Improving Environmental Sustainability by Characterizing Spatial and Temporal Concentrations of Ozone

Kyu Jong Lee 1, Hyungu Kahng 1, Seoung Bum Kim 1,2,* and Sun Kyoung Park 2,2*

1 School of Industrial Management Engineering, Korea University, Seoul 02841, Korea; xtorm2@gmail.com (K.J.L.); drlego99@gmail.com (H.K.)

2 School of ICT-Integrated Studies, Pyeongtaek University, Pyeongtaek 17869, Korea
* Correspondence: sbkim1@korea.ac.kr (S.B.K.); skpark@ptu.ac.kr (S.K.P.); Tel.: +82-2-3290-3397 (S.B.K.); +82-31-659-8310 (S.K.P.)

Received: 15 October 2018; Accepted: 30 November 2018; Published: 2 December 2018

Abstract: Statistical methods have been widely used to predict pollutant concentrations. However, few efforts have been made to examine spatial and temporal characteristics of ozone in Korea. Ozone monitoring stations are often geographically grouped, and the ozone concentrations are separately predicted for each group. Although geographic information is useful in grouping the monitoring stations, the accuracy of prediction can be improved if the temporal patterns of pollutant concentrations is incorporated into the grouping process. The goal of this research is to cluster the monitoring stations according to the temporal patterns of pollutant concentrations using a k-means clustering algorithm. In addition, this study characterizes the meteorology and various pollutant concentrations linked to high ozone concentrations (>0.08 ppm, 1-h average concentration) based on a decision tree algorithm. The data used include hourly meteorology (temperature, relative humidity, solar insolation, and wind speed) and pollutant concentrations (O$_3$, CO, NO$_x$, SO$_2$, and PM$_{10}$) monitored at 25 stations in Seoul, Korea between 2005 and 2010. Results demonstrated that 25 stations were grouped into four clusters, and PM$_{10}$, temperature, and relative humidity were the most important factors that characterize high ozone concentrations. This method can be extended to the characterization of other pollutant concentrations in other regions.

Keywords: ozone; k-means clustering; decision tree algorithm; PM$_{10}$; temperature; relative humidity

1. Introduction

Ozone (O$_3$) is a secondary pollutant formed by photochemical reactions. The presence of nitrogen oxides (NO$_x$) and volatile organic compounds (VOC$_x$) accelerates ozone formation [1]. Major sources of NO$_x$ and VOC$_x$ include emissions from cars and trucks and various industrial facilities [2]. Because strong sunlight plays a vital role in ozone formation, ozone concentrations increase during summer. It must be noted that ozone in the stratosphere protects ecosystems from ultraviolet rays that are harmful to human beings. However, high concentrations of ozone near the ground, which is studied in this paper, adversely affects vegetation growth and human health. Studies showed that vegetation exposed to ozone decreased photosynthesis and growth [3,4]. Fuhrer et al. (2016) indicated that the elevated surface ozone level could cause substantial reductions in the agricultural yields. In addition, it may affect the ecosystem through decreasing species diversity of plants, animals, insects, and fish, etc. [4]. Prolonged exposure to high levels of ozone is directly linked with eye irritation, respiratory and cardiovascular diseases, etc., especially among children, the elderly, and patients [5]. To minimize the negative effect on public health associated with exposure to high levels of ozone, an ozone warning system was developed in Korea in July 1995.
The ozone warning system informs the public of predicted high ozone concentrations. If the 1-h average ozone concentration is expected to be higher than 0.12 ppm, ozone warning is informed, and if it is expected to be higher than 0.3 ppm, ozone alarm is informed to the public. When ozone warning is informed, sensitive groups of people, such as these older than 65 years old or children, are recommended to decrease the outdoor activities. When ozone alarm is informed, all people are recommended to decrease the outdoor activities to minimize the harmful health effect caused by the exposure to high levels of ozone.

