Assessing the impact of air pollution on mortality rate from cardiovascular disease in Seoul, Korea

Sun Kyoung Park[†]

School of ICT Integrated Studies, Pyeongtaek University, Pyeongtaek 17869, Republic of Korea

ABSTRACT

The adverse health impact of air pollution is becoming more serious. The purpose of this study is twofold: One is to analyze the effect of air pollution and temperatures on human health by analyzing the number of deaths from cardiovascular disease in Seoul, Korea; the other is to determine what impact the location of a monitoring site has on the results of a health study. For this latter purpose, air pollution and temperature monitors are sited at three locations termed green, public, and residential. Then, a decision tree model is used to analyze factors linked with deaths occurring at each monitoring site. The results show that the environmental temperatures before death and the PM_{2.5} concentrations on the day of death are highly linked with the number of deaths regardless of the monitoring location. However, results are most accurate with residential data. The results of this study can be used as base data for a similar analysis and ultimately, as a guide to minimize the health impact of air pollution.

Keywords: Air pollution, Cardiovascular disease, Decision tree model, PM_{2.5}, Temperature

1. Introduction

Air pollution is a critical environmental issue. Various forms of air pollutants are known to have a direct impact on human health. Carbon dioxide has recently been recognized as a main cause of global climate change. An elevated level of PM_{2.5} leads to low visibility and inhibition of plant growth. When inhaled, PM_{2.5}, partly because of its small size, causes serious adverse health effects by penetrating blood vessels [1, 2]. Ozone (O₃), a secondary pollutant, is an important greenhouse gas in causing climate change. In addition, extended exposure to elevated ozone levels is a known cause of respiratory and ocular diseases [3].

The South Korean government has undertaken to reduce the adverse health effects of air pollution by designating eight pollutants – PM_{2.5}, PM₁₀, O₃, sulfur dioxide (SO₂), nitrogen dioxide (NO₂), carbon monoxide (CO), lead (Pb), and benzene (C₆H₆) – as targets of regulation [4]. However, effective reduction of air pollution is currently difficult in Korea because of domestic emissions and the inflow of external pollutants. In addition, the reduction of pollutants in Seoul, the capital of Korea, is complicated by a population nearing 10 million and the air pollutants emitted during commuting.

With the rising awareness of the negative impact of air pollution, epidemiological studies were being widely conducted [5, 6]. To improve the accuracy of health impact assessments of air pollution, the pollutant concentrations used in these analyses must accurately represent their levels in the corresponding region [7, 8]. In addition, air pollutant concentrations must be measured from areas of high population density [9].

The purpose of this study is to analyze the impact of air pollution and meteorological conditions on death caused by cardiovascular disease. Data used were hourly air pollutant concentrations (O₃, NO₂, PM_{2.5}, and PM₁₀) and temperatures in Seoul between 2008 and 2013. In addition, this study analyzed differences in factors linked with the death from cardiovascular disease of persons older than 65 when data from different locations were used. To achieve these goals, the monitoring stations from which air pollutant concentrations and temperatures were collected were divided into three types: Green, Public, and Residential areas. Green area included parks and mountains, and Public area covered the city hall, the public educational institutes, the health centers, the museum, and the district office. Residential area included the community service center, schools, and the public library school. Then, the factors affecting death were analyzed by using

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright © 2018 Korean Society of Environmental Engineers

Received February 6, 2018 Accepted April 20, 2018

[†] Corresponding author

Email: skpark@ptu.ac.kr

Tel: +82-31-659-8310

ORCID: 0000-0002-2896-7024

data from the three types of monitors to ascertain the impact of a monitoring location on an epidemiological study.

Social and economic factors other than air pollution and temperature can affect death from cardiovascular disease. Therefore, if comprehensive diverse factors were included, causes that influence such deaths could be more accurately analyzed. However, because considering all such factors were outside the scope of this study, the influence of air pollution and temperature were specifically considered in assessing the number of deaths caused by cardiovascular disease.

