J. Internet Comput. Serv.
ISSN 1598-0170 (Print) / ISSN 2287-1136 (Online)
http://www.jics.or.kr

Application of Urban Computing to Explore Living Environment Characteristics in Seoul : Integration of S-Dot Sensor and Urban Data

Daehwan Kim[1,2] Woomin Nam[1,3] Keon Chul Park[1*]

ABSTRACT

This paper identifies the aspects of living environment elements ($PM^{2.5}$, PM^{10}, Noise) throughout Seoul and the urban characteristics that affect them by utilizing the big data of the S-Dot sensors in Seoul, which has recently become a hot topic. In other words, it proposes a big data based urban computing research methodology and research direction to confirm the relationship between urban characteristics and living environments that directly affect citizens. The temporal range is from 2020 to 2021, which is the available range of time series data for S-Dot sensors, and the spatial range is throughout Seoul by 500mX500m GRID. First of all, as part of analyzing specific living environment patterns, simple trends through EDA are identified, and cluster analysis is conducted based on the trends. After that, in order to derive specific urban planning factors of each cluster, basic statistical analysis such as ANOVA, OLS and MNL analysis were conducted to confirm more specific characteristics. As a result of this study, cluster patterns of environment elements($PM^{2.5}$, PM^{10}, Noise) and urban factors that affect them are identified, and there are areas with relatively high or low long-term living environment values compared to other regions. The results of this study are believed to be a reference for urban planning management measures for vulnerable areas of living environment, and it is expected to be an exploratory study that can provide directions to urban computing field, especially related to environmental data in the future.

☞ keyword : Urban Computing; Living Environment; Urban Factor; Urban Data; S-Dot; IoT

1. Introduction

Nowadays, smart city strategies are receiving a lot of attention as an attempt to solve urban problems. In order to realize such a smart city strategy, it is necessary to understand the major technologies of the 4th industrial revolution, such as the Internet of Things, artificial intelligence, and drones, and one of the key elements is urban computing based on big data[1-2]. A system (urban computing) that accumulates, manages, and uses various big data (environment, transportation, urban planning, communication, SNS, etc.) in the city has become an important urban management strategy[3]. Urban computing is the process of collecting, integrating, and analyzing big data and heterogeneous data generated from various sources in urban space, which helps to understand the nature of urban phenomena and predict the future of cities, and is gradually growing as an interdisciplinary field that fuses traditional urban-related fields with computer science and information technology[4].

Since 2020, Seoul has been installing Seoul IoT sensor(S-Dot) throughout the city to collect various living environment data. Since then, it has been gradually producing high-quality and high-resolution data through the sophistication and calibration of the S-Dot device compared to the past, creating a favorable environment for research by using that data. However, in the case of previous studies

1 Urban Computing & Innovation Lab, Advanced Institute of Convergence Technology(AICT), Suwon, 16229, South Korea
2 Dept. of Civil and Environmental Engineering, Seoul National University, Seoul, 08826, South Korea
3 Dept. of Applied Artificial Intelligence, Sungkyunkwan University, Seoul, 03063, South Korea
* Corresponding author (parkkc07@snu.ac.kr)
[Received 26 June 2023, Reviewed 04 July 2023, Accepted 08 August 2023]
☆ This work is supported by Seoul Data Fellowship funded by the Seoul Metropolitan Government in 2022 and the Korea Agency for Infrastructure Technology Advancement(KAIA) grant funded by the Ministry of Land, Infrastructure and Transport (Grant RS-2022-00143404).
☆ This work is an extension of a paper that was selected as the best paper at the 2023 Korean Society for Internet Information(KSII) Spring Conference and recommended for publication in the Journal of the Korean Internet Information Society (JICS).

http://dx.doi.org/10.7472/jksii.2023.24.4.65

utilizing S-Dot living environment data, most of them only superficially analyze the data itself, or only analyze some phenomena, so multifaceted research has not been conducted. Also, very few studies have analyzed living environment data in conjunction with urban data, which does not lead to consideration of actual urban policies.

Inaddition, while various problems are occurring due to the development and concentration of cities, respiratory problems caused by fine dust and exposure to frequent noise in cities are a global problem and concern. South Korea's annual PM concentrations are often among the worst in the OECD[5], and it is no secret that noise and air pollution from traffic and other sources continue to threaten the health of modern people[6]. At this stage, it is indispensable to check the trend of environmental data such as fine dust and noise in cities and what factors affect the environment in the process of creating a huge urban organization.

