearity. VIF values must be well below 10 to conclude that there is no collinearity within the model, i.e. VIF < 3 or VIF < 5 (Field, 2009). Thus, the VIF value was estimated for each model in this study.

3. Results and Discussion

3.1 Exploratory result

Hourly measurements during 1997-2012 from 18 monitoring stations of O_3, co-pollutants and meteorological parameters were averaged at hours 1-24 and their histograms were illustrated in Fig. 2. From the histograms, we noticed that the diurnal changes for these 8 parameters had 3 patterns. For pattern 1, it showed that O_3, and T were at a maximum level per day at 14:00 h, an hour after SR level at 13:00 h and WS was at a maximum level per day at 15:00 h (see Fig. 2(a)-2(d)). For pattern 2, it showed that only RH came with a minimum level per day at 14:00 h coinciding with a maximum level time of O_3 (see Fig. 2(e)). For pattern 3, it was noticed for all co-pollutants, CO, NO_2 and SO_2 that they had two maximum levels a day, in the morning at 8:00 h or 9:00 h before the O_3 maximum level time and in the evening at 20:00 h, 21:00 h or 22:00 h and had a minimum level a day at 14:00 h matching the O_3 maximum level time (see Fig. 2(f)-2(h)). These histograms implied a sort of positive or negative linear relationship between O_3 and other variables during daytime. For a positive relationship example, O_3 increased and decreased when SR, T and WS did or O_3 increase in the afternoon followed co-pollutants increase in the morning. For a negative relationship example, O_3 increased when RH decreased or O_3 was maximized when co-pollutants were minimized. These patterns were also reported in other O_3 studies (Barrero *et al*., 2006; He and Lu, 2012; Nghiem and Oanh, 2004). From these exploratory results, we noticed and applied these relationships to build different metrics of O_3 predictors of co-pollutants and meteorological variables. We formed their average, maximum or minimum metric during daytime (7:00-18:00 h) or in the morning before the O_3 maximum level time regarding to histograms in Fig. 2 (7:00-10:00 h for co-pollutants and 7:00-12:00 h for meteorological variables). For example, NO_2 had 4 metrics: NO_2 maximum during daytime (NO_2max7-18); NO_2 minimum during daytime (NO_2min7-18); NO_2 average during daytime ($NO_2$7-18); and NO_2 average in the morning ($NO_2$7-10). There were 26 metrics formed for 8 parameters.

Mean and standard deviation of 16-year daily average of 8 parameters from all 18 stations by sub analyses were shown in Table 1. A daily O_3 average in Samutprakarn was 19.77 ± 11.30 ppb which was higher than that in Bangkok at an average of 14.06 ± 8.74 ppb. This has been also reported by PCD as more industrial VOC sources, emitting high concentration, present in Samutprakarn. Difference of O_3 concentration in the two areas may not depend only on emissions, but also as a role of atmospheric transport and diffusion of O_3 precursors and O_3 itself. In dry season, a daily O_3 average (17.71 ± 10.08 ppb) was higher and more fluctuated than that in wet season (10.84 ± 6.24 ppb). PCD has also reported the same result during October-April due to clearer sky with low RH, low wind speed, lack of precipitation and temperature inversion resulting in high NO_2 and ozone (Thai Pollution Control Department, 2015). These atmospheric conditions well supported the photochemical ozone formation reaction and an accumulation of ozone precursors (Moustris *et al*., 2012; Özbay *et al*., 2011; Singla *et al*., 2012). We

Table 1. Daily average of O_3 and their predictors from 18 stations (1997-2012)

Sub analyses		O_3 (ppb)	NO_2 (ppb)	SO_2 (ppb)	CO (ppm)	T (°C)	RH (%)	SR (W/m^2)	WS (m/s)
ALL	n	64100	87169	89053	73956	85058	84019	50405	83080
	mean	14.781	23.243	5.206	0.882	28.758	73.567	146.390	1.750
	SD	9.293	13.309	4.938	0.698	3.536	12.231	59.902	3.166
BKK	n	56029	60797	62444	62432	65946	64968	43879	65891
	mean	14.062	25.307	5.270	0.956	28.736	73.071	140.450	1.732
	SD	8.736	13.633	4.195	0.717	3.881	12.429	57.338	3.502
SPK	n	8071	26372	26609	11524	19112	19051	6526	17189
	mean	19.771	18.484	5.056	0.477	28.833	75.258	186.332	1.818
	SD	11.295	11.158	6.347	0.382	1.909	11.367	61.460	1.186
WET	n	27345	37398	38084	31173	35880	35430	21414	35257
	mean	10.845	18.455	5.081	0.784	29.017	77.177	147.077	1.731
	SD	6.243	9.448	4.237	0.638	3.248	10.741	55.389	3.049
DRY	n	36755	49771	50969	42783	49178	48589	28991	47823
	mean	17.709	26.840	5.300	0.953	28.569	70.934	145.883	1.764
	SD	10.075	14.594	5.402	0.730	3.720	12.574	63.024	3.249

noticed that all meteorological parameters were not greatly different between wet and dry seasons except RH. For co-pollutants, we found NO_2, CO and SO_2 were greater in Bangkok, especially for NO_2. In dry season, NO_2 was noticeably higher than in wet season while CO and SO_2 were slightly higher. NO_2 was known to be more fluently emitted from vehicle sources when compared with CO and SO_2 (Abdul-Wahab *et al*., 2005). SR intensity was lower in Bangkok than in Samutprakarn by the reason that Bangkok has significantly higher level of particulate matter and smog that can give some shade thus reducing ground SR intensity depending on monitoring station location and sun position. Furthermore, Bangkok's topography and boundary surface may play an important role supporting more stagnant airflow or less mixing capacity, which inhibit ozone formation.

