Journal of Environmental Science International
25(2); 191~216; February 2016

pISSN: 1225-4517 eISSN: 2287-3503
http://dx.doi.org/10.5322/JESI.2016.25.2.191

ORIGINAL ARTICLE

Assessing Factors Linked with Ozone Exceedances in Seoul, Korea through a Decision Tree Algorithm

Sun-Kyoung Park*

Hanyang Cyber University, Seoul 04763, Korea

Abstract

Since prolonged exposure to elevated ozone (O_3) concentrations is known to be harmful to human health, appropriate control strategies for ozone are needed for the non-attainment area such as Seoul, Korea. The goal of this research is to assess factors linked with the 1-hour ozone exceedance through a decision tree model. Since ozone is a secondary pollutant, lag times between ozone and explanatory variables for ozone formation are taken into account in the model to improve the accuracy of the simulation. Results show that while ozone concentrations of the previous day and NO_2 concentrations in the morning are major drivers for ozone exceedances in the early afternoon, meteorology plays more important role for ozone exceedances in the late afternoon. Results also show that a selection of lag times between ozone and explanatory variables affect the accuracy of predicting 1-hour ozone exceedances. The result analyzed in this study can be used for developing control strategies of ozone in Seoul, Korea.

Key words : Ozone, Exceedance, Decision tree algorithm, NO_2, Control strategy

1. Introduction

Epidemiological studies have shown that ground -level ozone (O_3) concentrations have adverse health effects. In addition, ozone is one of the critical components of photochemical smog, which decreases visibility and causes hazardous conditions for human activities. Seoul, the capital of South Korea, is one of the cities that needed control of ozone. Several previous studies have found steady increases in ozone concentrations in Seoul.

The number of cases in which ozone and NO_2 concentrations in Seoul exceeded the standard was analyzed using hourly data collected at 31 monitors between 1990 and 2000(Kim et al., 2005). A separate study showed that on-road mobile sources were a major source of NO_x in Seoul(Pandey et al., 2008). The mixing height of atmosphere and ground-level ozone concentrations measured in Seoul between May 24, 2000, and May 26, 2000, showed that the mixing height at night affected the ground level ozone concentrations the next day (Kim et al., 2007).

Ozone concentrations were simulated using several statistical models. A fuzzy expert system was applied to predict a high ozone concentration, and a neural network model was used to predict the daily maximum concentrations; both models were successful(Heo and Kim, 2004). In Austria, an autoregressive exogenous model was developed to

Received 18 January, 2016; **Revised** 31 January, 2016;
Accepted 15 February, 2016
*__Corresponding author__ : Sun-Kyoung Park, Hanyang Cyber University, Seoul 04763, Korea
Phone: +82-2-2290-0432
E-mail: helena@hycu.ac.kr

predict ozone levels in the country's east(Bauer et al., 2001). A linear regression model and a principal component analysis were applied to predict day as well as night ozone concentrations in Oman (Abdul-Wahab et al., 2005). Ozone concentrations have also been predicted in two ways by a feed forward artificial neural network model. One method used all the independent variables in the analysis; the other used only the principal components of the variables(Sousa et al., 2007). The results from principal components were better than those using all the independent variables because in the former data collinearity could be reduced.

A separate study compared the capabilities of linear time series, artificial neural networks, and fuzzy models in predicting high-level ozone episodes in Santiago, Chile(Jorquera et al., 1998). All three models performed similarly in forecasting ozone episodes with success rates of 70~95percent, but the fuzzy model was the most reliable because it had the lowest rate of false positives. Schlink et al.(2003) reported that neural network and generalized additive models as the best comprise for ozone forecasting after a model inter-comparison exercise using 15 different statistical techniques.

Among statistical methods, a decision tree algorithm, a nonparametric data-driven analysis method, and was also found to be suitable for predicting ground-level ozone levels. Gardner and Dorling(2000) compared between linear regression, regression tree and multilayer perception neural network models to predict hourly surface ozone concentrations in the UK. They indicated that regression tree models are more readily physically interpretable although multilayer perception models are more accurately capture the underlying relationship between the meteorological and temporal predictor variables and hourly ozone concentrations. Burrows et al.(1995) found that the summer season maximum surface ozone for the Vancouver, Montreal and Atlantic regions of Canada was successfully predicted using the classification and regression trees(CART). A decision tree was also applied to identify controlling factors of ground-level ozone levels over southwestern Taiwan(Chu et al., 2012). Since a decision tree algorithm was successful to classify the dependent variables, it was also applied to predict ozone exceedances in the Baton Rouge, Louisiana USA(Rohli et al., 2003).

The goal of this study is to assess factors linked with 1-hour ozone exceedances in Seoul, Korea. Ozone levels were mainly predicted using the photochemical air quality models, which were known to simulate air pollutant levels quite accurately. However, the accuracy of air quality models depended largely on the accuracy of the emissions, one of the important input variables of the air quality model. Thus, if the emissions were not accurately prepared, the accuracy of the results simulated through the air quality models was limited as well. In order to improve the accuracy of emissions, data assimilation process have been also used in previous studies(Marmur et al., 2006; Park et al., 2006a; Park et al., 2006b; Park et al., 2013; Park and Russell, 2013). Thus, as an alternative choice, a statistical model that uses the measured pollutant concentrations to predict ozone exceedances was used in this paper.

Among statistical models, a decision tree model was selected since it was easily interpretable to quantify the importance of factors affecting ozone formation. Since ozone was a secondarily formed pollutant, the lag time between ozone and explanatory variables for ozone formation was taken into account to improve the accuracy of predicting the ozone exceedances. Thus, this research was conducted in two steps. In the first step, lag times between ozone and the variables for its formation were investigated. Then, in the second step, these lag times were used to assess factors affecting 1-hour ozone exceedance (the hourly ozone concentration > 0.1 ppm) through a decision tree algorithm. The results analyzed in this study can be further used as a reference to develop control strategies for ground-level ozone concentrations in Seoul.

2. Materials and Methods

2.1. Data

The data used included hourly ozone [ppm], NO_2 [ppm], temperature [^oC], solar irradiance [W/m^2], relative humidity [%], and wind speed [m/sec] monitored at 25 stations in Seoul, Korea, between 2005 and 2012 (Fig. 1). Although volatile organic carbon (VOC) also plays an important role in ozone formation, hourly VOC data were not available. In their absence, the analysis relied on, the NO_2 concentrations and meteorological variables. The data were collected from the Air Quality Information Center in Seoul(Kim et al., 2005). UV photometric ozone analyzers and the chemi-luminescent method were used, respectively, to measure hourly ozone and NO_2 concentrations. Data not meeting quality standards were assumed to be missing.

2.2. Decision tree algorithm

A decision tree algorithm uses a tree structure to represent the decision rules used to classify or to predict dependent variables(Moon et al., 2012). The independent variables were selected based on a Gini index and were used to structure the decision tree(Breiman et al., 1984). The analysis was performed using CART (Classification and Regression Tree) software, which has been widely applied to predict and classify data(Chu et al., 2012).

CART is a binary tree algorithm in which data is divided into groups based on the value of independent variables that maximize the dissimilarity between groups. The decision tree grows through the successive divisions represented by the hierar -chical tree structure following the “if-then” rule. The growth of the tree stops when further division does not increase homogeneity in the nodes. The accuracy of the prediction does not always improve when the number of independent variables increases if the independent variables are correlated with each other.

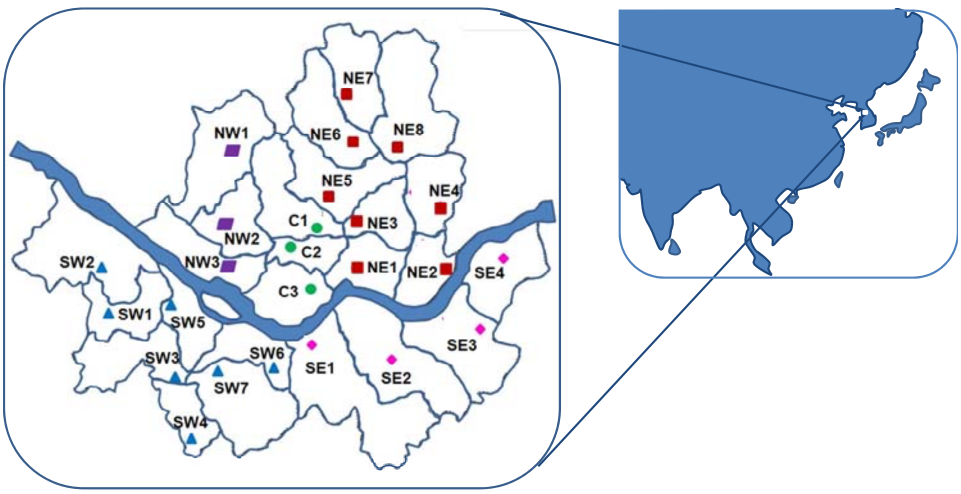

Fig. 1. Locations of monitoring stations in Seoul, Korea.

3. Results and Discussion

3.1. Spatial and temporal variations of air pollution and meteorology

The regional average ozone concentration between 2005 and 2012 was 0.019 ppm, with the highest and the lowest average ozone concentrations of 0.022 ppm (NW1) and 0.016 ppm (SE2), respectively. NO_2 concentrations were between 0.040 ppm (NE3) and 0.028 ppm (NW1), with a regional average concentration of 0.035 ppm. A spatial variation in meteorology was not apparent in the study area. The average temperature at each station was between 12.6^{o}C (SE3) and 13.9^{o}C (NE4), with a regional average of 13.3^{o}C; average solar irradiance was between 101.2 W/m^2 and 136.3 W/m^2, with an average of 124.5 W/m^2. Average relative humidity was between 56.6% (SW3) and 65.1% (NE2), with an average of 60.1%; average wind speed was between 0.9 m/sec and 2.0 m/sec, with an average of 1.3 m/sec (Table 1).

Annual average ozone varied little between 2005 and 2012. NO_2 concentrations increased slightly from 2005 to 2007 (Fig. 2). After 2007, it decreased until 2012. An increase in NO_2 concentration before 2007 was also observed in other areas of East Asia(He et al., 2007; Richter et al., 2005). A rapid increase in diesel fueled automobiles was suspected as a major cause of the increased NO_2 levels in Korea(Kim et al., 2005). The decrease in NO_2 concentrations after 2007 could be partly due to an achievement of a pollution abatement policy called the “Total Air Pollution Load Management System” that was established in 2005. The policy included a plan to reduce annual anthropogenic NO_x emissions by 2014 by as much as 53% of the annual anthropogenic emissions in 2001(Kim et al., 2013). Meteorological conditions apparently had little annual variations. Annual average temperature also varied little. The lowest temperature of 12.5^{o}C occurred in 2012 and the highest of 13.8^{o}C in 2007. The annual average solar irradiance was lowest in 2005 (102.8 W/m^2) and highest in 2009 (136.9 W/m^2). Annual average relative humidity was between 57.7% (2012) and 62.3% (2010), and annual average wind speed was between 1.26 m/sec (2006) and 1.40 m/sec (2012).