Against this background, this study has the following objectives. First, by attempting application study utilizing S-Dot big data,an urban data sensor in Seoul, we verify the current situation of S-Dot data and discuss possible directions for improvement and future utilization. Second, we closely identify the aspects of fine dust and noise data that are directly related to citizen' lives. Finally, we identify urban factors that affect each environmental data and propose directions for creating sustainable urban spaces. The temporal scope of the study is 2020.11.~2021.09. (11 months), a period when maximum data can be obtained through data exploration, and the spatial scope is 612 GRIDs(500m×500m) in Seoul, which were identified based on the available S-Dot locations during the temporal scope.

2. Related Works

2.1 Urban Computing Research related to Living Environments and Urban Elements

In managing the environment, past research has mainly focused on evaluating direct environmental hazards such as atmospheric factors. However, efforts have begun to incorporate factors such as land use and transportation into environmental analysis models to evaluate the impact of urban form on air quality.[7] The study of the environmental impacts of urban factors has gradually evolved into analyzing the environmental impacts of different scenarios of urban form[8,9] or to identify the relationship between urban spatial structure and environmental factors using direct statistical analysis such as regression models[10,11].

More specifically, several previous studies have identified a wide variety of urban planning factors that affect environmental variables such as fine dust(PM^{10}) and ultrafine dust($PM^{2.5}$). Several studies have proved the clear impact of road traffic on air pollution[12,13,14], others have demonstrated the impact of the total volume of roads on PM emissions[15,16,17], and still others have shown that roads with high traffic volumes are associated with particular matters[18]. In addition, a positive correlation between the frequency of bus stops and $PM^{2.5}$ has been found[19,20]. On the other hand, the role of land use in managing fine dust has been addressed in many studies, but there is no clear consensus except for industrial areas and green areas. Several studies have shown that industrial land use contributes to $PM^{2.5}$ emissions[15,17] and that green and park areas improve air quality[21,22]. However, studies differ on the impact of residential, commercial, and mixed-use on PM emissions and dispersion. Similarly, the air quality impacts of high-density population and housing, as envisioned by the Compact City concept, are controversial. Some studies have shown that higher density improves air quality by reducing emissions of air pollutants such as greenhouse gases[23,24], while others have shown that higher density increases air pollutants and worsens air quality[25].

There is also a large body of research interested in the relationship between noise environmental variables and urban factors. In order to identify the relationship between noise and urban space, studies mainly use noise maps to investigate the relationship between noise and urban density, shape, and traffic resilience[26,27]. In addition, there are studies that analyze the interrelationship between noise levels and urban form[28]. In particular, in noise-related studies, it is common to grid the urban area and model urban factor datas in each grid using GIS[29],and it is common practice to reflect spatiotemporally referenced noise values in the grid as representative [26,30,31].

Through a series of previous studies, we can confirm the results of studies to analyze the correlationbetween each living environment (fine dust, ultrafine dust, noise) and urban factors, and we set the variables for the analysis of this study based on them. In addition, according to the aim of this study, which identifies the living environment of Seoul, conduct the analysis by gridding the entire city of Seoul based on S-Dot.

2.2 Research of Time-series Data Analysis

K-Means Clustering is a well-known unsupervised machine learning algorithm that categorizes data into K types by measuring the distance between K user-set centroid points and the data[32,33]. Usually, this method has been applied to classify the types of data and predict the future value of data in various economic or social fields such as stock price and population[34].

In recent years, the research trend has been extended to environmental data, and K-Means cluster analysis has been applied to living environment data with time-series characteristics. For example, K-Means clustering of changes in environmental data such as barometric pressure, altitude, and wind has been used to categorize heat wave weather or to analyze the relationship between heat wave casualties and categorized weather[35]. In addition, there is also a study that utilizes the above cluster analysis method to classify types based on rainfall time series data to prepare a water resource management plan considering rainfall variability [36].

As there is a trend of studies that categorize time series data through K-Means clustering analysis and analyze the characteristics of the categorized data to solve problems or provide suggestions in the field, we expect that a similar methodology can be applied to this study using time series living environment data such as fine dust and noise.

3. Research Data and Process

3.1 Living Environmental Data from S-Dot

The living environments covered in this study are ‘ultra-fine dust($PM^{2.5}$)’, ‘fine dust(PM^{10})’, and ‘noise’. In the original data, S-Dot Data, there are additional living environment, but during the data review process, the three categories are finally determined as the dependent variables by considering the degree of missing values and outliers, the relevance to the actual lives of citizen, and the possible relationship with urban factors. The original dataset for the above three living environments is a bundle of hourly environmental data collected through S-Dot. For this, we use the average daily, monthly, seasonal, and annual data and the clustered values based on the GRID of each point to which the S-Dot sensor belongs as the analysis data. In addition, for ’temperature’ and ’humidity’ data, the annual average value is treated as an independent variable. All living environment data are from November 2020 to September 2021 (11 months) at the GRID level according to the spatial and temporal scope of this study.