Various statistical models have been used to predict ambient ozone concentrations. An autoregressive error model was developed to predict ozone levels in the southern area of Kyunggi Province, Korea [6]. Lee (2007) used eight meteorological variables (temperature, wind speed, cloud cover, solar insolation, humidity, precipitation, dew point, and vapor pressure) and four air pollutant concentrations (sulfur dioxide, nitrogen dioxide, carbon monoxide, and fine particulate matter) as independent variables. Results demonstrated that the autoregressive error model was successful in predicting short-term ozone concentrations [6]. An autoregressive exogeneous model was also developed to predict ozone levels in Eastern Austria [7]. Bauer et al. (2001) used present and past meteorology and air pollutant concentrations to predict ozone concentrations. El-Tahan (2018) analyzed the temporal and spatial distributions of ozone over Egypt for 12 years [8]. Analyses were conducted using data from both the Atmospheric Infrared Sounder (AIRS) and the model Modern-Era Retrospective Analysis for Research and Applications (MERRA). Results showed that seasonal variations of ozone concentrations from the AIRS were similar to those from MERRA. However, ozone concentrations from the MERRA were lower than those from the AIRS, indicating that the pollutant concentrations may have slightly different results depending on the choice of data caused by the uncertainty in the original data.

A linear regression model and principal component analysis were applied to predict daytime and night-time ozone concentrations [9,10]. Seven pollutant species (CH₄, NMHC, CO, CO₂, NO, NO₂, and SO₂) and five meteorological variables (wind speed and direction, air temperature, relative humidity, and solar radiation) were used. Results demonstrated that while temperature and solar insolation were directly linked to daytime ozone concentrations, night-time ozone concentrations were predominantly influenced by nitrogen oxides (NO + NO₂). The meteorological effect was minor for night-time ozone formation.

Ozone concentrations were also predicted by a feedforward artificial neural network model in two ways: using only independent variables in the analysis, and using only principal components of the variables [11]. Prediction results based on principal components were better than those based only on independent variables because data collinearity was reduced.

Several studies compared the performance of statistical models to predict ozone concentrations. A neural network model was compared to multiple regression analysis to predict summertime ozone concentrations in Seoul, Korea. Although both methods were useful in predicting ozone concentrations, predicted ozone concentrations from the neural network model were more accurate [12]. A separate study compared linear time series, artificial neural networks, and fuzzy models to predict high-level ozone concentrations in Santiago, Chile [13]. The study assessed model performance based on two indices: the ability of forecasting an episode and the tendency of a false positive (the tendency to report a high-level ozone episode that actually does not occur). Results demonstrated that all three models had similar performances in forecasting ozone episodes with a 70–95% success rate, but the fuzzy model was the most reliable because of its low false positive rate.

Some studies modeled ozone concentrations separately for each area to improve accuracy as temporal variations in pollutant concentrations, which differ from place to place. New data acquisition systems using “E-sensing” sensors could provide highly detailed information in particular cases where real-time data of ozone values are required [14]. For example, daytime maximum ozone concentrations were predicted separately in four areas of Seoul, Korea using a multiple linear regression model [15].
In addition, a transition function was developed separately for four areas in Seoul to predict ozone concentrations [16]. Both studies divided the modeling domain based on geography.

Although geographic information can be useful in dividing the modeling domain, a more accurate model can be developed if temporal patterns of pollutant concentrations were considered in the division of the domain. A previous study demonstrated that the modeling domain was successfully clustered solely by a temporal pattern of fine particulate matter (PM$_{2.5}$) concentrations [17]. The study used 24-h average PM$_{2.5}$ concentrations measured every third day between 2001 and 2005 and monitored at 522 sites in the United States. The study used a k-means clustering algorithm based on correlation distance to investigate the similarity between temporal patterns of PM$_{2.5}$ concentrations. In addition, the study demonstrated that a rotated principal component analysis (RPCA) was useful in characterizing spatial patterns of pollutant concentrations.

The goal of this study is to cluster monitoring stations based on temporal patterns of ozone concentrations. Thus, the hypothesis of this study is that ozone concentrations in Korea are associated with some meteorology and pollutants and show their own temporal and spatial patterns.