The monthly average ozone concentration was highest in June and lowest in December, reflecting that high temperatures favored the formation of ozone(Fig. 2). However, an abrupt decrease in monthly average ozone concentrations was observed in July despite the month’s high temperatures. This decrease was because heavy rainfall reduced solar irradiance. Monthly average rainfall in Seoul in July from 2005 to 2012 was 571 mm, more than 37% of the total annual rainfall(Qianet al., 2002). Monthly averages for NO_2 concentrations were higher in winter and lower in summer. Because a large percentage of NO_x emissions in Seoul originate from fuel burned for heating, NO_2 concent -rations increase in winter(Wang and Mauzerall, 2004). The monthly average temperature was highest (26.6°C) in August and lowest in January (-1.8°C); monthly average solar irradiance was highest in May (180.5 W/m^2) and lowest in December (68.8 W/m^2). Average relative humidity in Seoul was highest in July (76.2%) and lowest in February (51.3%), and monthly average wind speeds were between 1.12 m/sec (in October) and 1.65 m/sec (in March).

Hourly average ozone concentrations had a diurnal variation, with the highest concentration at 4 p.m. and a small bump at 4 a.m. (Fig. 2). Because high temperatures favored the formation of ozone, concentrations often peaked in the afternoon. The small bump in ozone concentration around 4 a.m. occurred because the nocturnal planetary boundary layer limited the amount of air pollution,

Table 1. Average ozone and NO_2 concentrations, temperature, solar radiation, relative humidity, and wind speed in 25 monitoring stations in Seoul from 2005 to 2012

Station Code	O_3 (ppb)		NO_2 (ppb)		Temperature (°C)		Solar Irradiance (W/m^2)		Relative Humidity (%)		Wind Speed (m/sec)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Stations in Central Seoul												
C1	19.2	17.2	33.5	16.7	13.5	10.8	125.1	197.0	58.4	18.7	1.3	0.7
C2	18.2	17.3	36.1	18.0	13.8	10.8	NA	NA	59.4	18.7	1.2	0.7
C3	17.5	16.4	36.6	18.2	12.7	10.8	136.3	213.9	60.1	20.1	1.9	1.1
Stations in NW Seoul												
NW1	21.7	17.4	27.8	14.2	13.1	10.8	NA	NA	63.7	19.9	1.1	0.8
NW2	18.8	16.1	34.0	19.0	13.0	11.0	NA	NA	58.3	20.7	1.7	1.1
NW3	17.4	15.4	35.9	19.0	13.6	10.7	NA	NA	61.4	19.5	1.8	1.1
Stations in NE Seoul												
NE1	17.9	17.1	32.0	17.6	12.7	10.7	NA	NA	62.7	22.4	1.0	0.8
NE2	20.2	18.4	31.7	16.7	12.8	10.9	NA	NA	65.1	21.3	2.0	1.1
NE3	18.9	18.1	39.7	19.9	13.8	10.8	NA	NA	60.4	19.4	1.5	1.0
NE4	18.0	15.8	35.1	18.3	13.9	10.8	119.7	191.9	60.9	20.0	1.1	0.7
NE5	18.4	17.4	34.5	17.7	13.6	11.0	NA	NA	60.1	18.4	1.6	0.8
NE6	21.2	18.6	34.6	16.9	13.4	10.7	124.8	202.5	58.8	19.6	1.1	0.6
NE7	21.2	18.3	30.2	17.9	12.7	10.9	101.2	170.4	58.8	21.1	1.1	0.7
NE8	21.1	18.9	31.3	17.6	13.1	10.8	129.4	203.8	57.4	19.5	1.4	0.8
Stations in SW Seoul												
SW1	19.3	17.7	35.9	18.5	13.4	10.5	127.4	201.1	59.7	18.7	1.4	0.7
SW2	18.4	16.7	35.9	19.3	13.1	10.9	NA	NA	56.8	18.5	1.1	0.6
SW3	19.8	17.5	35.5	18.6	13.5	10.6	NA	NA	56.6	18.4	1.7	0.9
SW4	20.0	18.1	37.1	18.8	13.3	10.6	128.2	202.5	59.8	18.7	1.3	0.9
SW5	17.5	16.8	37.2	18.1	13.8	10.7	NA	NA	60.9	19.0	0.9	0.5
SW6	20.3	18.1	35.4	18.7	13.2	10.7	111.0	171.2	58.3	18.8	1.3	0.8
SW7	19.8	17.9	37.5	19.1	13.8	10.6	132.5	202.6	60.9	19.0	1.4	0.9
Stations in SE Seoul												
SE1	18.4	17.0	35.6	17.6	13.1	10.7	122.1	190.9	59.9	19.7	0.9	0.5
SE2	15.5	16.5	39.1	19.1	13.9	10.6	NA	NA	62.9	20.9	1.0	0.7
SE3	19.6	19.3	32.2	17.9	12.6	11.0	133.0	211.8	63.7	21.4	1.3	0.8
SE4	18.4	17.3	32.8	17.4	13.1	10.7	127.7	199.4	57.4	20.0	1.4	0.8
Average	19.1		34.7		13.3		124.5		60.1		1.3	
SD	1.5		2.8		0.4		9.5		2.2		0.3	
Min	15.5		27.8		12.6		101.2		56.6		0.9	
Max	21.7		39.7		13.9		136.3		65.1		2.0	

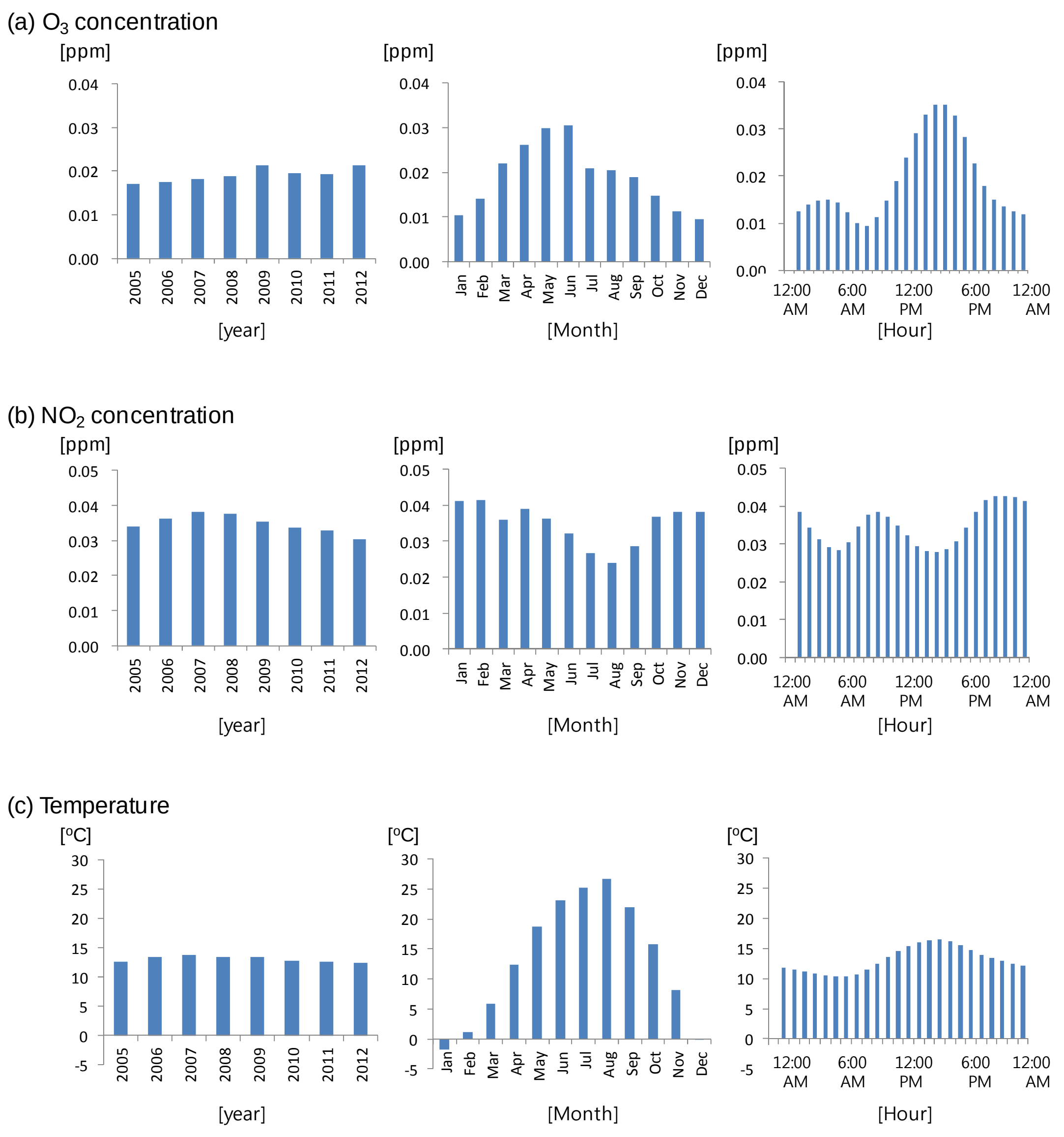

Fig. 2. Annual, monthly, and hourly average of (a) ozone concentrations, (b) NO_2 concentrations, (c) temperature, (d) solar radiation, (e) relative humidity, and (f) wind speed for 25 monitoring stations in Seoul from 2005 to 2012.