3.2 Urban Factor Datasets

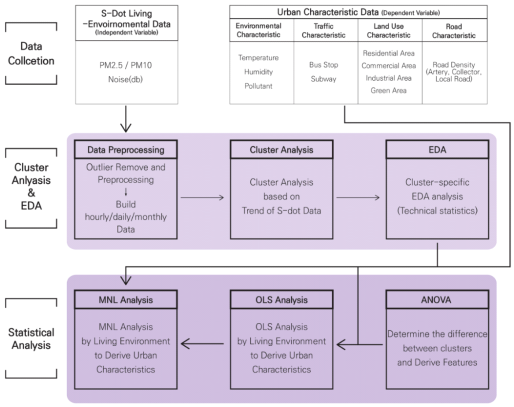
The urban factor datasets are composed of variables that have been described in previous studies as having a significant impact on the fine dust(including ultra-fine dust) and noise environmental categories. It consists of 14 variables in four main categories: ’environmental characteristics’, ’transportation characteristics’, ’land use characteristics’, and ’road characteristics’.

Density-related variables, which are often used in existing studies but have not been found to be significant, are excluded from this study because we believe that the variables in the ’transportation characteristics’ category contain some explanatory power related to density. Land use characteristics, which have differentimpacts on environment by each previous studies, are disaggregated and entered as variables, and road characteristics are also disaggregated to examine the hierarchical characteristics of roads (road types are classified into arterial roads, collector roads, and local roads in South Korea based on the function of roads). We also bundle temperature and humidity with facilities that act as pollution sources as variables in the ’environmental characteristics’.The detailed data sources, definitions, and measurement methods for each variable are shown in Table 1. These variables are used as independent variables for statistical determination of urban factor characteristics.

(Table 1) Dataset for urban factors

Characteristic	Variables	Raw Data (Data Source)	Definition	Measurement	Temporal Range
Environmental	Temperature	S-Dot Data (Seoul City)	Average annual temperature for each GRID	GRID representative valuesby preprocessing S-Dot data	2020.11. ~ 2021.09.
	Humi-dity		Average annual humidity for each GRID		
	Pollu-tant	Street Address Data (Ministry of the Interior and Safety)	Pollutanr building coverage ratio for each GRID	(Pollution source building area) ÷ (GRID Area)	2021.09.
Tranportation	Bus Stop	Bus stop locations (Seoul City)	Number of bus stops for each GRID	Count	2021.12.
	Subway Station	Subway Stationby by Line (Seoul City)	Number of subway stations for each GRID	Count	
Land Use	Low-rise Residential Area	Seoul Zoning Districts _Urban Areas (Seoul City)	Low-rise residential area for each GRID	(Area of each zoning districts) ÷ (GRID Area)	2021.12.
	High-rise Residential Area		High-rise residential area for each GRID		
	Commercial Area		Commercial area for each GRID		
	Industrial Area		Industrial area for each GRID		
	Green Area		Green area for each GRID		
Road	Total Road Density	Road Section (Ministry of the Interior and Safety)	Total road density for each GRID	(Total Road Area) ÷(GRID Area)	2021.12.
	Arterial Road Density		Arterial road density for each GRID	(Road Area by road level) ÷(GRID Area)	
	Collector Road Density		Collector road density for each GRID		
	Local Road Density		Local road density for each GRID		

3.3 Research Process



(Figure 1) Research process

Figure 1 summarizes the research process of this study. First, we conduct an exploratory data analysis (hereinafter referred to as EDA analysis) to check the cluster pattern of each category of living environment data. For this purpose, we check the optimal K through the Elbow Method for the S-Dot living environment time series data (ultra-fine dust K=8, fine dust K=11, noise K=7). Furthermore, by comparing and contrasting the values of each cluster, we identify significant clusters and their characteristics, and define reference category clusters for further statistical analysis.

Next, a multi-way ANOVA is performed as a statistical analysis to identify differences between clusters of living environment data and urban factor characteristics. The factor variable is the cluster value of each living environment sector, and the dependent variable is the average value of the entire period of the living environment data. This confirms the differences between the clusters derived from the living environment data and serves as logical evidence that the estimates determined in the previous EDA analysis may be significant. The data of the three living environment sectors used in this study are all confirmed to be ’not equally distributed’ through Levene’s analysis, so a ’Welch’ criterion ANOVA is performed.

Then, OLS analysis (Ordinay Least Squares) is conducted with 14 pre-processed urban factor characteristics data as independent variables. Since the purpose of the OLS analysis is to identify the general effects of urban factors on the living environment, all 612 GRIDs are sampled. Theconditions for regression analysis such as normality of each data used in the analysis were found to be satisfied. However, in the case of 'Total Road Density', multicollinearity was identified and removed from the final analysis.