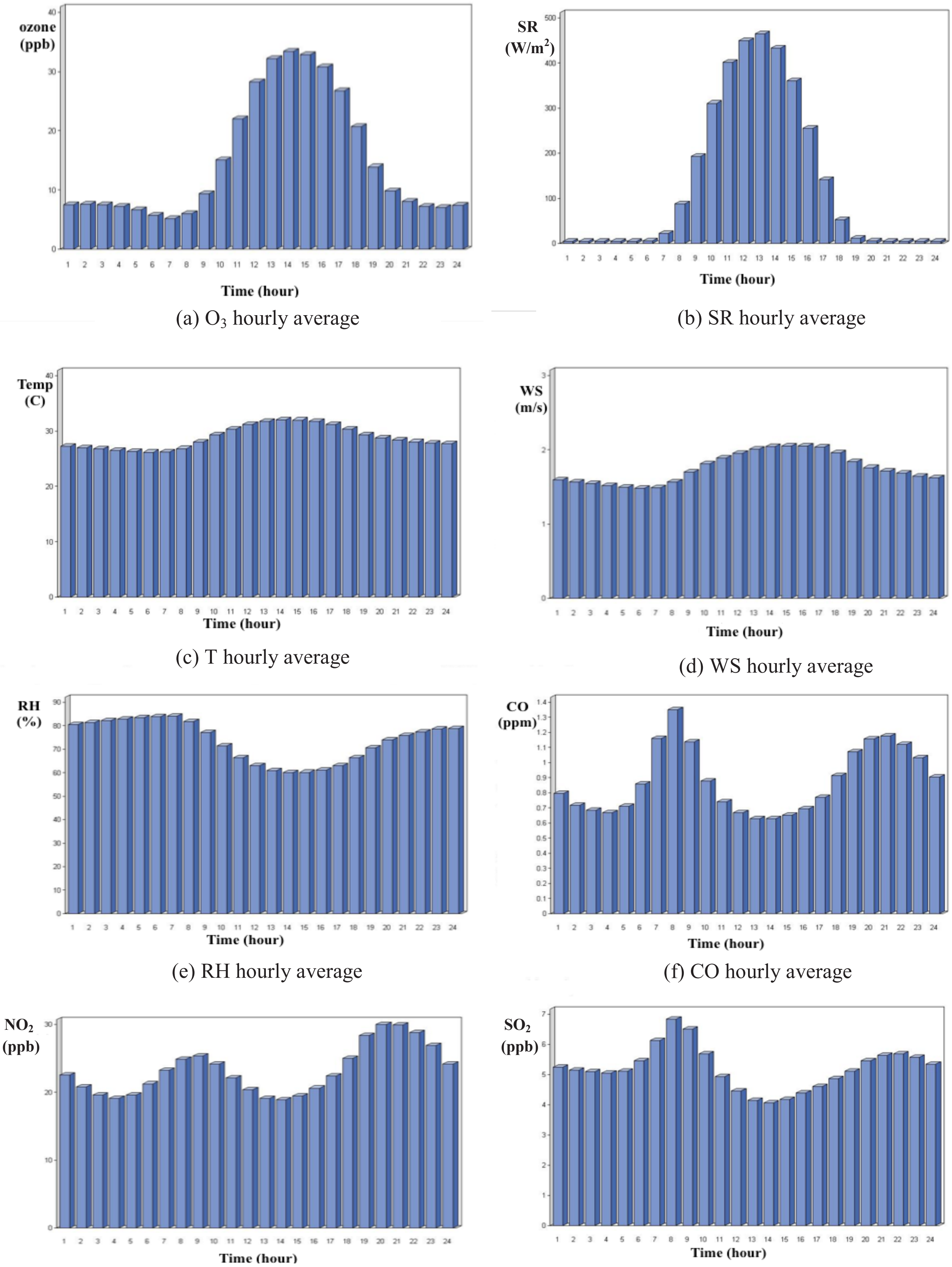
3.2 Correlation result

Pearson correlation *r* metrics of ozone and each predictor were presented in Table 2. Most correlation coefficients were statistically significant ($p<0.05$) except those marked with an asterisk symbol. This finding indicates a positive or a negative relationship between O_3 metrics and their predictors. Considering absolute value of *r*, we observed that O_3day and O_3max metrics showed the strongest correlation with O_3lagday (*r* at *0.77* and *0.68* respectively) and O_3lagmax (*r* at *0.68* and *0.66* respectively) following by RH, SR, CO, and NO_2. Previous day O_3 has showed a good correlation with current day O_3 due to an O_3 accumulation in the urban setting atmosphere. All RH metrics showed a negative correlation (*r* ranging from *-0.24* to *-0.42*) as it can decrease O_3

Table 2. Pearson correlation coefficients (*r*) between O_3 metrics and their predictors

Parameters	O_3day	O_3max
O_3lagmax	0.677	0.660
O_3lagday	0.768	0.683
CO7-10	-0.139	-0.069
CO7-18	-0.196	-0.120
COmax7-18	-0.171	-0.092
COmin7-18	-0.164	-0.119
$NO_2$7-10	0.109	0.193
$NO_2$7-18	-0.066	0.064
NO_2max7-18	0.049	0.175
NO_2min7-18	-0.158	-0.055
$SO_2$7-10	-0.050	-0.001*
$SO_2$7-18	-0.062	0.009
SO_2max7-18	-0.069	0.002
SO_2min7-18	-0.014	0.022
T7-12	-0.084	-0.080
T7-18	-0.004*	-0.005*
Tmax7-18	0.051	0.065
WS7-12	-0.016	-0.047
WS7-18	-0.016	-0.047
WSmax7-18	-0.006*	-0.034
RH7-12	-0.315	-0.239
RH7-18	-0.396	-0.317
RHmin7-18	-0.423	-0.375
SR7-12	0.185	0.120
SR7-18	0.235	0.146
SRmax7-18	0.233	0.162