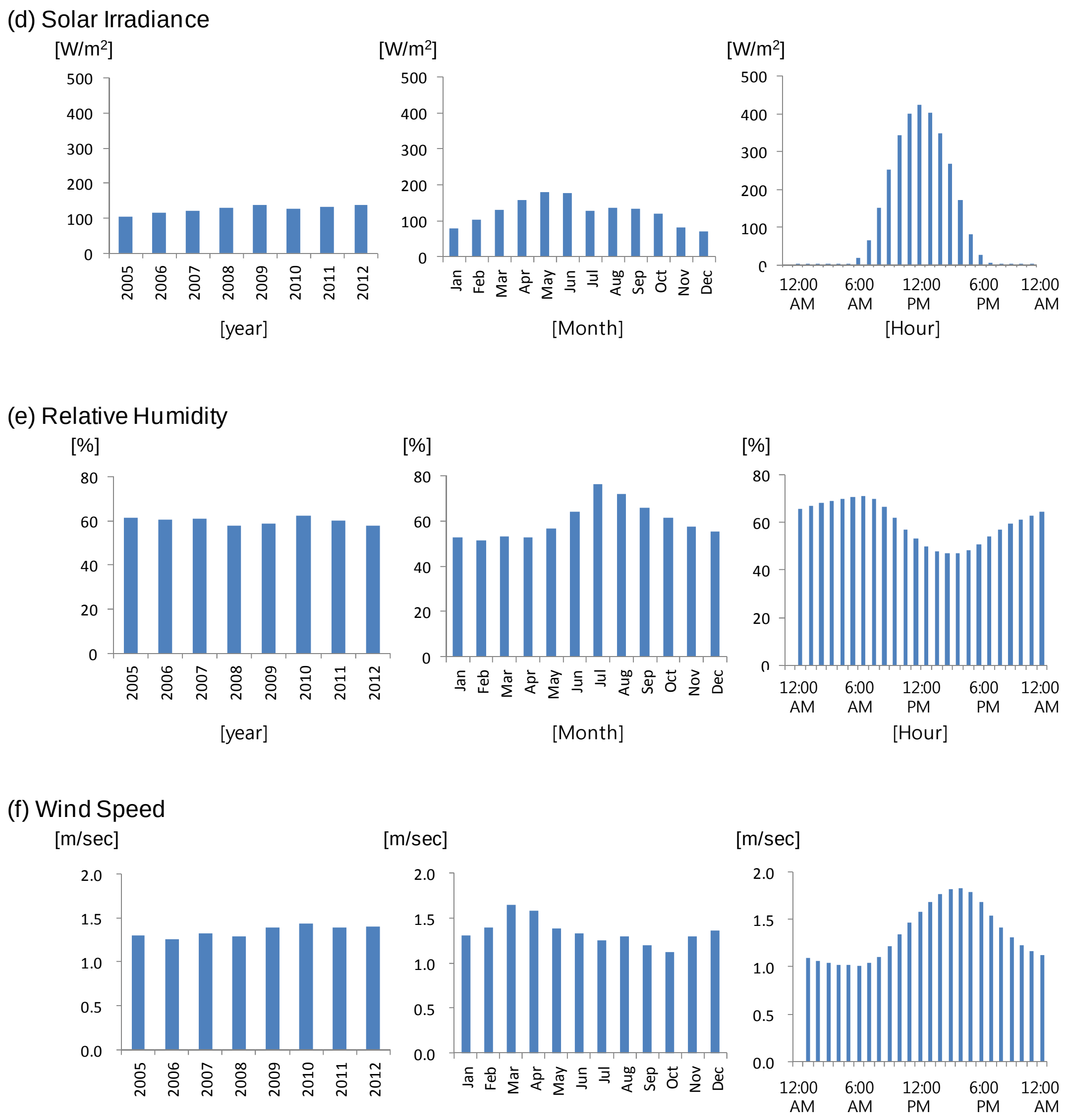

Fig. 2. Continued.

and this boundary layer was often lowest right before dawn when the air was relatively stable(Kim et al., 2007). Hourly NO_2 concentration had two peaks, one in the morning and the other in the evening. This was because morning and evening traffic surges emitted a large amount of NO_x. Studies have shown that because a large percentage of NO_2 is formed secondarily, morning and evening peaks of NO_2 often occur one or two hours after rush hours(Han et al., 2011). Hourly average temperatures and wind speeds were highest around 4 p.m., coinciding with occurrence of the lowest hourly relative humidity readings. Hourly average solar irradiance peaked at 1 p.m., a few hours before the maximum ozone concentrations.

3.2. Correlation between ozone and its potentially formative variables

A correlation coefficient is a useful statistics for finding the similarities and differences of a temporal trend between two variables. Among various correlation coefficients, Spearman's rank correlation coefficient (hereafter, correlation coefficient) measures how well two variables match with each other. Correlation coefficients between ozone and its formative variables between 2005 and 2012 were calculated with lag time from zero to 48 hours. The autocorrelation coefficients for hourly ozone concentrations increased and decreased with the period of 24 hours because of diurnal variations of ozone concentrations (Fig. 3). The correlation coefficient between ozone and NO_2 was -0.44 when lag time was zero. This negative correlation coefficient occurred not because low NO_2 concentrations favor ozone formation but because hourly ozone and NO_2 have opposite diurnal trends (Fig. 2). The correlation coefficient increased until the lag time reached -17 hours. However, this lag time of -17 hours had no direct link with the reaction time for ozone formation; instead it rather implied the time at which the temporal variations of hourly ozone and NO_2 were most similar with each other.

The correlation between ozone and temperature peaked when the lag time was zero(Fig. 3). This indicated that ozone and temperature have similar diurnal trends(Fig. 2). The correlation coefficient between ozone and solar irradiance was 0.53 when the lag time was zero. This correlation coefficient increased to 0.69 when lag time was -2 hours, demonstrating that maximum ozone concentrations were usually reached a few hours after peak solar irradiance. ozone and relative humidity were negatively correlated with the correlation coefficient of –0.37 when lag time was zero. Relative humidity was often low when solar irradiance was high and vice versa(Chu et al., 2012). The correlation coefficient between ozone and wind speed was 0.52 when lag time was zero. This high correlation between ozone and wind speed, however, occurred because wind speed was usually highest in the afternoon when ozone concentrations also were high, not because of any formative link between wind speed and ozone.

3.3. Lag time between ozone and its formative variables

The correlation coefficients calculated and discussed earlier between ozone concentrations and several variables showed the similarities of the diurnal variations of the two time series. This ruled out use of the correlation coefficient to select variables affecting high ozone formation. To find the factors affecting high ozone concentrations and the lag time between ozone and other variables, comparisons were made between the NO_2 concentrations and meteorology when the ozone concentrations exceeded the standard value and those when ozone concentrations were within the standard value. The number of 1-hour ozone

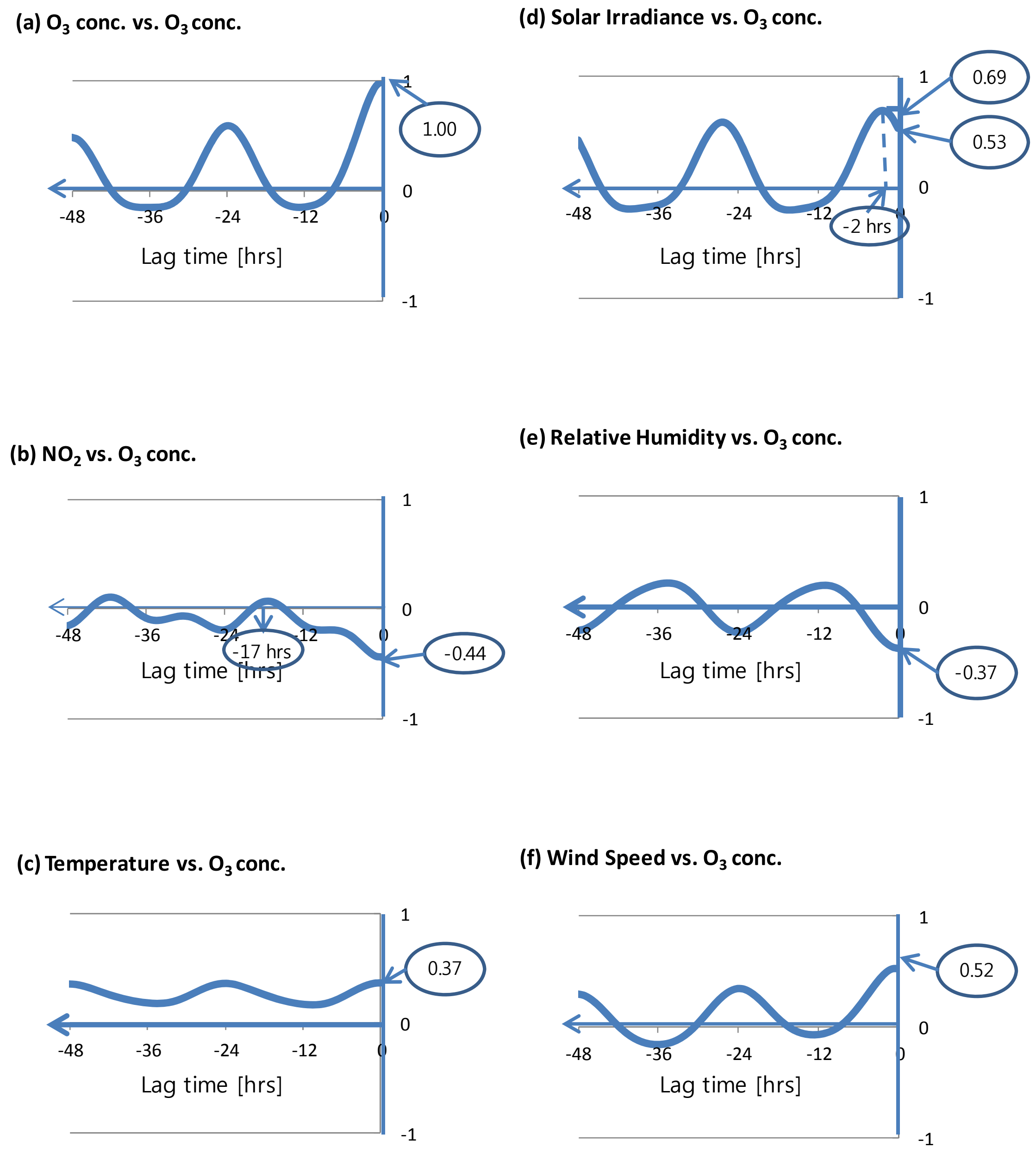

Fig. 3. Correlation coefficients with the lag time from zero to -48 hours between hourly ozone and nitrogen dioxide concentrations, temperature, solar radiation, relative humidity, and wind speed from 2005 to 2012.

exceedances occurred, on average, 382 times per year in Seoul based on the hourly concentrations monitored in 25 stations. The number of ozone exceedances was highest in 2009 (567 times) and lowest in 2006 (149 times) (Fig. 4(a)). The exceedances were observed from April to September, with the highest number occurring in June; more than 98% of these exceedances occurred between 1 p.m. and 7 p.m. (Figs. 4(b), 4(c)) More exceedances occurred in eastern Seoul, with the highest number of exceedances in SE3 (199 times); the fewest exceedances occurred in NW3 (51 times) (Fig. 4(d)).

To find the factors affecting ozone exceedances, average NO_2 concentrations and meteorological conditions when the ozone exceedances occurred were compared with those when ozone was within specified limits. Because ozone and NO_2 concent -rations and meteorology had obvious seasonal and diurnal variations, comparing the average value for a full year or a full day could not clearly represent the factors affecting the ozone exceedances. For example, NO_2 levels were relatively high in winter and low in summer because of the NO_x emissions generated from winter heating requirements (Fig. 2). Because ozone exceedances mostly occurred in summer when NO_2 concentrations were low, average concentrations of NO_2 when ozone exceedances occurred could be considerably lower than those when ozone exceedances did not occur. Thus, if seasonal variations were not taken into consideration, an analysis could conclude that low NO_2 concentrations favor high ozone formation, although $NO_2$was well known as a precursor material in ozone formation. Therefore, to minimize the effects of seasonal and diurnal variations, comparisons were made using data for each hour between May 1 and September 15 of each year, a period that corresponds to the ozone season in Korea.

The comparison was performed for each hour between 1 p.m. and 7 p.m., in which more than 95% of the ozone exceedances were observed (Fig. 4(c)). The NO_2 concentrations and wind speeds during ozone exceedances were similar to those periods when ozone exceedances did not occur (Figs. 5(b), 5(f)). Temperature was higher, but relative humidity was lower when ozone exceedances were observed (Figs. 5(c), 5(e)). Solar irradiance two hours before ozone exceedances was higher than when ozone exceedances did not occur. A lag time of -2 hours was given for solar irradiance because high ozone concentrations were observed a few hours after a period of strong sunlight (Fig. 2). Thus, solar irradiance was plotted between 11 a.m. and 5 p.m (Fig. 5(d)).