Monitors may be differently clustered depending on the choice of the algorithm. We have used the k-means clustering algorithm. In addition, the characteristics that were linked to high ozone concentrations were analyzed using a decision tree algorithm, which considers the correlation between variables to explain the characteristics of each cluster. The threshold to classify high ozone concentrations is 0.08 ppm of 1-h average concentrations. However, if a difference threshold is used, or the threshold is selected based on 8-h average concentrations, each cluster may have different characteristics. In summary, the limitation of this study is that similar analysis using the same data in this study can have slightly different results depending on the choice of the model or parameters.

This paper is organized as follows. Data used in this study are described in Section 2. Major analysis methods, the k-means clustering technique, and the decision tree algorithm are explained in Section 3. Finally, the results of the analysis and future applications of this research are presented in Section 4.

2. Materials and Methods

2.1. Data

This study used hourly air pollutant concentrations and meteorological variables measured in 25 monitoring stations in Seoul, Korea between 2005 and 2010 [18]. Monitors were operated by Korea Environment Corporation [19]. Heights of monitors were between 1.5 m and 10 m above the ground [20]. Pollutant species included hourly ozone (O$_3$), carbon monoxide (CO), nitrogen dioxide (NO$_2$), sulfur dioxide (SO$_2$), and fine particulate matter (PM$_{10}$). Meteorological variables included hourly temperature (in °C), relative humidity (in %), solar-insolation (in W·m$^{-2}$), and wind speed (in m·s$^{-1}$). All air pollutants were collected every 1 h. O$_3$, CO, NO$_x$, SO$_2$, and PM$_{10}$ were measured using the ultraviolet photometric method, non-dispersive infrared method, chemiluminescent method, pulsed UV fluorescence method, and β-ray absorption method, respectively. Monitors were regularly inspected following the operational guideline of air pollution monitoring system [20].

Missing data were estimated using the ordinary kriging technique, which calculates missing values by the weighted linear combination of known values (Equation (1)) [21]. Missing values were calculated using the “gstat” package of R software [22].

$$Y^*(S_0) = \sum_{i=1}^{n} a_i Y(s_i), \quad \sum_{i=1}^{n} a_i = 1 \quad (1)$$

Because high ozone concentrations usually occur in summer, the ozone warning period begins on 1 May and ends on 15 September of each year. Therefore, the analysis of this research focused on ozone concentrations between 1 May and 15 September from 2005 to 2010.
2.2. Methodology

The modeling domain was divided by a k-means clustering algorithm. The ozone concentrations of each clustered region were analyzed by a decision tree algorithm.

2.2.1. k-Means Clustering Algorithm

A k-means clustering algorithm systematically groups data by minimizing variances within the cluster while maximizing variances between clusters [23]. Variances are calculated based on the correlation distance between pollutant concentrations. While the Euclidean distance only measures the difference of the pollutant concentrations, the correlation distance also considers similarity among temporal patterns [17,24].

The k-means clustering algorithm groups monitoring stations in the following way. First, k number of centers are arbitrarily selected. Second, the correlation distance is calculated between the center and other stations. Third, measurement stations are clustered so that variances within clusters are minimized while variances between clusters are maximized. Fourth, the center of each cluster is re-selected so that the variance within the cluster is minimized. The third and fourth steps are then repeated until the center of each cluster is unchanged [25].

The analysis was performed on hourly ozone concentrations between 2000 and 2005, measured from 25 monitoring stations. The total data for each station were 19,872 (=24 h × 138 days × 6 years) because the hourly data from 1 May to 15 September of each year were used for the analysis. To ensure the appropriate clustering of stations through the k-means clustering algorithm, a locally linear embedding method was applied to the hourly ozone data with a dimension of 25 × 19,872. The locally linear embedding method, one of the widely used dimension reduction techniques, reduces the dimension by considering the spatial distribution of pollutant concentrations [26].