2. Research Methods

2.1. Data

2.1.1. Air pollution and temperature

Hourly air pollutant concentrations measured at 25 observation stations from 2008 to 2013 were used to analyze the health impact of air pollution (Fig. 1) [10]. The typical types of air pollutants in South Korea included PM_{2.5}, PM₁₀, ozone (O₃), sulfur dioxide (SO₂), nitrogen dioxide (NO₂), carbon monoxide (CO), lead (Pb), and benzene (C₆H₆). Among pollutant species, PM₁₀, PM_{2.5}, O₃, and NO₂, which exceeded the standard, were used in the analysis. Because PM₁₀ corresponded to particles smaller than 10 m, and PM_{2.5} referred to particles smaller than 2.5 m, the mass of PM₁₀ included that of PM_{2.5}. To accurately analyze the health impact of particles larger than 2.5 m, PMC (coarse particulate matter) was used instead of PM₁₀. PMC referred to particles with aerodynamic diameter between of 2.5 m and 10 m. Hourly air pollutant concentrations and temperatures were converted into daily values before their use in the analyses. In calculating daily averages with hourly data, only days with more than 20 h of data (> 80% out of 24 h) were used to increase the reliability of the analysis results.

Fig. 1. Air pollution and temperature monitoring locations in Seoul.

2.1.2. Number of deaths caused by cardiovascular disease

Daily deaths attributed to cardiovascular disease were collected from the National Statistical Office [11]. The number of deaths

from cardiovascular disease was taken as the number of deaths in the “I00~I99” category as based on the 10th International Classification of Diseases [12]. The number of deaths of persons older than 65, a population segment known as sensitive to elevated levels of air pollution and to climate change, were used for the analyses [13].

2.2. Classification and Regression Tree (CART) Model

A Classification and Regression Tree (CART) model is a decision tree model widely used for classification and prediction. For example, the CART model has been used extensively in predicting surface O₃ concentrations in the United Kingdom [14]. In addition, the CART model was also used in Taiwan in analyzing factors that influence ozone concentrations [15]. Because the CART model uses a tree structure to show the rules required in classifying or predicting dependent variables, it is advantageous in terms of visual understanding and interpretation [16]. In a CART model, not all independent variables are used to classify dependent variables. Instead, it uses only those independent variables appropriate for creating a decision tree [17]. The selection of independent variables to use to classify dependent variables is determined based on the Gini index [16].

Several impurity indices (Gini index, entropy, misclassification error) existed to measure degree of impurity for classification trees. In this study, Gini index, represented as the following equation, was used [16]:

$$Gini = \sum_j P_j(1 - P_j) \quad (1)$$

where P_j was the probability of each class. Gini index reached the maximum value when all classes had the same probability. For two-class problems, Gini index was maximized when $P_1 = P_2 = 0.5$. The value of Gini index was zero when $P_j = 1$, indicating that all observations belonged to only one class (i.e., perfectly pure state).

Independent variables used in this study were daily air pollutant concentrations (O₃, PM_{2.5}, PMC, NO₂) and temperatures. The dependent variable was expressed as a categorical value: “H(high)” or “L(low).” If the number of deaths on a specific day equals or exceeds the median daily number of deaths, the value of the dependent variable on that day is “H.” Conversely, if the number of deaths is lower than the median number of deaths, the value of the dependent variable is “L.”

In classifying a dependent variable as “H” or “L,” the CART model analyzes which independent variable is important to this determination. For example, when independent variables “A,” “B,” “C,” and “D” are given, the CART model finds an appropriate independent variable – here A – that best classifies as H or L, resulting in two branches (Fig. 2). The model then finds another appropriate independent variable to subdivide each branch into H and L. Because of such a classification method, the result is expressed in a tree-shaped hierarchical structure (Fig. 2). All observational data was given equal weight in this study [15]. Sub branch classification was allowed only when there were 10 or more data in an individual node, and the minimum number of data in an individual node was one. The accuracy of the results was analyzed through a 10-fold cross validation.

Fig. 2. Schematic diagram of classifications using CART (A, B, C, and D are independent variables, and H and L are dependent variables).