The comparisons showed the differences between NO_2 concentrations and meteorology when ozone concentrations exceeded the standard and those times when ozone concentrations were not excessive. However, comparisons at the time of exceedances alone could not show the degree of effect these variables had on the ozone exceedances. To assess the impact of other variables on high ozone formation requires analysis of the NO_2 concent -rations and meteorology before ozone exceedances occurred. Thus, NO_2 concentrations and meteorology were backtracked for 48 hours from the time at which ozone exceedances were observed (Figs. 6 through 11). To see ozone concentrations before ozone exceedances occurred, ozone concentrations were also backtracked for 48 hours.

The dotted line in Fig. 6 shows ozone concent -rations backtracked from the time at which 1-hour ozone exceedances were observed, whereas the solid line represents those backtracked from the time at which 1-hour ozone exceedances were not observed. The difference between the two lines varied as the lag time changed. The difference was

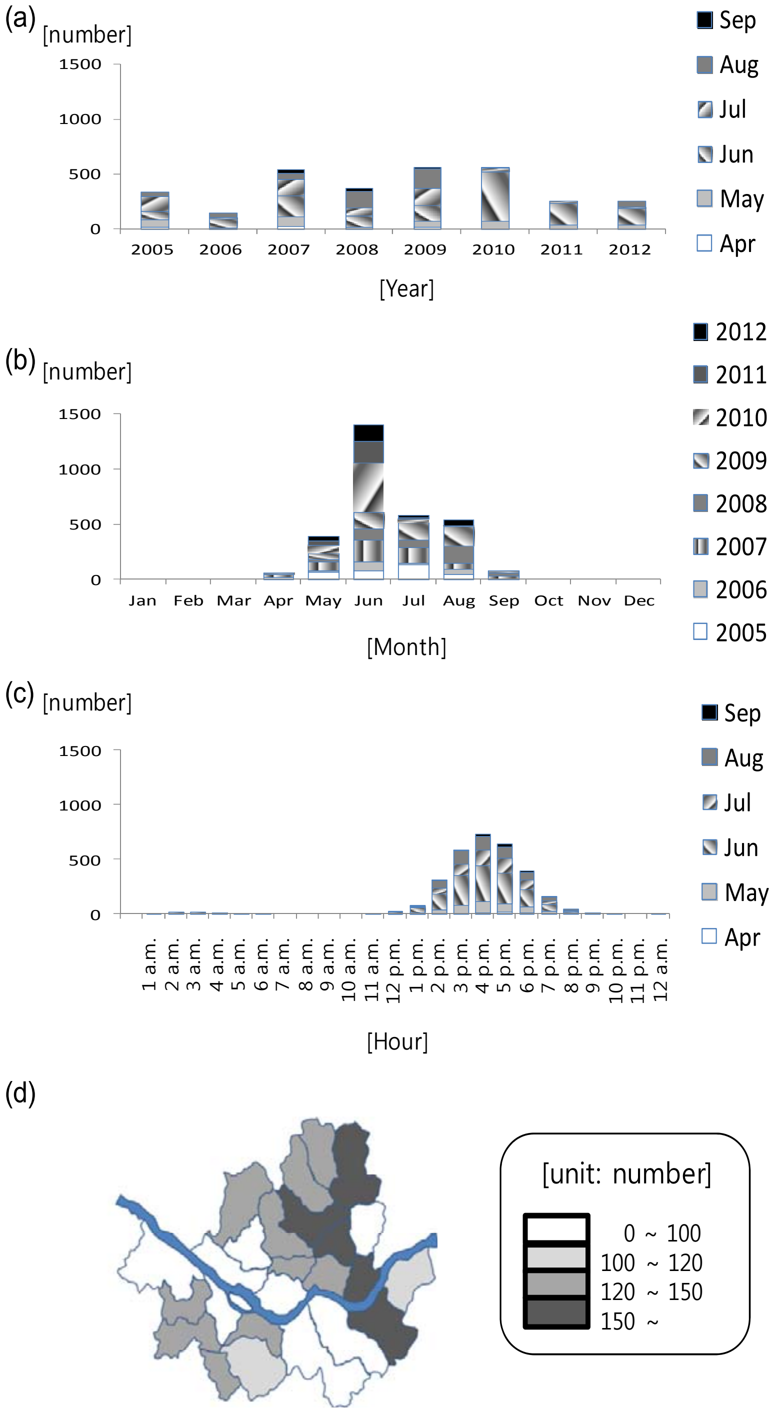

Fig. 4. The total number of 1-hour ozone exceedance (a) in each year, (b) in each month, (c) in each hour, and (d) in each region from 2005 to 2012.

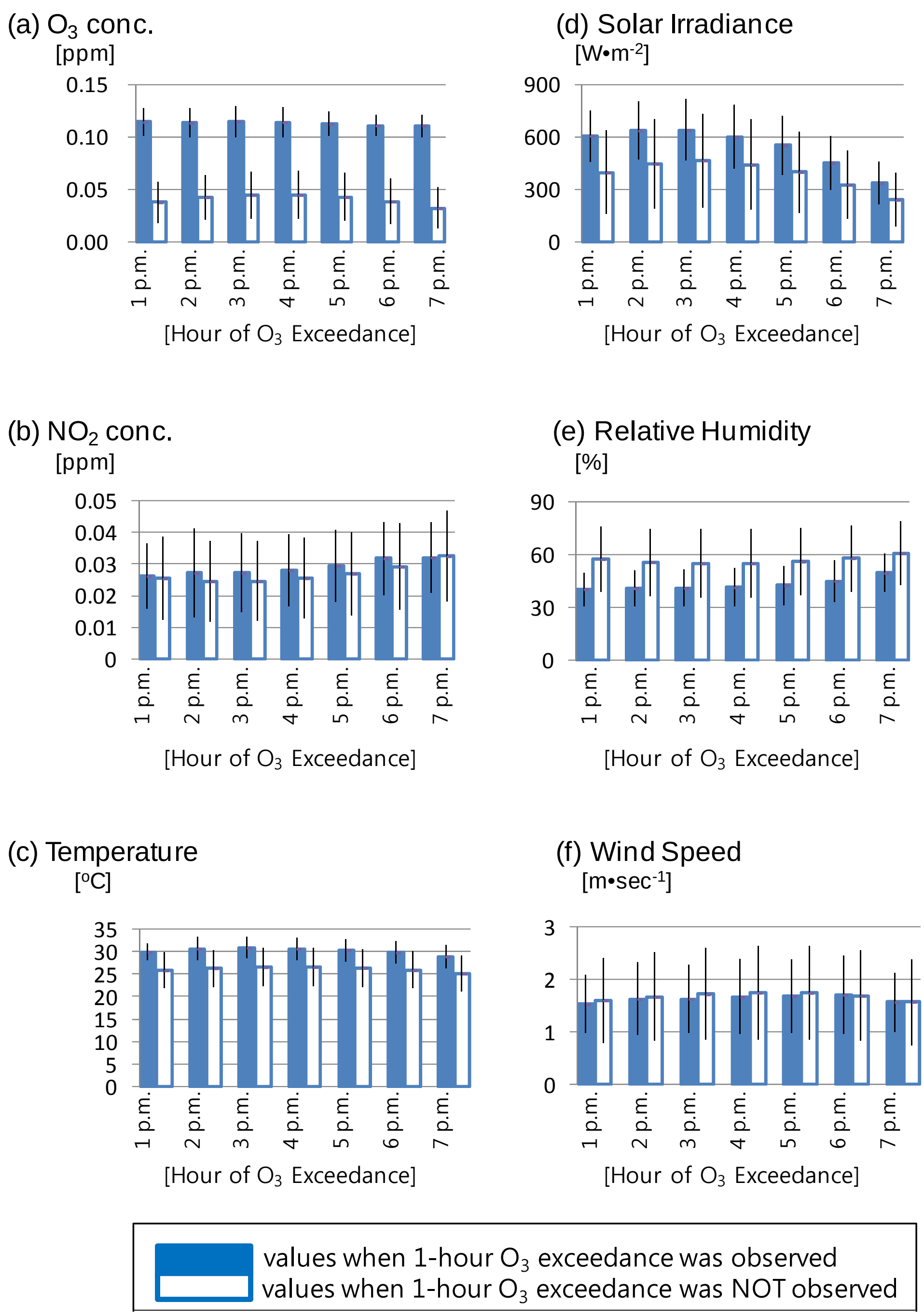

Fig. 5. Average and one standard deviation (vertical line on top of the bar) of ozone and NO_2 concentrations and meteorology when 1-hour ozone exceedance occurred and those when 1-hour ozone exceedance did not occur.

largest at the time of an exceedance, and the second peak of the difference was observed around 4 p.m. of the previous day, irrespective of the time of the exceedance. Thus, a high ozone concentration at 4 p.m. indicated a high possibility of an ozone exceedance the next day. Average ozone concent -rations at 4 p.m. of the previous day were 0.09 ppm when ozone exceedances were observed at 1 p.m. and 0.04 ppm when exceedances did not occur at 1 p.m. (Fig. 6(a)). Thus, the difference was as much as 0.05 ppm. Although average ozone concentrations at 4 p.m. of the previous day were 0.07 ppm when ozone exceedances occurred at 7 p.m., they were 0.04 ppm when ozone exceedances did not occur at 7 p.m. (Fig. 6(c)). In this case, the difference was only 0.03 ppm. Because the difference decreased with exceedances later in the afternoon, this indicates that high previous-day ozone concentrations affect ozone exceedances in the early afternoon more than in the late afternoon.

NO_2 concentrations apparently made no difference with or without ozone exceedances at the time of an ozone exceedance (Fig. 5(b)). However, when these NO_2 concentrations were backtracked for 48 hours, NO_2 concentrations before ozone exceedances were apparently higher than when ozone exceedances did not occur (Fig. 7). The difference was significant from the late evening of the previous day till the morning of the day in which the exceedance occurred. NO_2 levels more than 24 hours before exceedances made little difference. The results suggested that NO_2 concentrations on the late afternoon of the previous day and those in the morning of a day directly affect ozone exceedances.

The spatial distribution of NO_2 concentrations also showed higher NO_2 levels in most regions before ozone exceedances occurred (Fig. 12). As an example, spatial distributions of NO_2 levels backtracked for six hours from 4 p.m. when ozone exceedances occurred were compared with those backtracked from 4 p.m. without exceedances. The difference in NO_2 levels increased as the lag time shifted from zero to -6 hours. NO_2 levels were higher in the southwestern section of Seoul in the morning because of high on-road mobile NO_x emissions in the city's southwestern region. This shift of high NO_2 levels to the east occurred because of the prevailing westerly wind in Seoul, and ozone was formed a few hours later when sunlight became strong. Thus, the exceedances of ozone occurred more often in eastern Seoul (Fig. 4(d)). The largest part of NO_x emissions in Seoul came from vehicular sources. The major sources of traffic emissions in the southern and western parts of Seoul included commuters from these regions where numerous small cities are located. Other major sources of NO_x emissions included combustion and nonroad mobile sources. The major sources of nonroadNO_x emissions included Incheon International Airport and a waste treatment plant located in the west of Seoul (Fig. 13).