2.2.2. Decision Tree Algorithm

A decision tree algorithm uses a tree structure to represent a decision rule for classifying or predicting dependent variables [27]. The independent variables were selected based on the Gini index and were used to structure the decision tree [28]. The analysis was performed using the Classification and Regression Tree (CART) software, which has been widely used to predict and classify data [29].

Results were represented in a hierarchical structure following the “if-then” rule (Figure 1). The measured pollutant concentrations were distributed in the space composed of two independent variables (X₁, and X₂) (Figure 1a). Here, an example, which classified circles and rectangles, was represented. The intermediate node (in oval) shows which independent variables and what criteria are used to classify the independent variables in two groups (Figure 1b). The final nodes (in rectangle) in Figure 1b, which show the rules used for classifying data, correspond to each sector of Figure 1a, respectively.

Dependent variables included “high level (≥0.08 ppm, 1-h average concentration)” and “low level (≤0.08 ppm, 1-h average concentration)” ozone concentrations. A threshold of 0.08 ppm was selected as the ozone concentration, and concentrations higher than 0.08 ppm resulted in adverse health effects among children, the elderly, and patients [30].

The growth of the tree was stopped when the number of data points in the node reached 40, and the Gini index was used as a performance measure. The criteria that characterized high ozone concentrations were determined based on the Laplace accuracy, calculated by Equation (2) [31].

\[
\text{Laplace Accuracy} = \frac{n_c + 1}{n + p}
\]

where \( n \) is the number of observations in each node; \( n_c \) is the number of properly classified observations; \( p \) is the number of categories.
The number of observations \((n_i)\) indicated the total number of observations in each node, while the number of properly classified observations \((n_c)\) indicates the number of high ozone concentrations (>0.08 ppm). The number of categories \((p)\) is two because the dependent variable (ozone concentrations) are categorized into two classes: high and low concentrations.

\[
\text{Classification Error} = \frac{\text{Number of improperly classified observations}}{\text{Total number of observations}}
\]

\[
\text{Gini Index} = \frac{\sum_{i=1}^{p} n_i \cdot n_c}{\sum_{i=1}^{p} n_i^2}
\]

\[
\text{Information Gain} = \sum_{i=1}^{p} \frac{n_i}{n} \cdot \left( \frac{n_i}{n} \cdot \log \left( \frac{n_i}{n} \right) - \frac{n_c}{n} \cdot \log \left( \frac{n_c}{n} \right) \right)
\]

Figure 1. A sample diagram of a decision tree structure (a) two-dimensional representation, (b) hierarchical representation.

3. Results and Discussion

3.1. Spatial Characteristics of Ozone Concentrations

The results demonstrated that 25 stations in Seoul were clustered in four groups: the northern, central, southern, and eastern areas (Figure 2). It must be noted that stations in the same cluster were geographically closely located, even though the temporal pattern of ozone concentrations was primarily used for cluster stations.

| Cluster (Area) | No. of Station | Monitoring Station (Name) |
|---------------|----------------|----------------------------|
| Cluster 1 (North) | 5              | A (Dobong), B (Enpyung), C (Gangbook), D (Nowon), and E (Sungbook) |
| Cluster 2 (Center) | 7              | F (Dongdaemoon), G (Mapo), H (Jongro), I (Joong), J (Seodaemoon), K (Sungdong), and L (Yongsan) |
| Cluster 3 (South) | 8              | M (Gongduk), N (Gangseo), O (Geumcheon), P (Guro), Q (Gwak), R (Seocho), S (Yangcheon), and T (Yongdeungpo) |
| Cluster 4 (East)  | 5              | U (Gangdong), V (Gangnam), W (Gwangjin), X (Joonglang), and Y (Songpa) |

Figure 2. Monitoring stations clustered using the k-means clustering algorithm based on hourly ozone concentrations in Seoul.
Monitoring stations in each cluster from the k-means clustering method were closely located to each other in the reduced dimension (Figure 3). The results ensured that ozone concentrations of the same cluster exhibit similar temporal patterns. Ozone concentrations of other clusters were statistically different because the $p$-value of the F-test of the analysis of variance (ANOVA) was close to zero (Table 1). The ANOVA test allows for a comparison of more than two populations (groups). In other words, ANOVA is a statistical technique for testing whether more than two population means are all equal.