The last step is to conduct MNL(Multi-Nominal Logistic Regression). The MNL analysis identifies the urban factor characteristics that specifically affect the value of each cluster in a situation where all variables are controlled. In other words, it compares the impact (probability) of urban factors on each cluster relative to the cluster that serves as the reference category based on the trends identified in the previous EDA analysis. The reference category is Cluster 3 for ultra-fine dust($PM^{2.5}$), Cluster 11 for fine dust(PM^{10}), and Cluster 3 for noise. As with the OLS analysis, we use all 14 independent variables and exclude the "Total Road Density" variable. IBM SPSS Statistics26 is used for all EDA analysis and statistical analysis.

4. Result

4.1 Characteristics of Urban Factors in Lliving Environment identified by ANOVA and OLS

The results of the ANOVA analysis confirm that the differences in values are statistically significant in the seasonal and monthly averages for all three living environment categories (ultra-fine dust, fine dust, and noise). This indicates that the difference in trend between the clusters identified for each living environment is significant and gives us the justification to identify the factors that affect the difference in living environment values. This indicates that the difference in trend between the clusters identified for each living environment is significant and gives us the justification to identify the factors that affect the difference in living environment values. To this end, we conduct an OLS analysis with urban planning characteristics as independent variables, and the dependent variable is the average value of living environment values over the entire period for each GRID. The results of the analysis are shown in Figure 2 (explanatory power(R2) of the regression equation is 0.189 for ultrafine particulate matter, 0.183 for fine particulate matter, and 0.119 for noise).

〈Table 2〉 OLS analysis result for ultra-fine and fine dust

	Ultra-fine Dust			Fine Dust		
	β	t	p	β	t	p
(Constant)		-.956	.339		-.177	.860
Temperature	.163	1.821	*.069	.091	1.022	**.007
Humidity	.141	1.551	.121	.067	.735	.463
Pollutant	.017	.320	.749	.027	.504	.614
Bus Stop	.033	.703	.482	.017	.358	.721
Subway Station	-.050	-1.187	.236	-.032	-.754	.451
Low-rise Residential Area	-.010	-.167	.867	.021	.342	.732
High-rise Residential Area	-.133	-1.799	*.073	-.103	-1.392	.164
Commercial Area	-.052	-.897	.370	-.009	-.147	.883
Industrial Area	-.056	-.858	.391	-.028	-.432	.666
Green Area	.025	.364	.716	.048	.688	.492
Arterial Road Density	.028	.651	.516	.041	.951	.342
Collector Road Density	.129	2.499	*.013	.130	2.514	*.012
Local Road Density	.061	1.235	.217	.089	1.792	*.074
		R^2=0.189			R^2=0.183	

***p<0.001. **P<0.01. *P<0.1. ***p<0.001. **P<0.01. *P<0.1.

First, the urban factors that affect the concentration of ultra-fine dust are identified as 'Temperature', 'Collector Road', and 'High-rise Residential Area', with regression coefficients of 0.163, 0.129, and -0.133, respectively. Specifically, the higher the annual average temperature within a given GRID and the more roads that function as collector roads, the higher the concentration of ultra-fine dust, and the higher the area of high-rise residential areas within the GRID, the lower the concentration of that.

In the case of fine dust, the urban factors that affect its concentration are identified also as 'Temperature', 'Collector Road Density', and 'Local Road Density', with regression coefficients of 0.091, 0.130, and 0.089, respectively. It can be interpreted that the higher the average annual temperature within a given GRID, and the more roads with collector and local road functions, the higher the concentration of fine dust. In addition, a very similar result is found for ultra-fine dust, confirming that urban factors do not differ significantly in their impact on the two living environment sectors.

(Table 3) OLS analysis result for noise

	Noise		
	β	t	p
(Constant)		7.390	.000
Temperature	-.442	-5.179	***.000
Humidity	-.307	-3.531	***.000
Pollutant	.090	1.750	*.081
Bus Stop	.042	.958	.339
Subway Station	.011	.277	.782
Low-rise Residential Area	-.089	-1.542	.123
High-rise Residential Area	.011	.160	.873
Commercial Area	.073	1.329	.184
Industrial Area	.044	.698	.485
Green Area	.028	.420	.674
Arterial Road Density	.205	4.961	***.000
Collector Road Density	.083	1.685	*.093
Local Road Density	-.022	-.460	.646

R^2=0.119

***p<0.001. **P<0.01. *P<0.1.

Finally, for noise, five variables have a statistically significant impact on the magnitude of noise, and their regression coefficients are as follows: ‘Temperature’ is -0.442, ‘Humidity’ is -0.307, ‘Pollutant’ is 0.090, ‘Arterial Road Density’ is 0.205, and ‘Collector Road Density’ is 0.083. This means that the lower the average annual temperature and humidity, or the more building uses with potential pollution sources, and the more roads with arterial and collector functions, the higher the average noise within a given GRID.