Average temperatures backtracked 48 hours from the time of ozone exceedances were higher than similar periods without exceedances. The difference was largest at the time of exceedance (Fig. 8). Average solar irradiance was also higher when ozone exceedances occurred because strong sunlight favors ozone formation (Fig. 9). The correlation coefficient between solar irradiance and ozone concentrations was highest when the lag time was -2 hours (Fig. 3). However, solar irradiance backtracked from ozone exceedances showed that the difference in solar irradiance was largest between 12 p.m. and 2 p.m., irrespective of the time of ozone exceedance. Average relative humidity was apparently lower when ozone exceedances occurred than when ozone exceedances did not occur (Fig. 10). The difference in relative humidity was largest at the time of ozone exceedance.

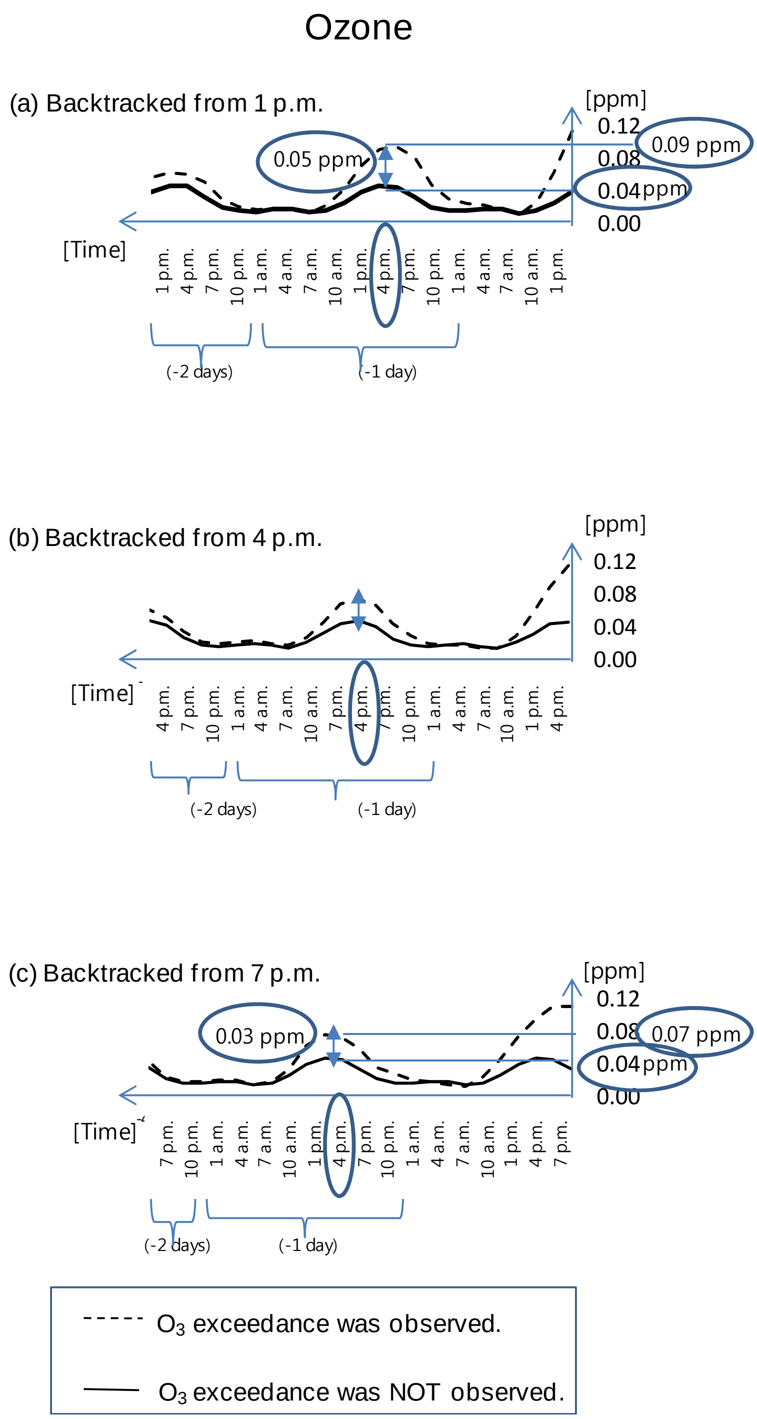

Fig. 6. Average ozone concentrations backtracked for 48 hour from the time 1-hour ozone exceedance occurred and those from the time 1-hour ozone exceedance did not occur from May 1 to September 15, between 2005 and 2012.

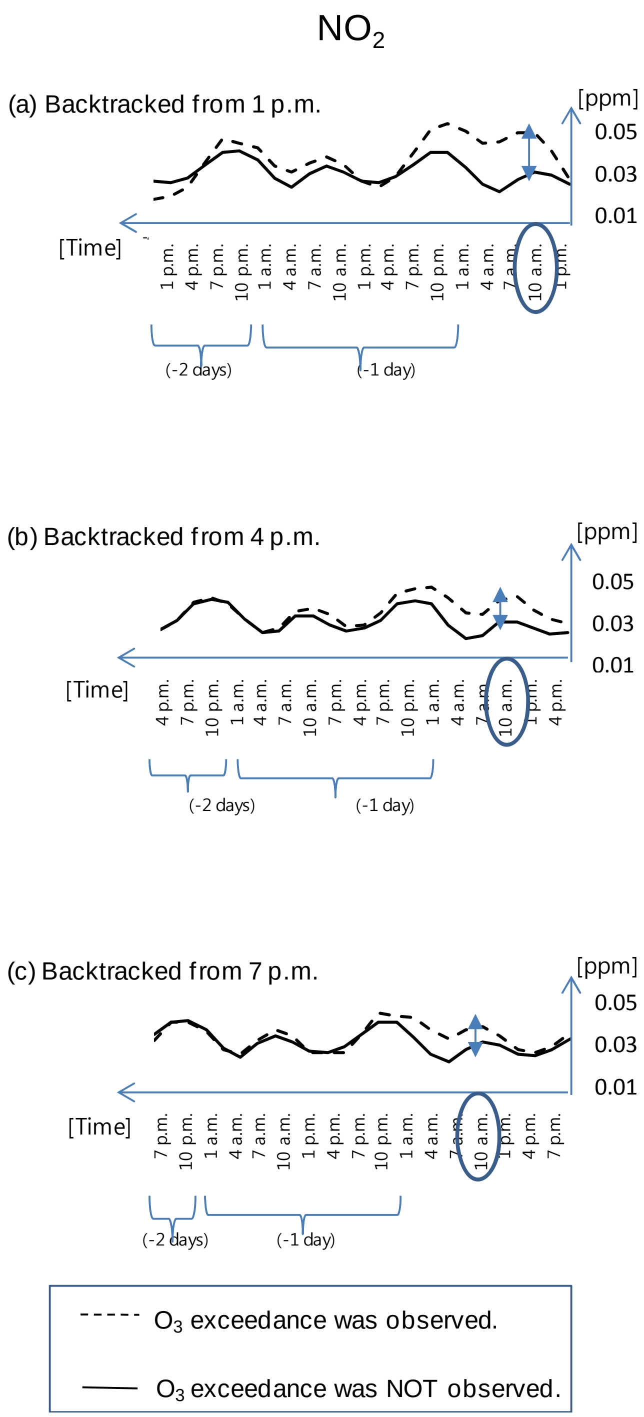

Fig. 7. Average NO_2 concentrations backtracked for 48 hour from the time 1-hour ozone exceedance occurred and those from the time 1-hour ozone exceedance did not occur from May 1 to September 15, between 2005 and 2012.

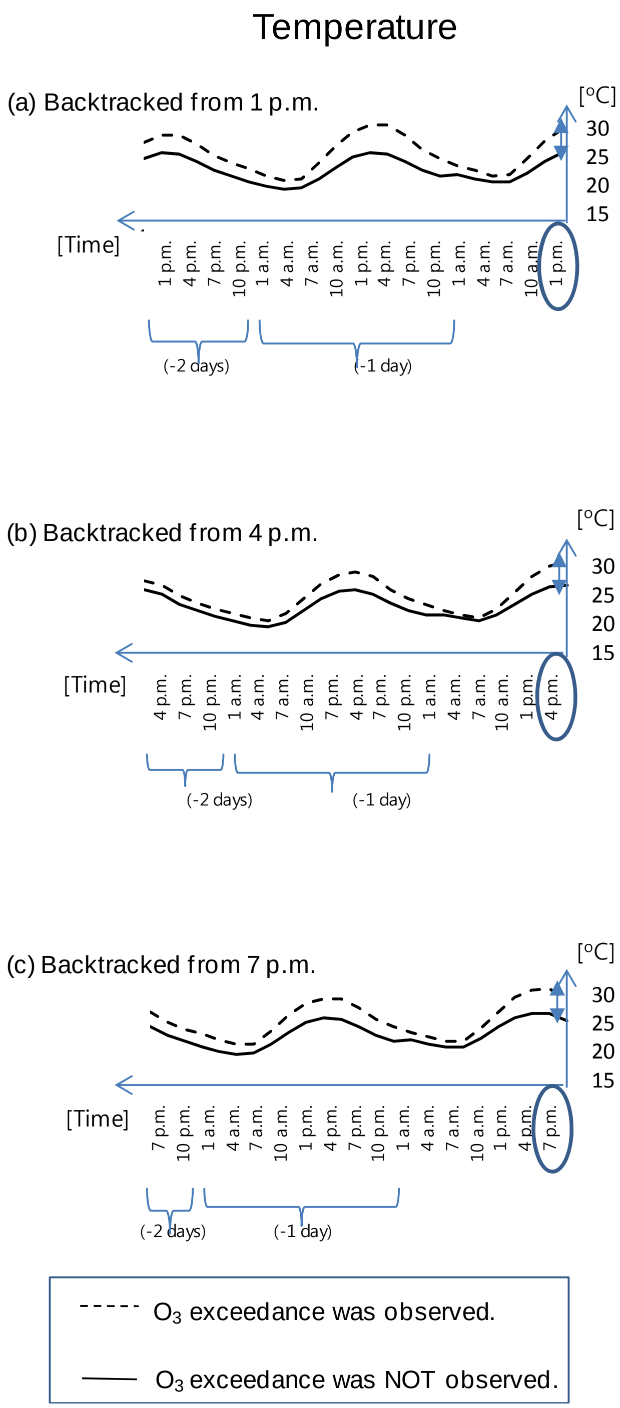

Fig. 8. Average temperature backtracked for 48 hour from the time 1-hour ozone exceedance occurred and those from the time 1-hour ozone exceedance did not occur from May 1 to September 15, between 2005 and 2012.