![Figure 3. Hourly average ozone concentrations plotted in the reduced dimension using the locally linear embedding.](image)

**Table 1.** ANOVA table for the test of mean difference between clusters.

| Source | DF | SS     | MS     | F      | $p$-value |
|--------|----|--------|--------|--------|-----------|
| Factor | 3  | 0.7064 | 0.2355 | 538.69 | 0         |
| Error  | 496,796 | 217.1545 | 0.0004 |        |           |
| Total  | 496,799 | 217.8609 |        |        |           |

### 3.2. Temporal Patterns of Ozone Concentrations

The annual mean ozone concentrations of each cluster between 2005 and 2010 illustrated that concentrations increased between 2005 and 2010 with significantly high concentrations in 2009 (Figure 4a). High ozone concentrations were expected in July and August as the ozone formation was directly related to the temperature. However, monthly average concentrations in May and June were obviously higher than those in July and August (Figure 4b). Relatively low concentrations in July and August were partly because of the low insolation, which was also one of the critical factors in ozone formation. Monthly average precipitation and monthly frequency of precipitation indicated that the total amount and the frequency of precipitation was significantly higher in July and August than in May and June (Figure 5). Therefore, average ozone concentrations were lower in July and August than in May even though the number of exceedance, in which the ozone levels are higher than the standard, was larger in July and August (Table 2).
Table 2. The number of exceeding environmental standards from May to September (1 h environmental standards: 0.1 ppm, issuing ozone warning standards: 0.12 ppm).

| Month     | No. of Exceeding 1 h Standard | No. of Issuing Ozone Warning |
|-----------|-------------------------------|-------------------------------|
| May       | 315                           | 27                            |
| June      | 1077                          | 280                           |
| July      | 549                           | 125                           |
| August    | 487                           | 148                           |
| September | 69                            | 7                             |

Figure 4. (a) Annual average ozone concentrations in each cluster. (b) Monthly average ozone concentrations in each cluster.

Figure 5. (a) Monthly average precipitation and (b) monthly frequency of precipitation in Seoul, Korea between 2005 and 2010.

Hourly average ozone concentrations in 25 stations had evident hourly variations. Ozone concentrations were the highest in the afternoon between 3:00 p.m. and 5:00 p.m., with a small bump at dawn around 4:00 a.m. (Figure 6). The afternoon peak was partly attributed to the reaction of ozone precursor materials emitted during the morning and afternoon traffic periods. Although ozone precursor materials, such as NOx, are emitted in the morning traffic period, peak ozone concentrations occur in the afternoon because ozone formation is favored in the presence of sunlight and high temperature. In addition, a small increase in ozone concentrations around 4:00 a.m. is caused by air pollutants, which are often isolated in the ground because of the low nocturnal planetary boundary layer [32].
3.3. Factors Determining High Ozone Concentrations

Meteorology and pollutant concentrations that characterized high ozone concentrations were analyzed using the decision tree algorithm. Independent variables included concentrations of four air pollutants (CO, NOx, PM10, and SO2) including four meteorological variables (temperature, relative humidity, solar insolation, and wind speed). The analysis using the decision tree algorithm in the northern area of Seoul was represented as an example (Figure 7). The criteria that resulted in high Laplace accuracy characterized high ozone concentrations (dotted line in Figure 7). The analysis demonstrated that high ozone concentrations in the northern area of Seoul were expected when the relative humidity was not more than 59.5%, the temperature was not lower than 22.75 °C, and the concentration of PM10 was not lower than 28.5 μg·m^{-3}.

Relative humidity, temperature, and PM10 concentrations were the key criteria for high ozone concentrations in all clusters in Seoul (Table 3). The thresholds of relative humidity and the temperature that characterized high ozone concentrations were similar between clusters, while those of the PM10 concentrations were apparently different, which indicates that PM10 was unevenly distributed in Seoul.