Most values were statistically significant $\alpha = 0.05$.
* Not statistically significant at $\alpha = 0.05$



(a) O_3 hourly average

(b) SR hourly average

(c) T hourly average

(d) WS hourly average

(e) RH hourly average

(f) CO hourly average

(g) NO_2 hourly average

(h) SO_2 hourly average

Figure 2. O_3, SR, T, WS, RH, CO, NO_2 and SO_2 hourly average from 18 stations (1997-2012)

level by wet deposition (Abdul-Wahab *et al*., 2005; Al-Alawi *et al*., 2008; Goswami and Midya, 2016; Özbay *et al*., 2011). Ozone levels may be decreased by water as it can absorb ozone which is water-soluble. Because this significant effect of water on ozone formation, many studies showed the effect of water in the form of humidity and rainfall (Goswami and Midya, 2016; Özbay *et al*., 2011; Rajab *et al*., 2013). Relative humidity can affect ozone in wet and dry season differently regarding other meteorological factors. In dry season, there are lower relative humidity, higher temperature, stronger sunlight and less cloud whereas in wet season there are higher relative humidity, lower temperature, weaker sun light and cloudier sky (Moustris *et al*., 2012; Özbay *et al*., 2011; Rubio and Eduardo, 2014).

All SR metrics were slightly positive correlated with 2 O_3 metrics (*r* ranging from *0.12* to *0.24*) as they promoted the photochemical formation of O_3 (Moustris *et al*., 2012; Pires and Martins, 2011). There were weak correlations for CO and NO_2. All CO metrics were negatively correlated. CO as a product from an incomplete combustion can be implied as having a positive correlation to VOCs, an O_3 precursor, released from incomplete combustion of vehicle engines as well, thus CO could be declined when O_3 was increased (Singla *et al*., 2012). O_3 was noticed for positive correlations with $NO_2$7-10 and NO_2max7-18 as they were O_3 precursors and recorded in the morning before the level O_3 maximum period at 14:00 h while NO_2min7-18 showed a negative correlation as observed in the afternoon and already dissociated to form O_3 (He and Lu, 2012). For other parameters, SO_2, T and WS, they showed very weak or no correlation as their metrics had *r* values close to 0. For T, only Tmax7-18 showed small positive *r* values as it was read at 14:00 h right just at the level O_3 time. In contrast, T7-12 and T7-18 showed small negative *r* values (not statistically significant for T7-18) as they were averaged over long continuous hours which may not capture O_3 fluctuation well.

3.3 Regression modeling result

We analyzed ozone data by 2 approaches (log-transformed O_3 and non log-transformed O_3 models) for 2 O_3 outcome metrics (O_3day and O_3max) using 2 log-transformed lagged O_3 predictors (lnO_3lagday and lnlagO_3lagmax) leading to 2^3 or 8 examinations. However, the obtained results showed higher model R^2 only when using log-transformed O_3 outcomes and predictors because the MLR method predicts better with variables having normal distribution. The natural logarithm can transform O_3 and adjust its distribution closer to the normal distribution. Forty models from 8 examinations and 5sub-analysses were fit with lagged O_3 predictors (lnO_3lagday or lnO_3lagmax) and showed that the lnO_3lagday predictor gave better model R^2 than lnO_3lagmax did. Therefore, only results of 10 final models predicting the natural logarithm transformed O_3day and O_3max outcomes using only the lnO_3lagday predictor were discussed here. Their 10 results of model R^2 and regression standardized coefficients (β) of lnO_3lagday and lnO_3lagmax models were demonstrated in Tables 3 and 4 respectively. Other results of log-transformed O_3 outcomes but with a lnO_3lagmax predictor and other results of non-transformed O_3 outcomes with a lnO_3lagday or a lnO_3lagmax predictor were not presented. To finalize a number of predictors in each model, we accounted for the multicollinearity problem by limiting a VIF value < 3 (Mustafa and Mohammed, 2012). We tested and learnt that if we choose a VIF value greater than 10, it will result in higher model R^2 and will include more predictors but will likely be bias in a regression coefficient, while if we choose a VIF value < 1, it will show a lower model R^2 but not much different from a R^2 value when tested at VIF value < 3.

As can be seen in Table 3, lnO_3day model R^2 and β coefficients of ALL and BKK sub analyses were almost identical as most observations in ALL data set were from BKK data set. The lnO_3day model R^2 ranged from *0.644-0.692* or about *64.40-69.20* % of daytime ozone was interpreted by its selected predictors. In forward stepwise method, the predictor of 5 sub analyses of the lnO_3day model that usually was introduced in the first order was lnO_3lagday (β ranging from 0.56-0.66) as to be the best predictor, which is similar to other studies (Barrero *et al*., 2006; Moustris *et al*., 2012; Pires and Martins, 2011). By only lnO_3lagday alone, it can provide high model R^2 ranging from *0.557 - 0.609* (not shown in the Table 3). The second predictor usually introduced was RHmin7-18. The others in subsequent orders were CO7-18 and SR7-12. Considering absolute values of β, in addition to lnO_3lagday and RHmin7-18 as major predictors in all sub analyses already, we can say that CO7-18 also showed a strong ability to predict lnO_3day in ALL, BKK, and WET. This was simply because CO and VOCs (O_3 precursors) were emitted simultaneously from same traffic sources. Other predictors with smaller β values found in some sub analyses were NO_2max7-18 following by SR7-18. For SO_2min7-18 and T7-12, they had tiny β values but were present in all sub analyses.