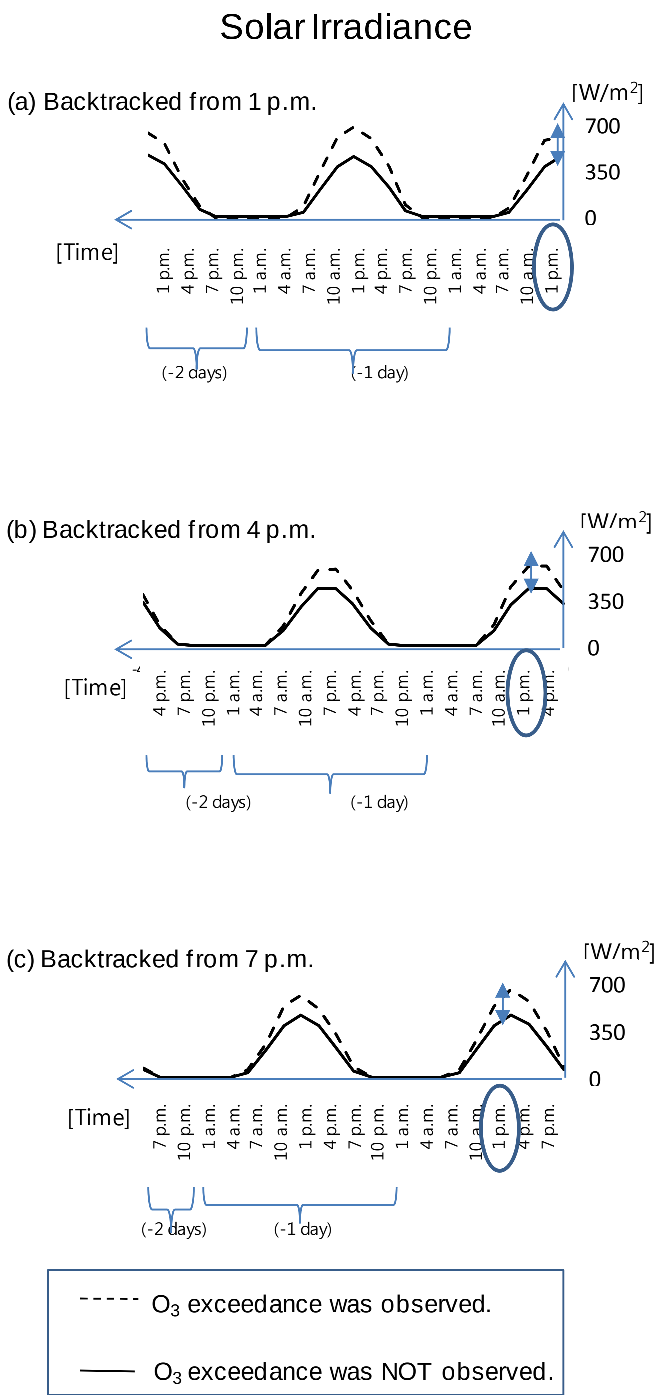

Fig. 9. Average solar irradiance backtracked for 48 hour from the time 1-hour ozone exceedance occurred and those from the time 1-hour ozone exceedance did not occur from May 1 to September 15, between 2005 and 2012.

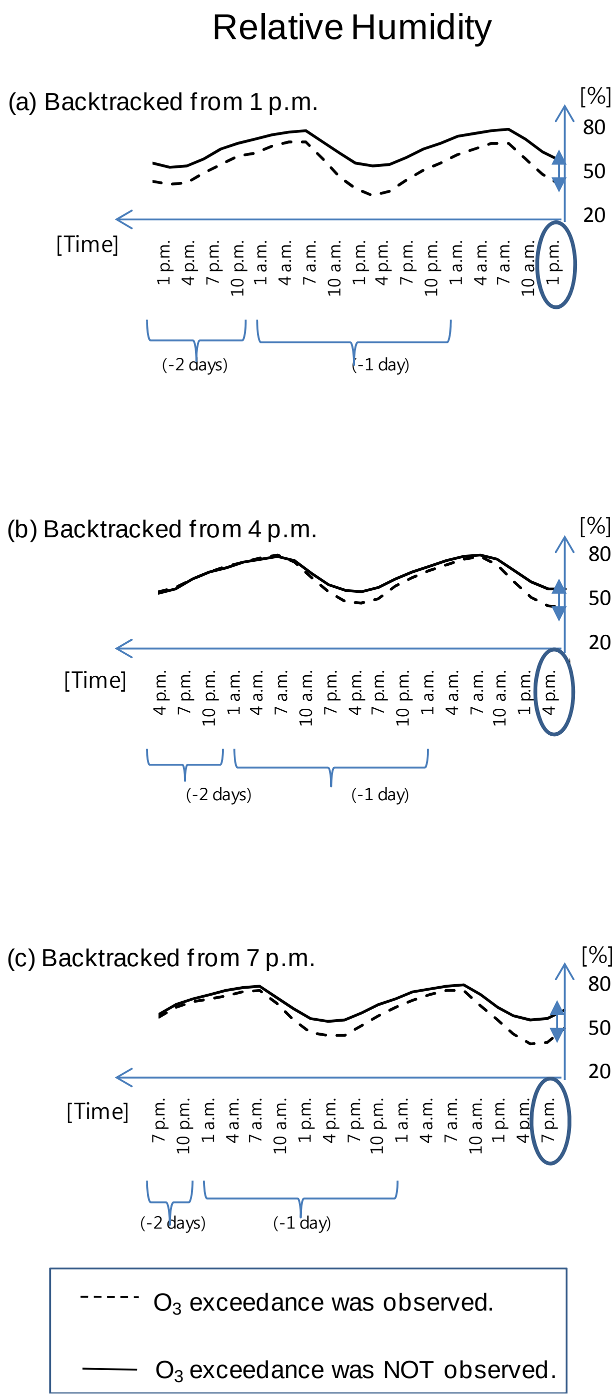

Fig. 10. Average relative humidity backtracked for 48 hour from the time 1-hour ozone exceedance occurred and those from the time 1-hour ozone exceedance did not occur from May 1 to September 15, between 2005 and 2012.

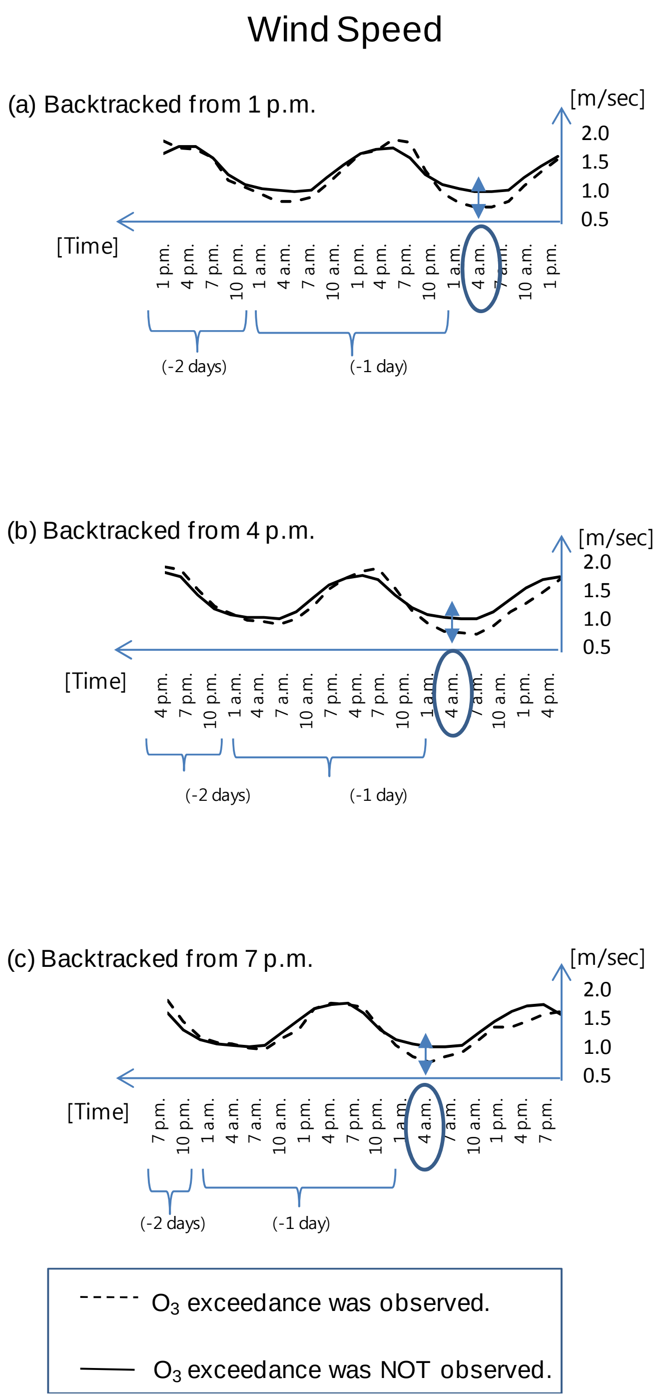

Fig. 11. Average wind speed backtracked for 48 hour from the time 1-hour ozone exceedance occurred and those from the time 1-hour ozone exceedance did not occur from May 1 to September 15, between 2005 and 2012.

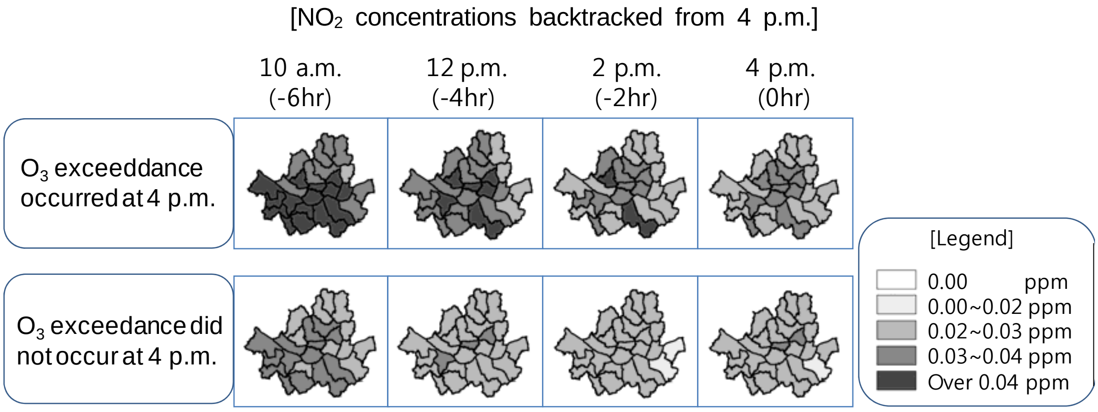

Fig. 12. Spatial distribution of average NO_2 concentrations backtracked from 4:00 p.m. between May 1 and September 15 from 2005 to 2012 when 1-hour ozone exceedance occurred at 4:00 p.m. or not.