The analysis results in this study are consistent with previous studies that used similar statistical approaches. Chu et al. (2012) used a decision tree to identify controlling factors of ground-level ozone measured in five monitors in Taiwan. The study used temperature, wind speed, relative humidity, NOx, alkanes, alkenes, and aromatic hydrocarbons as independent variables. Results varied depending on monitoring stations, but in general, temperature, wind speed, NOx, and aromatic hydrocarbons were important factors [33]. Park (2016) has used meteorological variables and NO2 to find factors characterizing high ozone concentrations in Seoul, Korea. Results indicated that relative humidity and temperature, as well and NO2 concentrations, were primary factors of high ozone concentrations [34].

Various studies used statistical models to classify high ozone concentrations. However, few efforts have been made to find the regional factors that characterize high ozone concentrations through clustering monitoring stations. This study clustered monitoring stations with similar temporal patterns of ozone concentrations. Then, factors linked to the high ozone episode were analyzed in each cluster by using a decision tree algorithm. In that way, the uncertainty in estimated factors can decrease.
Table 3. Overall high ozone episode in Seoul and each cluster based on the CART model.

| Region      | Rule                                                                 |
|-------------|----------------------------------------------------------------------|
| Overall Seoul | PM$_{10}$ $\geq$ 29.5 $\mu$g·m$^{-3}$, Temperature $\geq$ 25.05 °C, Relative Humidity $\leq$ 60.5% |
| North       | PM$_{10}$ $\geq$ 28.5 $\mu$g·m$^{-3}$, Temperature $\geq$ 26.35 °C, Relative Humidity $\leq$ 59.5% |
| Center      | PM$_{10}$ $\geq$ 41.5 $\mu$g·m$^{-3}$, Temperature $\geq$ 26.05 °C, Relative Humidity $\leq$ 65.5% |
| South       | PM$_{10}$ $\geq$ 45.5 $\mu$g·m$^{-3}$, Temperature $\geq$ 26.75 °C, Relative Humidity $\leq$ 60.5% |
| East        | PM$_{10}$ $\geq$ 36.5 $\mu$g·m$^{-3}$, Temperature $\geq$ 26.35 °C, Relative Humidity $\leq$ 58.5% |

Figure 7. The decision tree structure of the northern cluster that characterizes high ozone concentrations.
4. Conclusions

We analyzed hourly ozone concentrations measured in 25 monitoring stations in Seoul, Korea between 2005 and 2010. The k-means clustering algorithm was applied, and monitoring stations were clustered in four groups: the northern, central, southern, and eastern areas. The decision tree algorithm was useful in analyzing meteorology and various pollutant concentrations that characterized high ozone concentrations.

Ozone warning in Seoul was separately reported for five areas divided by geographic information: central, northeastern, southwestern, southeastern, and southwestern areas. The accuracy of prediction could be improved if the ozone concentrations were separately predicted for four clusters as ozone concentrations of the same cluster demonstrating similar temporal variations. In future research, separate time series models of hourly ozone concentrations will be developed for each cluster. The analysis performed in this study can be further applied to ozone prediction on a national level. In addition, this method can be applied to other pollutant species in any region and time period of interest.

Author Contributions: K.J.L. conceptualized the problem, and prepared the original manuscript. H.K. developed the methodology. S.K.P. investigated and collected resources. S.B.K. reviewed and edited the writing, and supervised the research.