4.2 Characteristics of Urban Factors in Lliving Environment identified by MNL Analysis

As mentioned above, in determining the reference category for MNL analysis through EDA analysis, the focus is on selecting clusters which annual trends in concentration (fine and ultrafine dust) and magnitude (noise) remain withinthe standard value (median) and comparing them to clusters with higher or lower trends. Accordingly, the finalized reference categories for each living environment sector are Cluster 3 for ultra-fine dust, Cluster 11 for fine dust, and Cluster 3 for noise.

(Table 4) MNL analysis result for ultra-fine dust

Reference Category = Cluster3	Cluster2			Cluster6		
	B	Wald	p	B	Wald	p
(Constant)	50.887	7.250	.007	84.912	20.132	.000
Temperature	-1.844	6.292	*.012	-3.920	28.089	***.000
Humidity	-.394	6.888	**.009	-.443	8.857	**.003
Pollutant	.000	.070	.792	.000	5.982	*.014
Bus Stop	.010	.070	.791	-.004	.008	.929
Subway Station	-.112	.227	.634	-.013	.003	.954
Low-rise Residential Area	-.002	.026	.873	-.004	.110	.741
High-rise Residential Area	.008	1.095	.295	.006	.637	.425
Commercial Area	.014	2.246	.134	.006	.430	.512
Industrial Area	-.003	.081	.776	-.030	3.309	*.069
Green Area	.015	1.565	.211	.014	1.194	.274
Arterial Road Density	-.036	1.318	.251	-.045	1.833	.176
Collector Road Density	-.078	7.332	**.007	-.058	3.752	*.053
Local Road Density	-.043	1.731	.188	.031	.894	.344

Chi-Square(Sig)=206.664(0.000)

***p<0.001. **P<0.01. *P<0.1.

For ultra-fine dust, the reference category (Cluster 3) and MNL analysis is conducted for Cluster 2 and Cluster 6, which have maintained a low annual trend in concentration. For Cluster 4, which has a high trend, no significant variables were identified in the analysis. In Cluster2 and Cluster6, the common significant variables are ’Temperature’, ’Humidity’, and ’Collector Road Density’. This is consistent with the previous OLS results, indicating that a given GRID is more likely to be classified as Cluster 3 (with higher annual average temperature and humidity and more roads that function as collector roads) than Cluster 2 or Cluster 6. Furthermore, the variables ’Pollutant’ and ’Industrial area’ are found to be significant in Cluster 6. In the case of ’pollutant’, the B-value is 0.000, which is almost insignificant, but the ’Industrial area’ can be interpreted to mean that the more industrial areas in a given GRID, the higher the probability of being classified as Cluster 3 (areas with higher annual ultra-fine dust concentrations than Cluster 6).

In the case of fine dust, MNL analysis is performed based on the reference category (Cluster 11) for Cluster 2 and Cluster 6, which are identified as having relatively low

(Table 5) MNL analysis result for fine dust

Reference Category = Cluster11	Cluster2			Cluster6			Cluster9		
	B	Wald	p	B	Wald	p	B	Wald	p
(Constant)	36.293	2.109	.146	62.733	8.116	.004	-4.781	.036	.849
Temperature	-1.235	1.504	.220	-2.848	11.069	***.001	-.554	.323	.570
Pollutant	.000	.482	.488	.000	2.872	*.090	.000	3.258	*.071
Bus Stop	.001	.001	.981	.011	.060	.807	.005	.009	.922
Subway Station	.210	.310	.578	.187	.336	.562	-.256	.390	.532
Low-rise Residential Area	-.027	2.209	.137	-.014	1.032	.310	-.014	.577	.448
Commercial Area	-.005	.165	.684	-.018	1.987	.159	-.019	1.517	.218
Industrial Area	-.017	.421	.517	-.025	1.889	.169	-.021	1.146	.284
Green Area	-.002	.016	.899	-.002	.017	.895	-.015	.705	.401
Arterial Road Density	.032	.509	.476	-.052	.991	.320	.111	7.105	**.008
Collector Road Density	-.122	7.058	**.008	-.051	2.015	.156	.038	.933	.334
Local Road Density	-.058	1.506	.220	.010	.060	.807	.063	1.848	.174

Chi-Square(Sig)=234.431(0.000) ***p<0.001, **P<0.01, *P<0.1.