3. Air Pollution, Temperature, and the Number of Deaths from Cardiovascular Disease in Seoul

3.1. Air Pollution and Temperatures in Seoul

Temporal and spatial variations of PM_{2.5}, PM₁₀, O₃, and NO₂ in Seoul were analyzed. Ozone concentrations were highest in spring because of strong sunlight. Ozone concentrations in northern Seoul were relatively higher than in its southern section (Fig. 3). Nitrogen dioxide levels, as well as PM_{2.5} and PM₁₀ concentrations, were relatively higher in winter, partly due to the lower mixing height and relatively frequent inversion in winter. NO₂ concentrations were relatively higher in southwestern Seoul. A relatively low level of particulate matter in summer was related to the season's frequent rainfall. In addition, increased vertical

Fig. 3. Spatial distribution of pollutant concentrations.

Fig. 4. Average seasonal temperatures in Seoul.

mixing and diffusion of the atmosphere because of the high altitude of the mixing layer was a factor in the relatively low levels of $PM_{2.5}$ and PM_{10} near the ground in summer [18].

The average summer temperature in Seoul has increased or decreased from 2008 to 2013 by approximately 2-3°C (Fig. 4). The average temperature did not consistently increase or decrease from 2008 to 2013. However, summer temperatures trended upward except in 2010. Such increases in temperatures in Seoul were similar to the pattern of global change in summer temperatures. As summers have trended warmer, winters have become colder. The trend of lower winter temperatures from 2008 to 2013 was explained, ironically, by global warming. Cold Siberian air, previously trapped in the northern latitudes by a once strong winter jet stream, could move south because of a now weakened winter jet stream caused by global warming. Thus, winter temperatures in Korea are colder [19].

3.2. Number of Deaths Caused by Cardiovascular Disease

Statistics showed that cancer was the number one cause of death in Korea (Fig. 5). Cardiovascular disease, the focus of this study, ranked second. The number of deaths from cardiovascular disease in Seoul, Korea from 2008 to 2013 was 47,839 [11]. The number of deaths per age group slowly increased up to age 65, after which the rate of increase became quite obvious, indicating that people older than 65 were those most vulnerable to cardiovascular disease (Fig. 6) [20]. Consequently, people age 65 and older were the targets of our analyses to determine the factors that affect death from the cardiovascular disease. In Seoul from 2008 to 2013, 38,112 people older than 65 died from cardiovascular disease.

Relatively more people older than 65 died from cardiovascular disease in winter (Fig. 7). The number of these deaths per 1,000 people was relatively higher in the western and northern sections of Seoul (Fig. 7). Here, the reason for accounting for deaths according to 1,000 people rather than by sectional population

Fig. 5. Cause-specific mortality rates in Korea in 2013.

Fig. 6. Total number of deaths from cardiovascular disease in Seoul, Korea, between 2008 and 2013.

was to show the ratio of deaths because the different sectional numbers of populations would influence the absolute number of deaths. A comparison of the spatio-temporal distribution of the number of deaths from air pollution and temperatures could yield useful information, but the linkages could not be analyzed accurately with average air pollution counts and the number of the deaths alone. Therefore, daily data were used to analyze the linkage of air pollution and temperatures with the number

Fig. 7. Mortality rates for cardiovascular disease (over 65 years old) per 1,000 people in Seoul between 2008 and 2013.

of deaths from cardiovascular disease. As noted earlier, the population studied consisted of persons 65 or older at the time of their death because this group was known as vulnerable to cardiovascular disease.

4. Results and Discussion

4.1. Factors Linked with Cardiovascular Disease

4.1.1. Number of deaths, air pollutant concentrations, and temperatures in Seoul

Each day between 2008 and 2013 was categorized as ranking

high (H) or low (L) according to the number of deaths on that day based on the median number of deaths of persons 65 and older from cardiovascular disease in Seoul. The monthly average temperature and air pollutant concentrations of O₃, NO₂, PM_{2.5}, and PMC were separately plotted for H and L days (Fig. 8).