Funding: This research was supported by Brain Korea PLUS, Basic Science Research Program through the National Research Foundation of Korea funded by the Ministry of Science, ICT and Future Planning (NRF-2016R1A2B1008994), the Ministry of Trade, Industry & Energy under Industrial Technology Innovation Program (R1623371), and by Institute for Information & communications Technology Promotion grant funded by the Korea government (No. 2018-0-00440, ICT-based Crime Risk Prediction and Response Platform Development for Early Awareness of Risk Situation).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Oh, I.B.; Kim, Y.K.; Hwang, M.K. Ozone Pollution Patterns and the Relation to Meteorological Conditions in the Greater Seoul Area. *J. Korean Soc. Atmos. Environ.* 2005, 21, 357–365.
2. Carrol, R.J.; Chen, R.; George, E.I.; Li, T.H.; Newton, J.; Schmiediche, H.; Wang, N. Ozone Exposure and Population Density in Harris County, Texas. *J. Am. Statist. Assoc.* 1997, 92, 392–404. [CrossRef]
3. Felzer, B.S.; Cronin, T.; Reily, J.M.; Melillo, J.M.; Wang, X. Impacts of Ozone on Trees and Crops. *C. R. Geosci.* 2016, 339, 784–798. [CrossRef]
4. Fuhrer, J.; Martin, M.; Mills, G.; Heald, C.; Harmens, H.; Hayes, F.; Sharps, K.; Bender, J.; Ashmore, M. Current and future ozone risks to global terrestrial biodiversity and ecosystem processes. *Ecol. Evol.* 2016, 6, 8785–8799. [CrossRef] [PubMed]
5. Huerta, G.; Sansó, B.; Stroud, J.R. A Spatiotemporal Model for Mexico City Ozone Levels. *J. R. Stat. Soc. Ser. C Appl. Stat.* 2004, 53, 231–248. [CrossRef]
6. Lee, H.J. Analysis of Time Series Models for Ozone at the Southern Part of Gyeonggi-do in Korea. *J. Korean Soc. Atmos. Environ.* 2007, 23, 364–372. [CrossRef]
7. Bauer, G.; Deistler, M.; Scherrer, W. Time Series Models for Short Term Forecasting of Ozone in the Eastern Part of Austria. *Environmetrics* 2001, 12, 117–130. [CrossRef]
8. El-Tahan, M. Temporal and Spatial Ozone Distribution over Egypt. *Climate* 2018, 6, 1–15.
9. Abdul-Wahab, S.A.; Bakheit, C.S.; Al-Alawi, S.M. Principal Component and Multiple Regression Analysis in Modelling of Ground-Level Ozone and Factors Affecting its Concentrations. *Environ. Modell. Softw.* 2005, 20, 1263–1271. [CrossRef]
10. Cannistraro, G.; Cannistraro, M.; Cannistraro, A.; Galvagno, A. Analysis of Air Pollution in the Urban Center of Four Cities Sicilian. *Int. J. Heat Technol.* 2016, 34, S219–S225. [CrossRef]
11. Sousa, S.I.V.; Martins, F.G.; Alvim-Ferraz, M.C.M.; Pereira, M.C. Multiple Linear Regression and Artificial Neural Networks Based on Principal Components to Predict Ozone Concentrations. *Environ. Modell. Softw.* 2007, 22, 97–103. [CrossRef]
12. Kim, Y.G.; Lee, C.B. Development of Neural Network Model for Prediction of Daily Maximum Ozone Concentrations in Summer. *J. Korean Soc. Atmos. Environ.* **1984**, *10*, 224–232.

13. Jorquera, H.; Pérez, R.; Cipriano, A.; Espejo, A.; Letelier, M.V.; Acuna, G. Forecasting Ozone Daily Maximum Levels at Santiago, Chile. *Atmos. Environ.* **1998**, *32*, 3415–3424. [CrossRef]

14. Cannistraro, M.; Lorenzini, E. The Application of the New Technologies “E-Sensing” in Hospitals. *IJMT* **2016**, *34*, 551–557. [CrossRef]

15. Kim, Y.J. A Study on the Development of Operable Models Predicting Tomorrow’s Maximum Hourly Concentrations of Air Pollutants in Seoul. *J. Korean Soc. Atmos. Environ.* **1997**, *13*, 79–89.

16. Kim, Y.K.; Sohn, K.T.; Moon, Y.S.; Oh, I.B. Development of Transfer Function Model to Frecased Ground-level Concentration in Seoul. *J. Korean Soc. Atmos. Environ.* **1999**, *15*, 779–789.