We saw that predictors included in the final lnO_3day models for each sub analysis were different but found that they commonly included lnO_3lagday, CO7-18 and RHmin7-18 likely associating with a day-to-day accumulation of O_3, a correlation between CO and VOC from incomplete traffic combustion and a wet deposition of O_3 onto atmospheric water respectively (Singla *et al*., 2012). Commonly chosen predictors in BKK and SPK were lnO_3lagday and RHmin7-18, while uncommonly chosen predictors were CO7-18, COmin7-18, T7-12, WS7-18 and SR7-12. For WET and DRY, commonly chosen predictors were lnO_3lagday, RHmin7-18 and NO_2min7-18, whereas uncommonly chosen predictors were CO7-18, NO_2max7-18 and SR7-18.

For lnO_3max model R^2 and β coefficients, they were shown in Table 4. The lnO_3max model R^2 ranged from *0.551-0.661* meaning that 55.10-66.10 % of ozone maximum was interpreted by the selected predictors listed. The predictor that usually was introduced in the first order was lnO_3lagday (β ranging from 0.42-0.64) as to be the best predictor which is similar to lnO_3day model and another study (Barrero *et al*., 2006). Considering absolute values of β, we noticed that the strongest positive predictor of lnO_3max model was lnO_3lagday (β ranging from 0.42 to 0.64), following by COmin7-18 (not included in DRY) and RHmin7-18. The strength order of the last two predictors was different from lnO_3day model. Other important predictors were also

Table 3. LnO_3day model standardized coefficients predicted with lnO_3lagday

LnO_3day model	Regression coefficients (β)				
	ALL	BKK	SPK	WET	DRY
R^2 model	0.676	0.676	0.692	0.644	0.657
Predictors					
lnO_3lagday	0.655	0.653	0.562	0.597	0.625
CO7-10	-	-	-	-	-
CO7-18	-0.180	-0.182	-	-0.239	-0.033
COmax7-18	-	-	-0.100	-	-
COmin7-18	-	-	0.131	-	-
$NO_2$7-10	-	-	-	-	-
$NO_2$7-18	-	-	-	-	-
NO_2max7-18	0.090	0.090	-	0.213	-
NO_2min7-18	-	-	-	-0.139	-0.130
$SO_2$7-10	-	-	-	-	-
$SO_2$7-18	-	-	-	-	-
SO_2max7-18	-0.009	-0.009	-	-	-0.020
SO_2min7-18	0.026	0.025	0.053	0.061	0.030
T7-12	-0.064	-0.062	-0.132	-0.046	-0.068
T7-18	-	-	-	-	-
Tmax7-18	-	-	-	-	-
WS7-12	-0.076	-0.084	-	-	-0.087
WS7-18	-	-	-0.115	-0.048	-
WSmax7-18	0.054	0.064		-	0.057
RH7-12	-	-	-	-	-
RH7-18	-	-	-	-	-
RHmin7-18	-0.172	-0.110	-0.229	-0.143	-0.206
SR7-12	0.107	0.171	-	-	-
SR7-18	-	-	0.081	0.150	0.090
SRmax7-18	-	-	-	-	-

different including WS7-18 and NO_2max7-18 (not included in SPK and DRY). As a level of O_3 maximum can be more sensitive and fluctuating than daytime O_3, from the mentioned predictors we noticed these predictors were in sensitive metrics as well, e.g. COmin7-18 and NO_2max7-18. For WS7-18 with small β values, it was also introduced here as likely having an influence of diluting the O_3 maximum level because the wind speed can affect ozone peak by transport process.

For O_3max, the predictors included in the final lnO_3max models were different from the final lnO_3day models but we saw that they commonly included lnO_3lagday, COmin7-18 and RHmin7-18. This finding can be explained similarly to those found in lnO_3day models as mentioned earlier. Common predictors found in BKK and SPK were lnO_3lagday, COmin7-18 and RHmin7-18, while others were COmax7-18, NO_2max7-18, T7-12, WS7-18 and SR7-12. For WET and DRY, similar predictors were lnO_3lagday, while others were COmin7-18, NO_2max7-18, WS7-12 and WS7-18.

3.4 Modeling validation result

First, for model validation we checked through the standardized residuals values of all 10 models. They all showed similar distribution, closely to normal distribution with a mean and a standard deviation values were 0.00 and 1.00 respectively. Fig. 3 showed an example plot of standardized residuals of predicted and observed values estimated from the lnO_3day model of DRY sub analysis.