Monthly average temperatures and O₃ concentrations were similar between H and L days (Fig. 8). However, monthly average NO₂, PM_{2.5}, and PMC concentrations were relatively higher on H days than on L days (Fig. 8). As such, air pollution was assumed to be related to the number of deaths because several elements of air pollution concentrations were quite

Fig. 8. Temperature and pollutant concentrations when the number of deaths from cardiovascular disease is higher (H) or lower (L) than the daily median number of deaths.

different on H and L days. However, factors that influenced the number of the deaths could not be isolated by comparing monthly average pollutant concentrations alone. To resolve such a limitation, the CART model was used to analyze daily data. The daily data were also used to analyze the delay time between the point of exposure to air pollution and the time of death.

The number of deaths analyzed included deaths from both acute and chronic cardiovascular disease. If the analysis was conducted separately for acute and chronic death cases, the accuracy of the results could be improved. However, since the separate data was not available, the analysis was conducted with the total number of deaths. In addition, by expanding the analytical target to the number of visitations to emergency rooms for specific diseases, the influence of air pollution and temperature could be more accurately identified.

4.1.2. Lag time between deaths and exposure to air pollution

Elevated air pollutant levels and temperature changes can directly lead to the death of patients, but patients can also die a few days after the exposure. Related studies showed that

elevated levels of air pollution had a direct impact on health for up to two days after exposure [18]. Studies had also showed a delay between a change in temperature and a spike in the number of deaths. The lag time varied depending on the season, but the lag time in summer was relatively shorter than in winter [21].

Air pollutant concentrations and temperatures a few days before death were also examined to see the lag time. Differences in air pollution and temperatures between H and L days were largest in May (Fig. 8). Air pollution and temperatures before deaths in May were represented in Fig. 9. Thus, the bar for a lag time of zero in Fig. 9 corresponded to the value in May in Fig. 8. The bar on the lag time for one day showed air pollution and temperatures one day before deaths. Differences in the number of deaths between H and L days decreased as the lag time increased from zero to four days. Temperatures on the lag time of a zero day differed little between H and L days. However, the difference increased as the lag time increased, indicating that temperature before death could have direct connection with deaths.

Fig. 9. Average temperatures (lag: zero ~ 25 d) and air pollutant concentrations (lag: zero ~ 4 d) in May between 2008 and 2013 before deaths. H is temperature and air pollution before a day with a higher probability of death, and L is temperature and air pollution before a day with a lower probability of death.

4.1.3. Factors linked with deaths due to cardiovascular diseases in Seoul analyzed through CART

Days of higher or lower numbers of deaths were classified based on the median number of deaths throughout Seoul. Since the mean number was more prone to be affected by the extreme value, the median number was used. The use of median number, instead of the median number of daily deaths from cardiovascular disease between 2008 and 2013 was calculated for each season. The median number of daily deaths for spring, summer, fall, and winter was 17, 15, 17, and 19, respectively (Table 1).

The CART model was used to analyze factors that influence the daily number of deaths from cardiovascular disease. The dependent variable was categorized as days with higher and lower numbers of daily deaths compared with the median numbers of deaths. Independent variables were daily levels of air pollution ($PM_{2.5}$, PM_{10} , O_3 , and NO_2) and temperatures. To see the differences in air pollution and temperatures on the day of deaths and before death, air pollution up to four days before the day of death was used as the independent variable. Temperature was known to have a relatively long-term effect compared with air pollution in terms of time before the day of death [13]; consequently, average temperatures between 11 and 20 d before death were also used as variables. Here, the span of 11 to 20 d was selected in order to decrease the number of independent variables in CART. Ideally, temperature in each day should be separately use, but if so, the number of variables became too large. Then, the accuracy of classification may decrease [16]. Thus, temperature between 11 and 20 d were grouped before it was used in CART.