Wind speed at the time of ozone exceedances made little difference with or without ozone exceedances (Fig. 5). However, when wind speed was backtracked for 48 hours, an apparently low wind speed was observed in the early morning on the day of exceedance (Fig. 11). Low wind speed slowed the horizontal mixing of pollutants, and the boundary layer in the early morning was relatively low. Thus, low wind speed in the early morning created good conditions for undiluted NO_2. Therefore, when wind speed was low in the early morning, relatively high NO_2 concentrations were observed and eventually, high ozone concentrations.

3.4. Prediction of ozone exceedances using a decision tree algorithm

Factors characterizing ozone exceedances were predicted using a decision tree algorithm. The independent variables were the ozone concentrations of the previous day, NO_2 concentrations and meteorological variables (temperature, relative humidity, solar irradiance, and wind speed). The dependent variable was hourly ozone concentrations classified as “high level” (> 0.1 ppm) and “low level” (≤ 0.1 ppm). A threshold of 0.1 ppm was selected as the hourly ozone standard in Korea. Because there were more instances when ozonedid not exceed the standard than when it did, the data classified as “high level” were duplicated to balance the numbers in the two categories. The terminal node minimum cases were given as the number of the duplication, and the parent node minimum cases were put as 10 times of the terminal node minimum cases.

Two models, named as Model [A] and Model [B] were constructed depending on times chosen as the independent variables selected. The times of the NO_2 concentration, temperature, relative humidity, and wind speed in Model [A] were the same as that of the dependent variable. The time of solar irradiance in Model [A] was selected as two hours earlier than that of the dependent variable. The lag time of two hours for solar irradiance was selected as the correlation coefficient between hourly solar irradiance and ozone concentration was highest when the lag time was -2 hours.The ozone concentration at 4 p.m. of the previous day was also included as the independent variable in Model [A].

The times of temperature and relative humidity in Model [B] were the same as that of the

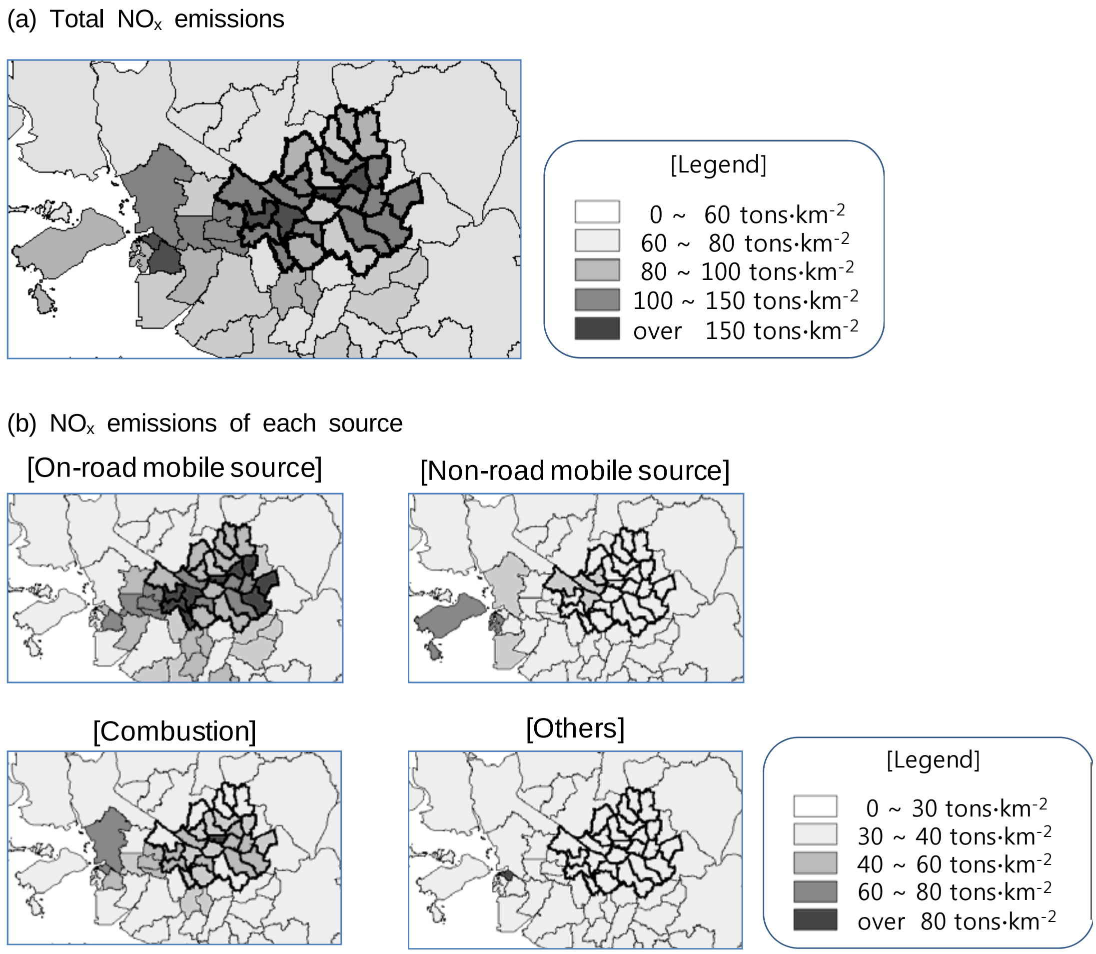

Fig. 13. Annual NOx emissions per unit area in Seoul and surrounding regions in 2010(Lee et al., 2011).

dependent variable. However, the times of solar irradiance, the NO_2 concentration, and wind speed were 1 p.m., 10 a.m., and 4 p.m., respectively. The choice of the specific times for the independent variables was based on the comparison of the two time series; one backtracked from the ozone exceedances and the other backtracked from the lack of occurrences of ozone exceedances. When the two time series were compared, solar irradiance around 1 p.m. and the NO_2 concentrations around 10 a.m. were much higher when the ozone concentrations exceeded the standard than when they did not (Figs. 7 and 9). Wind speed around 4 a.m. when the ozone concentrations exceeded the standard was much lower than in their absence (Fig. 11). The ozone concentration at 4 p.m. of the previous day was also included as the independent variable in Model [B].

The performance of the model was evaluated through misclassification errors in 10-fold cross validation. Although the misclassification errors of both Model [A] and Model [B] were less than 10%, the performance of Model [B]was better than that of Model [A] from 0% to 4% depending on the

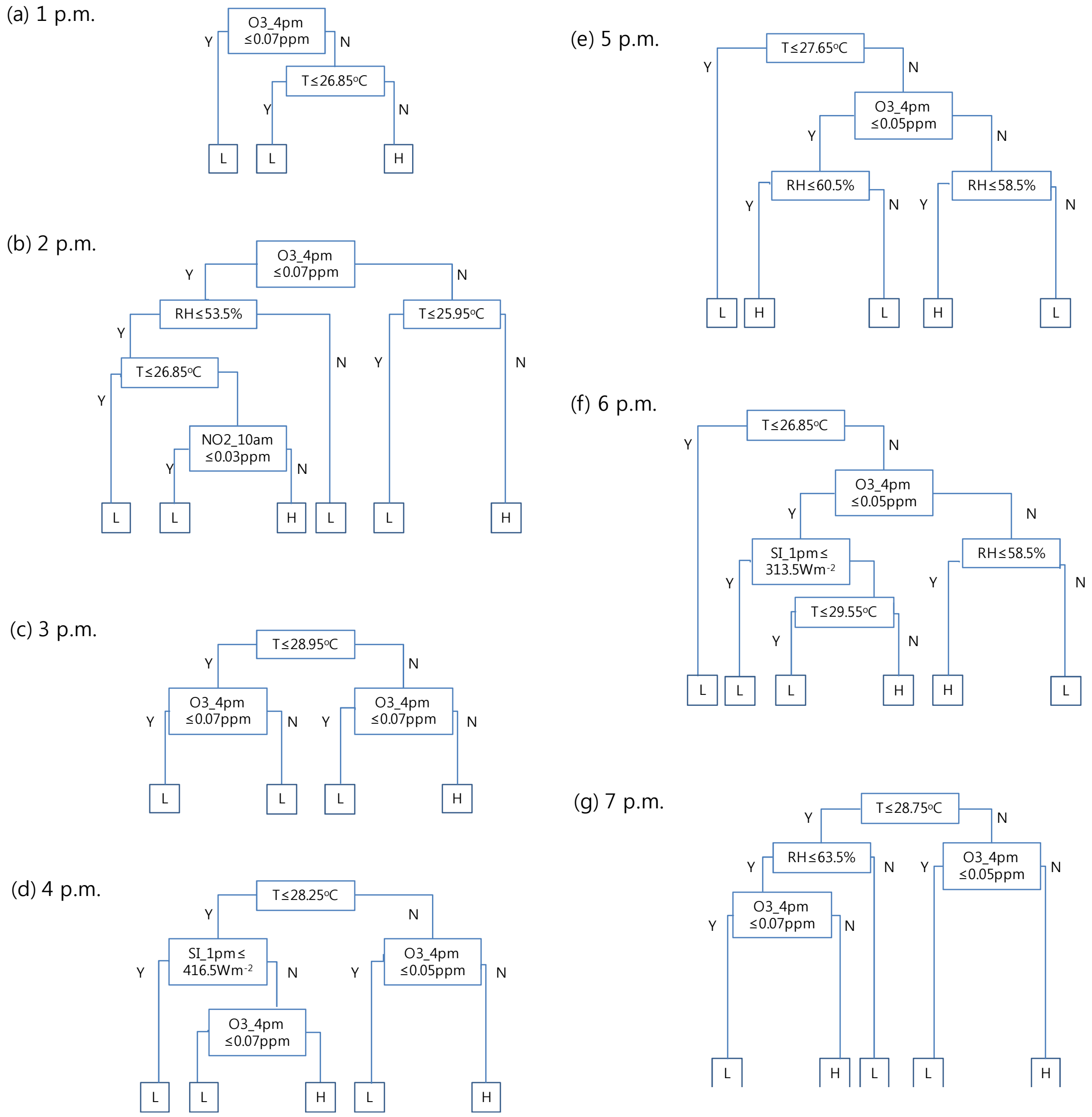

Fig. 14. Schematic diagram of primary splitters of classification trees for high level (H) and low level (L) ozone concentrations.

time of the ozone exceedance. However, the actual improvement should be bigger than that since measured NO_2 concentrations and wind speed could be used in Model [B], in which the times of NO_2 and wind speed were earlier than that of the dependent variable. Whereas, the times of NO_2 and wind speed in Model [A] were the same with that of dependent variable. The classification trees for high level (H) and low level (L) ozone concentration in Model [B] were represented in Fig. 14. The primary splitters between high and low level ozone concent -rations in Model [B] were illustrated in Table 2. The relative importance of variables for ozone

Table 2. Sets of primary splitters for classifying high and low ozone concentrations

Time of the ozone exceedance	Primary Splitters*
1:00 PM	O_3_4pm, T
2:00 PM	O_3_4pm, RH, T, NO_2_10am
3:00 PM	T, O_3_4pm
4:00 PM	T, SI_1pm, O_3_4pm
5:00 PM	T, O_3_4pm, RH
6:00 PM	T, O_3_4pm, SI_1pm, RH
7:00 PM	T, RH, O_3_4pm

*T: temperature
RH: relative humidity
O_3_4pm: ozone concentration at 4 p.m. of the previous day
SI_1pm: solar irradiance at 1 p.m.
NO_2_10am: NO2 concentration at 10 a.m.