Second, the coefficient of determination (R^2) values between the observed O_3 and the predicted O_3 for the testing data set were estimated from *0.370* to *0.659*. From 10 plots, lnO_3day showed higher R^2 values than lnO_3max (averaged R^2 at *0.599* and *0.435* respectively). The lnO_3day model and lnO_3max model from the SPK sub analysis gave the highest R^2 at *0.659* and *0.480* respectively in 2 scatter plots as shown in Fig. 4(a)-4(b). In contrast, the lnO_3day model and lnO_3max model in WET sub analyses gave the lowest R^2 at *0.509* and *0.370* respectively (scatter plots of other 8 models were not shown). This finding

Table 4. LnO_3max model standardized coefficients predicted with lnO_3lagday

LnO_3max model	Regression coefficients (β)				
	ALL	BKK	SPK	WET	DRY
R^2 model	0.595	0.595	0.661	0.551	0.561
Predictors					
lnO_3lagday	0.639	0.639	0.422	0.598	0.604
CO7-10	-	-	-	-	-
CO7-18	-	-	-	-	-0.033
COmax7-18	-	-	-0.102	-	-
COmin7-18	-0.170	-0.176	0.130	-0.248	-
$NO_2$7-10	-	-	0.180	-	-
$NO_2$7-18	-	-	-	-	-
NO_2max7-18	0.156	0.159	-	0.173	-
NO_2min7-18	-	-	-	-	-0.086
$SO_2$7-10	-0.022	-0.023	-	-	-
$SO_2$7-18	-	-	-	0.049	-
SO_2max7-18	-	-	-	-	-
SO_2min7-18	0.032	0.034	0.033	0.039	0.026
T7-12	-0.068	-0.062	-0.104	-	-0.079
T7-18	-	-	-	-	-
Tmax7-18	-	-	-	0.086	-
WS7-12	-	-	-	-0.174	-
WS7-18	-0.092	-0.086	-0.250	-	-0.132
WSmax7-18	-	-	-	0.096	-
RH7-12	-	-	-	-	-
RH7-18	-	-	-	-	-
RHmin7-18	-0.127	-0.125	-0.169	-0.079	-0.177
SR7-12	0.101	0.102	-	0.055	0.074
SR7-18	-	-	0.083	-	-
SRmax7-18	-	-	-	-	-

was associated with their lowest model R^2 for WET sub analysis as it was difficult to predict O_3 due to its fluctuation by strong moonsoon wind, precipitation, and atmospheric turbulance. Our WET validation results under an influence of high humidity were not found in other cold, dry area investigations (Barrero *et al*., 2006; Moustris *et al*., 2012; Pires and Martins, 2011).

Third, Fig. 5 showed an example of the predicted and observed O_3 values from lnO_3day model of the ALL sub analysis during a testing period in years 2009-2012 (Fig. 5(a)) and in September-December 2012 (Fig. 5(b)). Both observed and predicted dotted lines seemed to fit together fairly well except at extreme values. The model had some difficulties in predicting high ozone concentrations because they were dramatically fluctuating while the lower observed ozone levels were more possible to predict and were resembled clearly with predicted values. In addition, these predicted values were not lower or higher than a range of the observed values. Effected by an influence of lagged O_3, the most predicted values likely moved slightly to a right side of the observed values (see Fig. 5(b)) as the predicted ones trended to follow lagged O_3 that was a dominant predictor (Barrero *et al*., 2006). These three validations were fairly in agreement so we can generally accept the performance of the models.

However our obtained model R^2 values of both models (roughly *0.6-0.7*) were lower than other studies (roughly *0.7-0.8*) conducted in cold cities (Barrero *et al*., 2006; Moustris *et al*., 2012; Pires and Martins, 2011). This may be due to unlike meteorlogy condition of cold-dry vs. hot-humid atmosphere making dissimilarly fluctuating O_3 levels and due to different predictors used in those studies. In addition, some models may have less accuracy probably due to some parameters in the model having no linear relationship with ozone or not including important variables that effected O_3 concentrations such as total VOCs, a boundary layer, wind direction, natural O_3 precursor sources, etc. Model fitting could be improved in future study if other such important predictors that relating to O_3 formation were introduced to the models. Unfortunately, total VOCs has not been monitored comprehensively in Bangkok and Samutprakarn.