The CART model was used to separately analyze each season (Fig. 10). The results of the CART model in Fig. 10 could be interpreted as follows [16-17]. The top branch was created based on the average temperature from 11 d to 20 d before deaths: $[T(11-20\text{ d})]$ as the most important factor in determining higher or lower numbers of deaths in spring (Fig. 10). The risk of death was higher when $[T(11-20\text{ d})]$ was lower or equal to 7°C . In addition, if $[T(11-20\text{ d})]$ were higher than 7°C , the branch was subdivided based on $PM_{2.5}$ on the day of death: $[PM_{2.5}(0\text{ d})]$. If $[PM_{2.5}(0\text{ d})]$ was lower than $28\text{ }\mu\text{g}\cdot\text{m}^{-3}$, CART classified such days as carrying a lower risk of death. In that way, CART added more branches until 10 or more data were in each node. Results showed that temperature, $PM_{2.5}$, and NO_2 were closely related to the number of deaths caused in spring by cardiovascular disease (Fig. 10).

Among temperature and the four forms of pollutants, temperature was found to be the most important factor in deaths in

summer (Fig. 10). However, unlike spring, the risk of death was higher when the average temperature one day before death: $[T(1\text{ d})]$ was higher than 29°C . The risk of death was also higher when the $[PM_{2.5}(0\text{ d})]$ was higher than $31\text{ }\mu\text{g}\cdot\text{m}^{-3}$ although $[T(1\text{ d})]$ was lower than 29°C . Temperature was also the most important factor in fall. The risk of death was higher when the average temperature two days before death: $[T(2\text{ d})]$ was lower or equal to 9°C (Fig. 10). The risk of death was lower when the $[PM_{2.5}(0\text{ d})]$ was lower or equal to $29\text{ }\mu\text{g}\cdot\text{m}^{-3}$ although the $[T(2\text{ d})]$ was higher than 9°C (Fig. 10).

Fig. 10. Factors affecting the high (H) and low (L) probability of death from cardiovascular disease. H and L days are classified based on the median number of daily deaths in all of Seoul.

Table 1. Basic Statistics of the Daily Number of Deaths from Cardiovascular Disease in Seoul 2008-2013

	Spring	Summer	Fall	Winter
Median	17	15	17	19
Mean	17.6	15.5	17	19.4
Max	33	30	32	34
Min	7	4	7	8

The result of CART analysis showed that temperature was also the most influential factor on deaths from cardiovascular diseases in winter. The risk of death was higher when $[T (11-20 \text{ d})]$ was lower than minus 7°C (Fig. 10). The risk of death was also higher when $[PM_{2.5} (0 \text{ d})]$ exceeded 35 gm^{-3} , although $[T (11-20 \text{ d})]$ was higher than 7°C . Through this analysis, temperatures and air pollutant concentrations were found to be closely related with the daily number of deaths from cardiovascular disease.

The accuracy of the CART analysis was evaluated by using a 10-fold cross validation. Because relevant documents with detailed descriptions of 10-fold cross validation are readily available [22, 23], only a concise description of the concept is given here. First, data were divided into 10 groups, and a CART model was constructed using data from only nine groups. The model's error was calculated by testing the constructed model with the data of the unused group. The average error calculated after repeating the above process ten times becomes the 10-fold cross validation error.

The 10-fold cross validation errors per season were 12% (spring), 29% (summer), 20% (fall), and 13% (winter), respectively. The errors were acceptable when compared with other studies in the environmental fields. When CART was applied to predict ground-level O_3 concentrations over southwestern Taiwan, the errors were between 17 and 28% [15]. Another study that used CART to predict summer season maximum surface O_3 for the Vancouver, Montreal, and Atlantic Regions of Canada, errors were between 40 and 80% [24]. Park (2016) also used CART to predict O_3 concentrations, and errors were around 10% [25]. Marshall used classification and regression trees in clinical epidemiology, and indicated that the limitations of the tree model to decrease errors below a certain point [26].

4.2. Effect of Geographic Location on Health Study Result

4.2.1. Geographic properties of monitoring stations

Factors linked with deaths from cardiovascular disease could differ depending on the choice of monitoring stations. The 25 air pollution and temperature monitoring stations in Seoul were categorized according to their geographical type: residential, public, and green area. Then, factors affecting H and L days were analyzed separately for these three geographical types to ascertain the different effects associated with data from different monitoring stations (Fig. 11).