17. Kim, S.B.; Temiyasathit, C.; Chen, V.C.P.; Park, S.K.; Sattler, M.; Russell, A.G. Characterization of Spatially Homogeneous Regions Based on Temporal Patterns of Fine Particulate Matter in the Continental United States. *J. Air Waste Manag. Assoc.* **2008**, *58*, 965–975. [CrossRef]

18. Yoo, E.C.; Park, O.H. The Assessment of Air Quality Monitoring Network Considering the Change of various Environmental Factors in Busan. *J. Korean Soc. Atmos. Environ.* **2006**, *22*, 405–420.

19. Korea Environment Corporation. Air Quality and Environment Management. Available online: [https://www.keco.or.kr/en/core/climate_air1/contentsid/1946/index.do](https://www.keco.or.kr/en/core/climate_air1/contentsid/1946/index.do) (accessed on 11 January 2018).

20. Ministry of Environment. Operational Guideline of Air Pollution Monitoring System; Ministry of Environment: Sejong City, Korea, 2016; NIER-GP2016-086.

21. Le, N.D.; Zidek, J.V. *Statistical Analysis of Environmental Space-Time Processes*; Springer Science and Business Media: Berlin, Germany, 2006.

22. The R Project for Statistical Computing. Available online: [https://www.r-project.org/](https://www.r-project.org/) (accessed on 12 February 2018).

23. Gordon, A.D. *Classification*; Chapman and Hall/CRC: Boca Raton, FL, USA, 1999.

24. Lee, K.J.; Kim, S.B.; Park, S. Daily, Seasonal, and Spatial Patterns of PM10 in Seoul, Korea. In Proceedings of the International Symposium on System Informatics and Engineering, Qingdao, China, 11–13 July 2011.

25. Tan, P.; Steinbach, M.; Kumar, V. *Introduction to Data Mining*, 2nd ed.; Pearson Addison-Wesley: Boston, MA, USA, 2016.

26. Roweis, S.T.; Saul, L.K. Nonlinear Dimensionality Reduction by Locally Linear Embedding. *Science* **2000**, *290*, 2323–2326. [CrossRef]

27. Moon, S.S.; Kang, S.; Jitpitaklert, W.; Kim, S.B. Decision Tree Models for Characterizing Smoking Patterns of Older Adults. *Expert Syst. Appl.* **2012**, *39*, 445–451. [CrossRef]

28. Breiman, L. *Classification and Regression Trees*, 1st ed.; Routledge: Abingdon, UK, 2017.

29. CADDIS Volume 4: Data Analysis. Classification and Regression Tree (CART) Analysis. Available online: [https://www.epa.gov/caddis-vol4/](https://www.epa.gov/caddis-vol4/) (accessed on 10 February 2018).

30. What’s CAI. Available online: [https://www.airkorea.or.kr/eng/cai/cai1](https://www.airkorea.or.kr/eng/cai/cai1) (accessed on 7 February 2018).

31. Wang, Y.; Xin, Q.; Coenen, F. A Novel Rule Ordering Approach in Classification Association Rule Mining. In *Machine Learning and Data Mining in Pattern Recognition*; Lecture Notes in Computer Science; Perner, P., Ed.; Springer: Berling/Heidelberg, Germany, 2007; Volume 4571, pp. 339–348.

32. Kim, S.; Yoon, S.; Won, J.; Choi, S. 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. [CrossRef]

33. Chu, H.; Lin, C.; Liau, C.; Kuo, Y. Identifying Controlling Factors of Ground-level Ozone Levels over Southwestern Taiwan using a Decision Tree. *Atmos. Environ.* **2012**, *60*, 142–152. [CrossRef]

34. Park, S. Assessing Factors Linked with Ozone Exceedances in Seoul, Korea through a Decision Tree Algorithm. *J. Environ. Sci. Int.* **2016**, *25*, 191–216. [CrossRef]