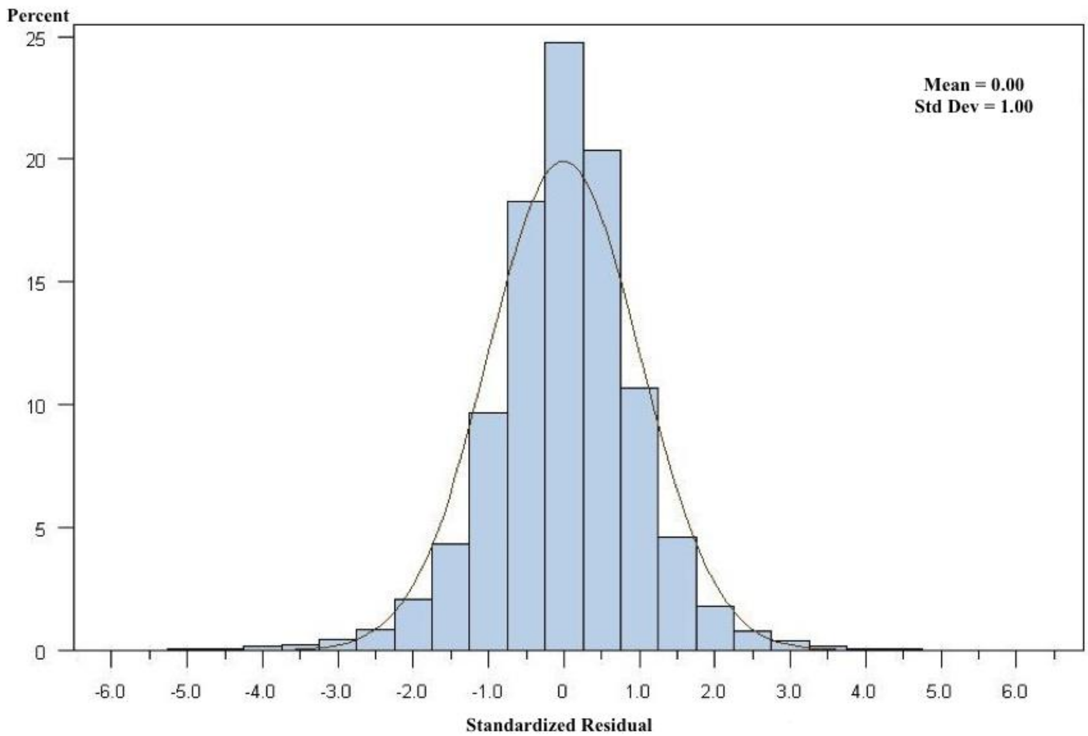
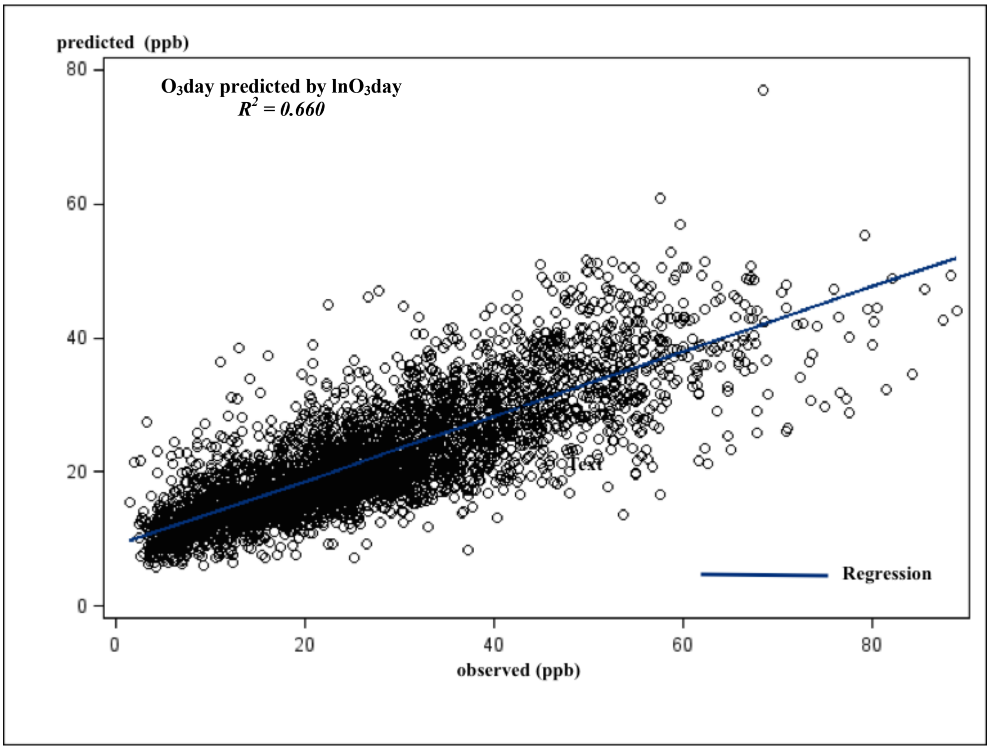
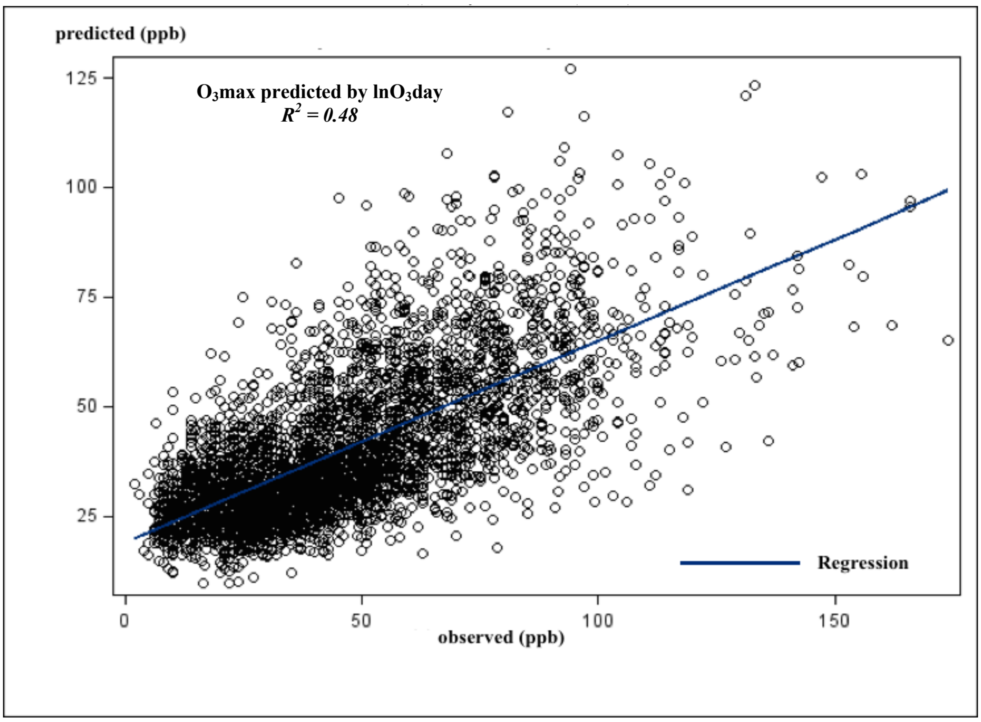