Fig. 11. Ten-fold cross validation error.

4.2.2. Comparisons of monthly average air pollution and temperatures at different monitors

Monthly average air pollutant concentrations and temperatures on H days and those on L days were compared separately for different monitoring locations. Differences in monthly average O_3 concentrations and temperatures between H and L days were insignificant for all three types of monitors. However, the differences were obvious for NO_2 , $PM_{2.5}$, and PMC (Fig. 12). Air pollution and temperatures on the day of death up until three days before the day of death were also observed. The plot for May was an example. Although the differences in air pollution and temperatures between H and L days were obvious, differences among residential, public, and green stations were not obvious (Fig. 13).

Similar studies have been conducted; some of them categorized stations as roadside or non-roadside monitors to find the health impact of traffic-related air pollution [27-29]. Those studies indicated that especially NO_x concentrations were relatively higher in roadside monitors since a large part of NO_x emissions was originated from the on-road mobile source. However, this study divided the 25 monitoring stations, which were all among the non-roadside monitors, into residential, public, and green areas. Thus, the differences were not significant.

4.2.3. Comparisons of factors linked with deaths among data used from different monitors

Factors that influence the number of deaths was analyzed separately using CART with data from three locations (Fig. 14). Those factors were similar, but differences existed in the threshold of the independent variables that distinguished high and low numbers of deaths. In spring, the primary factor that distinguished H and L days was $T (11-20 \text{ d})$ regardless of the monitoring location. When $T (11-20 \text{ d})$ was lower than the threshold, higher chances of deaths from the cardiovascular disease were expected. Although $T (11-20 \text{ d})$ was higher than 6.5°C , if $PM_{2.5} (0 \text{ d})$ was higher than 24.5 gm^{-3} , the probability of deaths increased in Green area. Besides $T (11-20 \text{ d})$, factors affecting the number of deaths in Public area included $NO_2 (0 \text{ d})$ and $PM_{2.5} (0 \text{ d})$, and those in Residential area were $PM_{2.5} (0 \text{ d})$, $T (5-10 \text{ d})$, $NO_2 (0 \text{ d})$.

In summer, the major factor that determine the number of deaths was $T (1 \text{ d})$. If $T (1 \text{ d})$ was higher than $29-30.5^{\circ}\text{C}$ depending on the monitoring location, the probability of deaths increased. Other important factors in summer were $PM_{2.5} (0 \text{ d})$, $PM_{2.5} (2 \text{ d})$, and $T (2 \text{ d})$. The most important factor in fall was $T (2 \text{ d})$. If $T (2 \text{ d})$ was lower than $8-9^{\circ}\text{C}$ depending on the monitoring location, the probability of deaths increased. Besides $T (2 \text{ d})$, $PM_{2.5} (0 \text{ d})$ and $T (5-10 \text{ d})$ were also important in fall. $T (11-20 \text{ d})$ and $PM_{2.5} (0 \text{ d})$ were important factors to determine the number of deaths in winter.

In addition to these differences, there were differences in the 10-fold cross validation errors for different station locations (Table 2). Errors were relatively low for green areas and residential areas in spring. Errors for residential areas were lowest in summer and fall. Errors for green and public areas were lowest in winter; errors for residential areas were the lowest across all seasons. Based on the results, the air pollutant concentrations in resi-

Fig. 12. Average pollutant concentrations and temperatures in residential, public, and green areas between 2008 and 2013. H/L days are days with higher/lower numbers of deaths compared with the median number of daily deaths in Seoul. (a) O_3 , (b) NO_2 , (c) $PM_{2.5}$, (d) PMC , and (e) temperature.

dential area were most directly affecting factors to the number of deaths. The results could be used to select the monitoring locations of air pollution and temperature for the epidemiological studies.