annual trends in concentration, and Cluster 9, which is identified as having high annual trends. Cluster 3 also belongs to the group with a high annual trend, but the number of samples is quite small and no significant variables are identified in the analysis, so it is excluded from the analysis results. First of all, the significant variable in Cluster 2 is 'Collector Road Density', which means that if there are many roads with collector road function in a particular GRID, the probability of being classified as Cluster 11 (areas with higher annual fine dust concentration than Cluster 2) is higher, as shown in the previous OLS results. For Cluster 6, three variables are found to be significant: Temperature', Humidity', and 'Pollutant'. However, "Pollutant" is found to be a non-significant variable, with a B-value of 0.000, as was the case for ultra-fine dust. From this, it can be interpreted that if the average annual temperature and humidity within a particular GRID is high, the probability of being classified as Cluster 11 (areas with higher annual fine dust than Cluster6) is high. On the other hand, in the case of Cluster 9, 'Pollutant' and 'Arterial Road Density' are significant, but the former variable is less plausible for the reasons mentioned above. Therefore, through the latter variable, it can be interpreted thatif there are many roads with arterial road functions in a specific GRID, the probability of being classified as Cluster 9 (an area with higher annual fine dust concentration than Cluster 11) is higher.

(Table 6) MNL analysis result for noise

Reference Category = Cluster3	Cluster2			Cluster6		
	B	Wald	p	B	Wald	p
(Constant)	60.634	6.558	.010	32.985	3.101	.078
Temperature	-3.100	10.824	***.001	-1.246	2.809	*.094
Humidity	-.299	2.704	.100	-.289	4.042	*.044
Pollutant	.000	2.606	.106	.000	.296	.586
Bus Stop	.018	.093	.761	.046	1.022	.312
Subway Station	-.191	.262	.609	.110	.128	.720
Low-rise Residential Area	-.010	.187	.665	.021	1.986	.159
High-rise Residential Area	.001	.008	.928	.007	.561	.454
Commercial Area	.017	1.007	.316	.012	.752	.386
Industrial Area	-.010	.232	.630	.004	.088	.766
Green Area	.000	.000	.985	.019	2.292	.130
Arterial Road Density	.151	12.638	***.000	.041	1.073	.300
Collector Road Density	.081	3.747	*.053	-.039	1.287	.257
Local Road Density	.154	8.469	**.004	.161	15.776	***.000

	Cluster1			Cluster7		
	B	Wald	p	B	Wald	p
(Constant)	-25.682	2.207	.137	14.796	.400	.527
Temperature	1.165	2.893	*.089	-.651	.511	.475
Humidity	.106	.645	.422	-.133	.521	.470
Pollutant	.000	.028	.867	.000	.849	.357
Bus Stop	-.002	.001	.971	.046	.718	.397
Subway Station	.024	.007	.933	-.219	.261	.610
Low-rise Residential Area	.030	5.294	*.021	.031	3.306	*.069
High-rise Residential Area	.009	.901	.343	.009	.509	.476
Commercial Area	.009	.539	.463	-.004	.038	.845
Industrial Area	-.005	.165	.685	-57.588	.002	.960
Green Area	.015	1.780	.182	-.001	.003	.954
Arterial Road Density	-.036	.792	.374	.049	1.006	.316
Collector Road Density	-.023	.555	.456	.023	.260	.610
Local Road Density	.213	32.930	***.000	.097	3.824	*.051

Chi-Square(Sig)=221.705(0.000) ***p<0.001. **P<0.01. *P<0.1.

Lastly, for noise, we conduct MNL analysis with the reference category (Cluster3) for Cluste1 and Cluster7, which have a low average annual magnitude of noise, and Cluster2 and Cluster4, which have a high trend in noise magnitude. First, Cluster1 and Cluster7 share a common characteristic: both the variables 'Local Road Density' and 'Low-rise Residential Area' are significant. This means that the more local roads and the larger the area of low-rise residential areas in a given GRID, the more likely it is to be classified as Cluster 1 or Cluster 7 (areas with lower annual noise than Cluster 3). In other words, residential areas that do not have a large number of arterial or collector road features are classified as GRIDs with lower annual noise, which is the same result as in OLS. Meanwhile, the significant variables in Cluster 2 are 'Temperature', 'Arterial Road Density', 'Collector Road Density', and 'Local Road Density'. This means that, similar to the previous OLS results, the lower the average annual temperature or the more arterial and collector roads within a given GRID, the more likely it is to be classified as Cluster 2 (i.e., areas with higher annual noise than Cluster 3). In addition, "Local Road Density" was also found to be a significant variable in Cluster 2, suggesting that areas with a significantly higher road density would have been classified in Cluster 2 regardless of the hierarchy. Furthermore, in Cluster 4, the significant variables are 'Temperature', 'Humidity', and 'Local Road Density', which is consistent with the OLS result that lower average annual temperature and humidity within a given GRID are more likely to have higher annual noise.