Figure 3. Distribution of standardized residuals of Daytime O_3 (DRY)

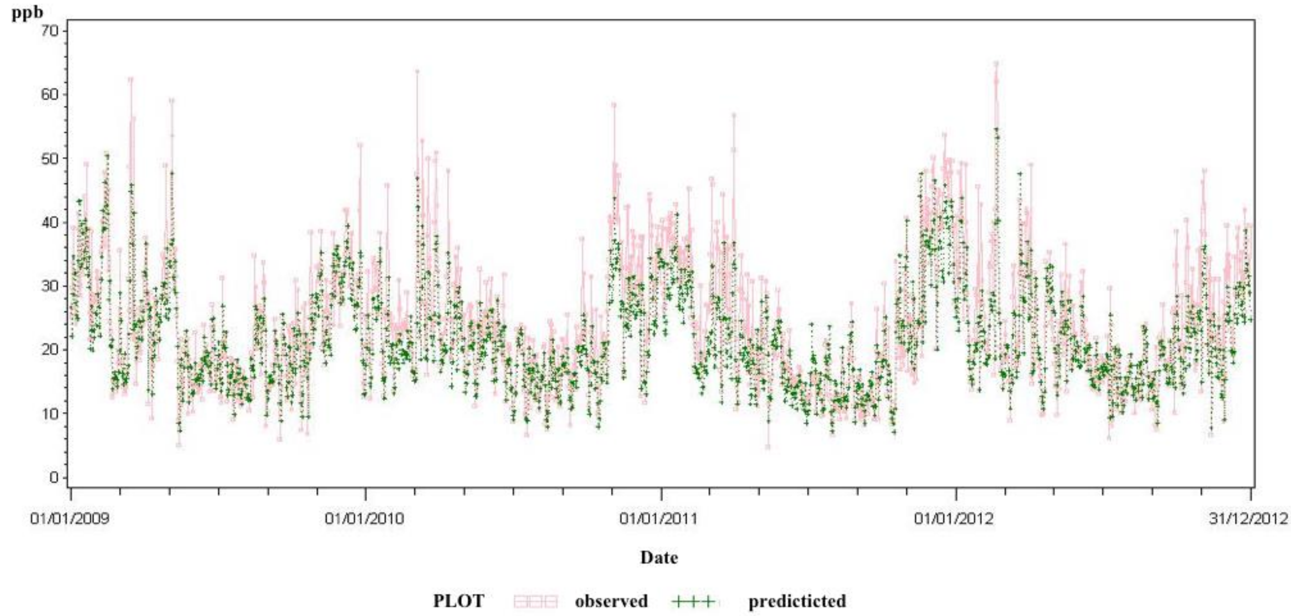


(a) Daytime O_3 (SPK)

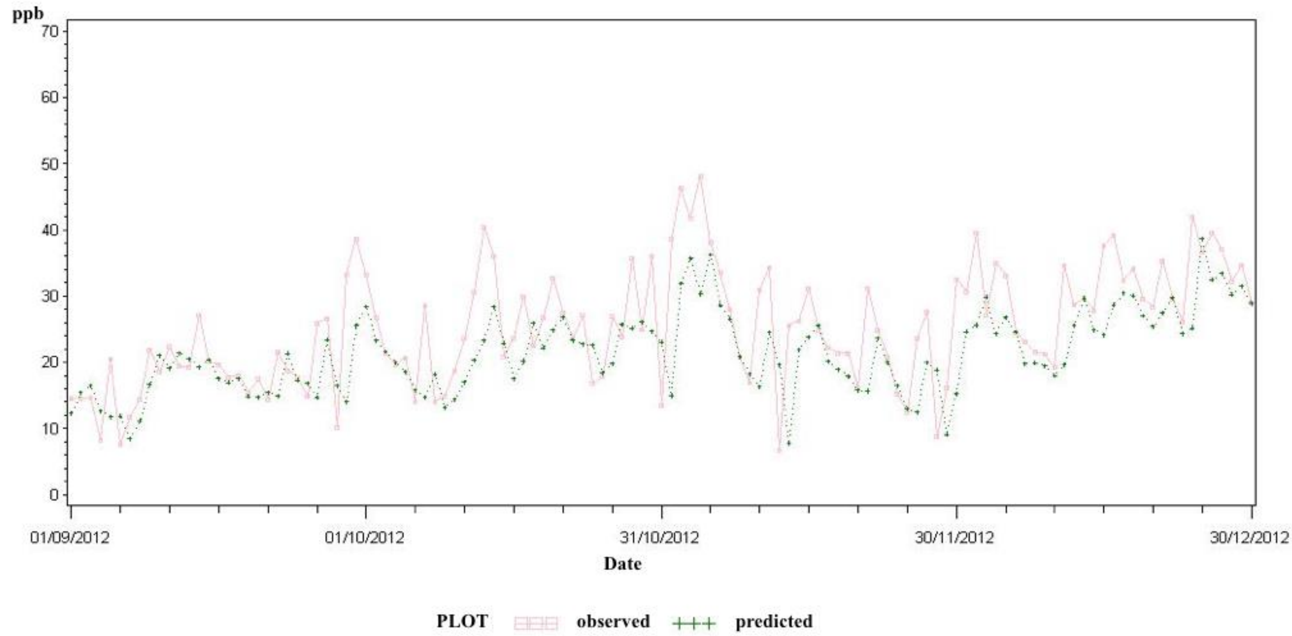


(b) Daily maximum O_3 (SPK)

Figure 4. Scatter plots of observed and predicted O_3



(a) Observed vs. predicted O_3 (2009-2012)