Table 2. Ten-fold Cross Validation Errors When Using Data from Different Locations

	Spring	Summer	Fall	Winter
Green Area	11.6%	23.5%	19.8%	10.9%
Public Area	20.2%	25.7%	21.0%	10.4%
Residential Area	13.1%	19.8%	16.62%	13.9%

Fig. 13. Average pollutant concentrations (lag: zero ~ 3 d) and temperatures (lag: zero ~ 20 d) in residential, public, and green areas in May between 2008 and 2013 before deaths. H/L days are days with higher/lower numbers of deaths compared with the median numbers of daily deaths in Seoul. (a) O_3 , (b) NO_2 , (c) $\text{PM}_{2.5}$, (d) PMC, and (e) temperature.

Fig. 14. Factors affecting the high (H) and low (L) probability of death from cardiovascular disease. H and L days are classified based on the median number of deaths in Seoul. Air pollution and temperature data are from (a) green, (b) public, and (c) residential areas.

5. Conclusions

Factors linked with the number of deaths from 2008 to 2013 of persons over 65 y old from cardiovascular disease were analyzed in relation to air pollution and temperature. For this purpose, the analysis was conducted per season using a decision tree model. Independent variables were daily air pollutant concentrations (O₃, PM_{2.5}, PM₁₀, NO₂) and temperatures. The dependent variables were categorized as days with higher and lower numbers of daily deaths compared with the median numbers of daily deaths. Results of the analyses showed that the most influential factors on the number of deaths caused by cardiovascular disease were temperatures and PM_{2.5} concentrations. The temperature one day or two days before death was one of the important factors in deaths in summer and in fall, whereas, temperatures more than 10 days before death were a primary factor for deaths in winter and in spring. PM_{2.5} concentrations on the day of death were an important factor in deaths across all seasons.

When the analysis was conducted, air pollution and temperature monitors were categorized as three geographical types – residential, public, and green areas – and data collected from each type were analyzed separately. The accuracy of analysis through 10-fold cross validation errors was better with pollution and temperature data from monitors in residential areas compared with data used from green or public areas. As such, site selection for observation stations could influence the results

of a health impact assessment. The accuracy of analytical results was better when using data from the residential areas, indicating that people are highly affected by air pollution. Studies showed that accuracy could increase by using air pollutant concentrations from area of highest population density rather than average concentrations in the area.

The results of this study could be used as a guidance to select monitoring stations for epidemiological studies. In addition, this study could provide useful information for the health impact assessment analyses for similar studies of other causes of death, such as respiratory diseases. Furthermore, this study could be used as the base data for establishing environmental policies.

Acknowledgments

This paper was supported by the Research Fund, 2017, Pyeongtaek University in Korea.

This article was presented at the 2017 International Environmental Engineering Conference (IEEC2017) held on 15-17 November 2017, Jeju, Korea.