5. Conclusion and Implication

5.1 Conclusion

This study categorized clusters according to the trends of ultra-fine dust, fine dust, and noise in the living environment in Seoul and identified urban factors that affect those clusters. It is inevitable that there will be areas in the city that have relatively high or low levels oflong-term living environment compared to other areas, and while there are many factors that contribute to the development of such areas, it is possible to identify the influence of some urban factors (environmental characteristics, land use characteristics, road characteristics, etc.).

Based on the results of this study, it can be concluded that there are certain characteristics of urban factors that affect the concentration of fine dust and the loudness of noise in a situation where several characteristics are controlled. First of all, it can be observed that the living environment (ultrafine particles, fine particles, and temperature) is sensitive to basic and natural environmental characteristics (i.e., temperature and humidity). All three categories of living environment are significantly affected by 'temperature' and 'humidity'(positive correlation for ultrafine and fine dust concentrations and negative correlation for noise), which is a well-known natural phenomenon, and the results of this study also confirm this fact. In addition, noise is positively correlated with 'pollutant'. In this study, pollutant refers to the density of buildings of a given use that are potential sources of air and noise pollution. This confirms that areas of the city with a high density of buildings of that use do indeed have higher average annual noise levels.

However, in addition to natural phenomena, we should pay more attention to the effects of urban organizations (elements) created through urban planning. In particular, when it comes to roads, smaller roads (collector roads, local roads) that are directly connected to city traffic have a significant impact on the living environment, rather than roads for high-speed driving with a large scale. In the case of 'collector road', a positive correlation is found for all three living environments. A collector road is a road that connects traffic on local roads to secondary arterial roads that connect to residences and various facilities at the end of traffic, and is a kind of node that connects large and small roads through an appropriate road width. In other words, areas with a high density of collector roads are bound to have high traffic volume and frequent population movement, and it can be interpreted that such characteristics are closely related to the living environment such as ultra-fine dust, fine dust, and noise. In addition, the density of local roads is positively correlated with the concentration of fine dust, and it can be assumed that the characteristics of local roads that connect multiple buildings within a block generate as much traffic and pedestrian traffic as collector roads, and the average concentration of fine dust increases accordingly. On the other hand, for noise, all types of road densities are correlated, confirming that the presence of roads has a significant impact on noise.

Some effects of land use characteristics are also identified on this study. First, we find that areas with large areas of 'industrial zones' are associated with higher concentrations of fine dust, which supports previous research findings. It is noteworthy that the extent of 'high-rise residential area' is negatively correlated with ultra-fine dust. High-rise residential areas, as defined in this study, are residential areas with high residential density and a relatively comfortable living environment according to the Korean Land Planning Act. Therefore, even within the same city, the larger the area of such residential zones, the better the concentration of fine particulate matter is managed.

5.2 Implication

The relationships between urban factors and the living environment identified in this study suggest that there are parts that need to be further explored beyond those that have been demonstrated in previous studies. However, actions such as removing collector roads or allocating more low-rise residential areas to reduce fine dust concentrations or noise are impossible to control at the city level, and one-size-fits-all urban policies will never be effective. However, this studyhas identified clusters of environmentally vulnerable neighborhoods, and it is necessary to take appropriate measures to reduce fine dust and noise in these areas. Furthermore, for newly developed urban areas, it is expected that the results of this study can be used to proactively prevent the layout and design of roads, temperature and humidity management, and the layout and management of buildings that are potential sources of pollution.

This study also enhances the possibility and importance of linking urban data with living environment data. This study categorized clusters based on time series data of three environmental sectors directly related to the lives of citizen and identified their patterns. Furthermore, it can be seen as a cornerstone or direction for the production of living environment maps. By analyzing the situation of Seoul by linking multiple data instead of a single living environment data, we can better understand the city and provide directions for improvement. Finally, based on the analysis results, we suggest directions for reducing fine dust and noise from an urban planning and policy perspective.

Above all, in the future, more detailed analysis of the living environment in Seoul can be achieved by inputting more urban factor datas orchanging the analysis model. It would also be possible to analyze the structure of the grid in more detail. We hope that the quality research will continue in the future.