(b) Observed vs predicted O_3 (Sep-Dec 2012)

Figure 5. Daytime O_3 (ALL) comparison of observed vs. predicted O_3

4. Conclusions

This study presented the multiple linear regression method to forecast O_3day and O_3max. The highest daily ozone averages were found in Samutprakarn and in dry season. O_3day and O_3max were strongly correlated with daytime lagged O_3 (positive) following by RH (negative) and SR (positive). By applying natural logarithm transformation of O_3 outcomes along with a lnO_3lagday predictor, it showed model R^2 improvement. The lnO_3day models gave better R^2 than lnO_3max models in all sub analyses. In both lnO_3day and lnO_3max models, lnO_3lagday was the strongest positive predictor and RHmin7-18 was the strongest negative predictor. SPK sub analysis showed highest R^2 in lnO_3day model and in lnO_3max model, *0.6916* and *0.6612* respectively. An application of the obtained models can be used to predict tomorrow O_3 levels for comunity health alert or future O_3 variability regarding to meterological variable fluctuation due to climate change.

Acknowledgements

The authors would like to thank the Thai Pollution Control Department for providing air pollutant and meteorological data. This work was partially supported by the Sci Super III Fund (Ratchadapiseksomphot Endowment Fund), Chulalongkorn University.

References

Abdul-Wahab SA, Bakheit CS, Al-Alawi SM. Principal component and multiple regression analysis in modelling of ground-level ozone and factors affecting its concentrations. Environmental Modelling and Software 2005; 20(10): 1263-71.

Al-Alawi SM, Abdul-Wahab SA, Bakheit CS. Combining principal component regression and artificial neural networks for more accurate predictions of ground-level ozone. Environmental Modelling and Software 2008; 23(4): 396-403.

Barrero MA, Grimalt JO, Cantón L. Prediction of daily ozone concentration maxima in the urban atmosphere. Chemometrics and Intelligent Laboratory Systems 2006; 80(1): 67-76.

Brönnimann S, Neu U. Weekend-weekday differences of near-surface ozone concentrations in Switzerland for different meteorological conditions. Atmospheric Environment 1997; 31(8): 1127-35.

Field AP. Discovering statistics using SPSS: (and sex and drugs and rock 'n' roll) SAGE Publications: Thousand Oaks, Calif. ; London. 2009; 135-39.

Goswami S, Midya SK. Seasonal variation of daily total column ozone (TCO), its depletion and formation role on surface temperature and rainfall over Chennai, India. Journal of Indian Geophysical Union 2016; 20(1): 101-11.

He Hd, Lu WZ. Decomposition of pollution contributors to urban ozone levels concerning regional and local scales. Building and Environment 2012; 49: 97-103.

Kuzma JW, Bohnenblust SE. Correlation and linear regresion. Mayfield Publishing Company: Singapore. 2001; 13.

Moustris KP, Nastos PT, Larissi IK, Paliatsos AG. Application of multiple linear regression models and artificial neural networks on the surface ozone forecast in the greater Athens area, Greece. Advances in Meteorology 2012; 2012: 1-8.

Mustafa YA, Mohammed SJ. Measurement of ground level ozone at different locations. American Journal of Environmental Sciences 2012; 8(3): 311-21.

Nghiem LH, Oanh NTK. Prediction of maximum daily ozone levels using neural network model in Bangkok, Asian Institute of Technology. 2004.

O'Rourke N, Hatcher L, Stepanski EJ, Hatcher L. A step-by-step approach to using SAS for univariate & multivariate statistics. Wiley-Interscience, New York. 2005; 277-85.

Özbay B, Keskin GA, Doğruparmak ŞÇ, Ayberk S. Multivariate methods for ground-level ozone modeling. Atmospheric Research 2011; 102(1-2): 57-65.

Pires JCM, Martins FG. Correction methods for statistical models in tropospheric ozone forecasting. Atmospheric Environment 2011; 45(14): 2413-17.

Rajab JM, MatJafri MZ, Lim HS. Combining multiple regression and principal component analysis for accurate predictions for column ozone in Peninsular Malaysia. Atmospheric Environment 2013; 71: 36-43.

Rubio MA, Eduardo LG. Temperature as thumb rule predictor of ozone levels in Santiago De Chile ground air. Journal of the Chilean Chemical Society 2014; 59(2): 2427-31.

Singla V, Pachauri T, Satsangi A, Kumari KM, Lakhani A. Surface ozone concentrations in Agra: links with the prevailing meteorological parameters. Theoretical and Applied Climatology 2012; 110(3): 409-21.

Sousa SIV, Martins FG, Alvim-Ferraz MCM, Pereira MC. Multiple linear regression and artificial neural net works based on principal components to predict ozone concentrations. Environmental Modelling and Software 2007; 22(1): 97-103.

Thai Pollution Control Department. Thailand state of pollution report 2014 text and journal publication. Bangkok, Thailnd. 2015; 13-14.

Wilson A, Reich BJ, Nolte CG, Spero TL, Hubbell B, Rappold AG. Climate change impacts on projections of excess mortality at 2030 using spatially varying ozone-temperature risk surfaces. Journal of Exposure Science and Environmental Epidemiology 2017; 27(1): 118-24.

Received 4 April 2017
Accepted 24 May 2017

Correspondence to
Dr. Sitthichok Puangthongthub
Department of Environmental Science,
Faculty of Science
Chulalongkorn University,
Bangkok 10330
Thailand
Tel/Fax: +662 218-5198
E-mail: Sitthichok.p@chula.ac.th