References

1. Billet S, Garcon G, Paget V, et al. Mutational pattern of TP53

- tumor suppressor gene in human lung cells exposed to air pollution. *Pollut. Atmos.* 2012;131-144.
2. Mauzerall DL, Tong Q. A preliminary estimate of the total impact of ozone and PM_{2.5} air pollution on premature mortalities in the United States. In: 27th international technical meeting on air pollution modeling and its application; 24-29 October 2004; Baniff Center, Baniff, Alberta, Canada. p. 102-108.
 3. Hoeppe P. Lung function and prevalence of irritations of eyes and airways on days with elevated ozone concentrations. *Immun. Infekt.* 1995;23:161-165.
 4. Clarke K, Kwon HO, Choi SD. Fast and reliable source identification of criteria air pollutants in an industrial city. *Atmos. Environ.* 2014;95:239-248.
 5. Anderson PK, Cunningham AA, Patel NG, Morales FJ, Epstein PR, Daszak P. Emerging infectious diseases of plants: Pathogen pollution, climate change and agrotechnology drivers. *Trends Ecol. Evol.* 2004;19:535-544.
 6. Katsouyanni K, Touloumi G, Spix C, et al. Short-term effects of ambient sulphur dioxide and particulate matter on mortality in 12 European cities: Results from time series data from the APHEA project. *Brit. Med. J.* 1994;314:1658-1663.
 7. Park SK, Cobb CE, Wade K, Mulholland J, Hu Y, Russell AG. Uncertainty in air quality model evaluation for particulate matter due to spatial variations in pollutant concentrations. *Atmos. Environ.* 2006;40:S563-S573.
 8. Patz JA, Campbell-Lendrum D, Holloway T, Foley JA. Impact of regional climate change on human health. *Nature* 2005;438:310-317.
 9. Son JY, Bell ML, Lee JT. Individual exposure to air pollution and lung function in Korea: Spatial analysis using multiple exposure approaches. *Environ. Res.* 2010;110:739-749.
 10. City of Seoul. Seoul air quality information [internet]. c2017 [cited 12 February 2017]. Available from: <http://cleanair.seoul.go.kr>.
 11. Statistics Korea. Korean statistical information service [internet]. c2017 [cited 24 March 2017]. Available from: <http://kostat.go.kr>.
 12. WHO. Classification of diseases [internet]. c2017 [cited 22 January 2017]. Available from: <http://www.who.int/classifications/icd/en>.
 13. Park J, Bae HJ, Seo YW. A study of environmental welfare policy for climate and environment-susceptible populations (I). *Korea Environ. Inst.* 2013;2013:1398-1632.
 14. Gardner MW, Dorling SR. Statistical surface ozone models: An improved methodology to account for non-linear behavior. *Atmos. Environ.* 2000;34:21-34.
 15. Chu HJ, Lin CY, Liao CJ, Kuo YM. Identifying controlling factors of ground-level ozone levels over southwestern Taiwan using a decision tree. *Atmos. Environ.* 2012;60:142-152.
 16. Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and regression trees. Wadsworth International Group, Belmont; 1984.
 17. Moon SS, Kang SY, Jitpitaklert W, Kim SB. Decision tree models for characterizing smoking patterns of older adults. *Expert Syst. Appl.* 2012;39:445-451.
 18. Kim SW, Yoon SC, Won JG, Choi SC. 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. Environ.* 2007;41:7069-7081.
 19. Choi Y, Park C. Distributions of cold surges and their changes in the Joongbu Region, the Republic of Korea. *Geogr. J. Korea* 2010;44:713-725.
 20. Bae HJ, Ha JS, Lim YL. Health impacts of climate change and air pollution – Effects of socioeconomic factors on mortality. *Korea Environ. Inst.* 2011;2011:1-117.
 21. Gong SY, Bae HJ, Yoon DO, Hong SP, Park HY. A study on the health impact and management policy on PM_{2.5} in Korea. *Korea Environ. Inst.* 2012;2012:1-209.
 22. Camacho J, Ferrer A. Cross-validation in PCA models with the element-wise *k*-fold (*ekf*) algorithm: Theoretical aspects. *J. Chemometr.* 2012;26:361-373.
 23. Meijer RJ, Geoman JJ. Efficient approximate *k*-fold and leave-one-out cross-validation for ridge regression. *Biometrical J.* 2013;55:141-155.
 24. Burrows WR, Benjamin M, Beauchamp S, et al. CART decision-tree statistical analysis and prediction of summer season maximum surface ozone for the Vancouver, Montreal, and Atlantic regions of Canada. *J. Appl. Meteor.* 1995;34:1848-1862.
 25. Park SK. Assessing factors linked with ozone exceedances in Seoul, Korea through a decision tree algorithm. *J. Environ. Sci. Int.* 2016;25:191-216.
 26. Marshall RJ. The use of classification and regression trees in clinical epidemiology. *J. Clin. Epidemiol.* 2001;54:603-609.
 27. Brucher W, Kessler C, Kerschgens MJ, Ebel A. Simulation of traffic-induced air pollution on regional to local scales. *Atmos. Environ.* 2000;34:4675-4682.
 28. Cai H, Xie S. Traffic-related air pollution modeling during the 2008 Beijing Olympic Games: The effects of an odd-even day traffic restriction scheme. *Sci. Total Environ.* 2011;409:1935-1948.
 29. Leary PJ, Kaufman JD, Barr RG, et al. Traffic-related air pollution and the right ventricle: The multi-ethnic study of atherosclerosis. *Am. J. Respir. Crit. Care Med.* 2014;189:1093-1100.