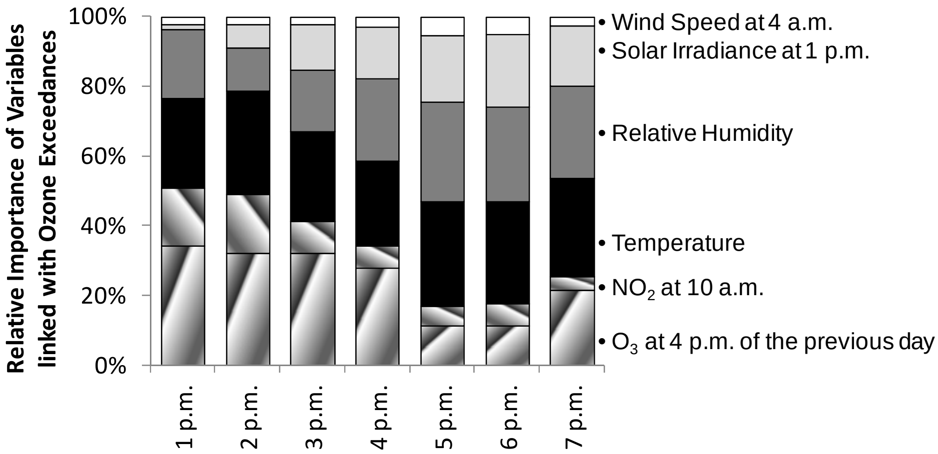

Fig. 15. Relative importance of variables linked with ozone exceedances.

exceedance of Model [B] showed that ozone concentration of the previous day and NO_2 concentration in the morning were the most important drivers of ozone exceedance in the early afternoon (1 p.m. ~ 4 p.m.). Whereas, meteorological variables such as temperature, relative humidity, and solar irradiance were major drivers of ozone exceedance in the late afternoon (5 p.m. ~ 7 p.m.) (Fig. 15).

4. Conclusions

As the evidence linking air pollution and human health increases, air pollutant concentrations have been monitored in 25 stations in Seoul, Korea. Among pollutants, ozone (O_3) concentrations in Seoul often exceeded the standard. In order to minimize the adverse health effect of ambient ozone, Korean government has been operating an ozone warning system, which informed the public of appropriate behavior when high ozone levels were expected. The goal of this research was to assess factors linked with ozone exceedances to be used for developing control strategies of ozone in Seoul, Korea.

The data used included hourly ozone, NO_2, temperature, solar irradiance, relative humidity, and wind speed monitored in Seoul, Korea, between 2005 and 2012. The factors were assessed using a decision tree model, which is easily interpretable to quantify the importance of factors affecting ozone exceedances. Appropriate lag times between ozone and explanatory variables, used for the decision tree model, were also taken into account. The lag time was investigated by constructing two sets of 48-hour time series: One with meteorology and NO_2 backtracked 48 hours from the time at which 1-hour ozone exceeds the standard, and the second with meteorology and NO2 backtracked from when ozone did not exceed the standard. Although ozone exceedances mostly occurred in the afternoon, the comparison of the two time series showed that solar irradiance differed most around 1 p.m., and NO_2 concentrations and wind speeds differ most in the early morning. The differences in temperature and relative humidity were greatest at the time of exceedance indicating no significant lag time exists for temperature and relative humidity.

The factors characterizing ozone exceedances were analyzed through the CART model. Two models, Model [A] and Model [B], were constructed, depending on the times chosen as independent variables. Model [A] used the independent variables concurrent with the dependent variable, while Model [B] used the time lagged independent variable. Results showed that Model [B] simulated 1-hour ozone exceedances better than Model [A] did. The above results indicated that when a decision tree algorithm was applied to find the ozone exceedances, using independent variables of certain lag times was superior to using those occurring at the same time as the dependent variables. In addition, ozone exceedances in the early afternoon were much affected by ozone concentrations of the previous day and NO_2 concentrations in the morning, while ozone exceedances in the late afternoon were more driven by the meteorology. The results analyzed in this study showed the importance of factors affecting ozone exceedances at each hour, and these results can be used to develop control strategies of ozone in Seoul, Korea.

The limitation of this study includes that VOC data were not included to predict hourly ozone concent -rations. VOC is one of the most important precursor materials for ozone formation. However, this study did not use VOC data to predict hourly ozone concentrations since hourly VOC data were not available. In the future study, the accuracy of predicting ozone concentrations will improve by incorporating more ozone precursor materials including hourly VOC concentrations.

Acknowledgement

This work was supported by the research fund of Hanyang Cyber University.

REFERENCES

Abdul-Wahab, S. A., Bakheit, C. S., Al-Alawi, S. M., 2005, Principal component and multiple regression analysis in modelling of ground-level ozone and factors affecting its concentrations, Env. Model. Softw., 20(10), 1263-1271.

Bauer, G., Deistler, M., Scherrer, W., 2001, Time series models for short term forecasting of ozone in the eastern part of Austria, Environmetrics, 12(2), 117-130.

Breiman, L., Friedman, J. H., Olshen, R. A., Stone, C. J., 1984, Classification and Regression Trees, Wadsworth International Group, Belmont CA, U. S. A.

Burrows, W. R., Benjamin, M., Beauchamp, S., Lord, E. R., McCollor, D., Thomson, B., 1995, CART Decision

-tree statistical-analysis and prediction of summer season maximum surface ozone for the Vancouver, Montreal, and Atlantic regions of Canada, J. App. Met., 34(8), 1848-1862.

Chu, H. J., Lin, C. Y., Liau, C. J., Kuo, Y. M., 2012, Identifying controlling factors of ground-level ozone levels over southwestern Taiwan using a decision tree, Atmos. Env., 60, 142-152.

Gardner, M. W., Dorling, S. R., 2000, Statistical surface ozone models: an improved methodology to account for non-linear behaviour. Atmos. Env., 34(1), 21-34.

Han, S. Q., Bian, H., Feng, Y. C., Liu, A., Li, X., Zeng, F., Zhang, X., 2011, Analysis of the Relationship between O_3, NO and NO_2 in Tianjin, China, Aeros. Air Quality Res., 11(2), 128-139.

He, Y., Uno, I., Wang, X., Ohara, T., Sugirnoto, N., Shimizu, A., Richter, A., Burrows, J., 2007, Variations of the increasing trend of tropospheric NO_2 over central east China during the past decade, Atmos. Env., 41(23), 4865-4876.

Heo, J. S., Kim, D., 2004, A new method of ozone forecasting using fuzzy expert and neural network systems, Sci. Tot. Env., 325(1-3), 221-237.

Jorquera, H., Perez, R., Cipriano, A., Espejo, A., Letelier, M., Acuna, G., 1998, Forecasting ozone daily maximum levels at Santiago, Chile, Atmos. Env., 32(20), 3415-3424.

Kim, K. H., Choi, Y., Kim, M., 2005, The exceedance patterns of air quality criteria: a case study of ozone and nitrogen dioxide in Seoul, Korea between 1990 and 2000, Chemosphere, 60(4), 441-452.

Kim, N. K., Kim, Y., Morino, Y., Kurokawa, J., Ohara, T., 2013, Verification of NOx emission inventory over South Korea using sectoral activity data and satellite observation of NO_2 vertical column densities, Atmos. Env., 77, 496-508.

Kim, S. W., Yoon, S., Won, J., Choi, S., 2007, Ground -based remote sensing measurements of aerosol and ozone in an urban area: A case study of mixing height evolution and its effect on ground-level ozone concentrations, Atmos. Env., 41(33), 7069-7081.

Lee, D., Lee, Y., Jang, K., Yoo, C., Kang, K., Lee, J., Jung, S., Park, J., Lee, S., Han, J., Hong, J., Lee, S., 2011, Korean National Emissions Inventory System and 2007 Air Pollutant Emissions, Asian J. Atmos. Env., 5(4), 278-291.

Marmur, A., Park, S., Mulholland, J., Tolbert, P., Russell, A. G., 2006, Source apportionment of $PM_{2.5}$ in the southeastern United States using receptor and emissions-based models: Conceptual differences and implications for time-series health studies, Atmos. Env., 40(14), 2533-2551.

Moon, S. S., Kang, S., Jitpitaklert, W., Kim, S., 2012, Decision tree models for characterizing smoking patterns of older adults, Exp. Syst. App., 39(1), 445-451.

Pandey, S. K., Kim, K., Chung, S., Cho, S., Kim, M., Shon, Z., 2008, Long-term study of NOx behavior at urban roadside and background locations in Seoul, Korea, Atmos. Env., 42(4), 607-622.

Park, S. K., Cobb, C., Wade, K., Mulholland, J., Hu, Y., Russell, A. G., 2006a, Uncertainty in air quality model evaluation for particulate matter due to spatial variations in pollutant concentrations, Atmos. Env., 40, S563-S573.

Park, S. K., Marmur, A., Kim, S., Tian, D., Hu, Y., McMurry, P., Russell, A. G., 2006b, Evaluation of fine particle number concentrations in CMAQ, Aeros. Sci. Tech., 40(11), 985-996.

Park, S. K., Marmur, A., Russell, A. G., 2013, Environ -mental Risk Assessment: Comparison of Receptor and Air Quality Models for Source Apportionment, Human& Eco. Risk Assess., 19(5), 1385-1403.

Park, S. K., Russell, A. G., 2013, Regional adjustment of emission strengths via four dimensional data assimilation, Asia-Pacific J. Atmos. Sci., 49(3), 361-374.

Qian, W., Kang, H., Lee, D., 2002, Distribution of seasonal rainfall in the East Asian monsoon region, Theor. App. Climat., 73(3-4), 151-168.

Richter, A., Burrows, J., Nuss, H., Granier, C., Niemeier, U., 2005, Increase in tropospheric nitrogen dioxide over China observed from space, Nature, 437(7055), 129-132.

Rohli, R. V., Hsu, S., Blanchard, B., Fontenot, R., 2003, Short-range prediction of tropospheric ozone concentrations and exceedances for Baton Rouge, Louisiana, Weather & Forecast., 18(2), 371-383.

Schlink, U., Dorling, S., Pelikan, E., Nunnari, G., Cawley, G., Junninen, H., Greig, A., Foxall, R., Eben, K., Chatterton, T., Vondracek, J., Richter, M., Dostal, M., Bertucco, L., Kolehmainen, M., Doyle, M., 2003, A rigorous inter-comparison of ground-level ozone predictions, Atmos. Env., 37(23), 3237-3253.

Sousa, S. I. V., Martins, F., Alvim-Ferraz, M., Pereira, M., 2007, Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations, Env. Model. Softw., 22(1), 97-103.

Wang, X. P., Mauzerall, D., 2004, Characterizing distributions of surface ozone and its impact on grain production in China, Japan and South Korea: 1990 and 2020, Atmos. Env., 38(26), 4383-4402.