References

[1] M. H. Chung and J. H. Kim, "The Internet Information and Technology Research Directions based on the Fourth Industrial Revolution," *KSII Transactions on Internet and Information Systems*, vol. 10, no. 3, pp. 1311-1320, 2016.
https://doi.org/10.3837/tiis.2016.03.020

[2] H. A. Almashhadani, X. Deng, O. R. Al-Hwaidi, S. T. Abdul-Samad, M. M. Ibrahm and S. N. A. Latif, "Design of A new Algorithm by Using Standard Deviation Techniques in Multi Edge Computing with IoT Application," *KSII Transactions on Internet and Information Systems*, vol. 17, no. 4, pp. 1147-1161, 2023.
https://doi.org/10.3837/tiis.2023.04.006

[3] C. Rabari and M. Storper, "The digital skin of cities: Urban theory and research in the age of the sensored and metered city, ubiquitous computing and big data," *Cambridge Journal of Regions, Economy and Society*, Vol. 8, No. 1, pp. 27-42, 2015.
https://doi.org/10.1093/cjres/rsu021

[4] Y. ZHENG, "Urban computing: enabling urban intelligence with big data," *Frontiers of Computer Science*, Vol. 11, No. 1, pp. 1-3, 2017.
https://doi.org/10.1007/s11704-016-6907-2

[5] IQAir, "2019 world air quality report. Region & City PM2. 5 Ranking," 2020.

[6] World Health Organization, "Burden of Disease from Environmental Noise: Quantification of Healthy Life Years Lost in Europe," *World Health Organization. Regional Office for Europe*, 2011.

[7] L. O. Marquez and N. C. Smith, "A framework for linking urban form and air quality," Environmental Modelling & Software, vol. 14, no. 6, pp. 541-548, 1999.https://doi.org/10.1016/s1364-8152(99)00018-3

[8] C. Borrego, H. Martins, O. Tchepel, L. Salmim, A. Monteiro and & A. I. Miranda, "How urban structure can affect city sustainability from an air quality perspective," *Environmental modelling & software*, vol. 21, no. 4, pp. 461-467, 2006.
https://doi.org/10.1016/j.envsoft.2004.07.009

[9] U. W. Tang and Z. S. Wang, "Influences of urban forms on traffic-induced noise and air pollution: Results from a modelling system," *Environmental Modelling & Software*, vol. 22, no. 12, pp. 1750-1764, 2007.
https://doi.org/10.1016/j.envsoft.2007.02.003

[10] B. Stone Jr, "Urban sprawl and air quality in large US cities," *Journal of environmental management*, vol. 86, no. 4, pp. 688-698, 2008.
https://doi.org/10.1016/j.jenvman.2006.12.034

[11] L. P. Clark, D. B. Millet and J. D. Marshall, "Air quality and urban form in US urban areas: Evidence from regulatory monitors," *Environmental Science & Technology*, vol. 45, no. 16, pp. 7028-7035, 2011.
https://doi.org/10.1021/es2006786

[12] J. P. Shi, D. E. Evans, A. A. Khan and R. M. Harrison, "Sources and concentration of nanoparticles (< 10 nm diameter) in the urban atmosphere," *Atmospheric environment*, vol. 35, no. 7, pp. 1193-1202, 2001.
https://doi.org/10.1016/s1352-2310(00)00418-0

[13] P. Pant, and R. M. Harrison, "Estimation of the contribution of road traffic emissions to particulate matter concentrations from field measurements: A review," *Atmospheric environment*, vol. 77, pp. 78-97, 2013.
https://doi.org/10.1016/j.atmosenv.2013.04.028

[14] A. Rakowska, K. C. Wong, T. Townsend, K. L. Chan, D. Westerdah *et al.*, "Impact of traffic volume and composition on the air quality and pedestrian exposure in urban street canyon," *Atmospheric Environment*, vol. 98, pp. 260-270, 2014.
https://doi.org/10.1016/j.atmosenv.2014.08.073

[15] D. K. Moore, M. Jerrett, W. J. Mack and N. Künzli, "A land use regression model for predicting ambient fine particulate matter across Los Angeles, CA," *Journal of Environmental Monitoring*, vol. 9, no. 3, pp. 246-252, 2007.
https://doi.org/10.1039/b615795e

[16] I. Aguilera, J. Sunyer, R. Fernández-Patier, G. Hoek, A. Aguirre-Alfaro *et al.*, "Estimation of outdoor NO x, NO2, and BTEX exposure in a cohort of pregnant women using land use regression modeling," *Environmental science & technology*, vol. 42, no. 3, pp. 815-821, 2008.
https://doi.org/10.1021/es0715492

[17] W. Liu, X. Li, Z. Chen, G. Zeng, T. León *et al.*, "Land use regression models coupled with meteorology to model spatial and temporal variability of NO2 and PM10 in Changsha, China," *Atmospheric Environment*, vol. 116, pp. 272-280, 2015.
https://doi.org/10.1016/j.atmosenv.2015.06.056

[18] G. S. Hagler, E. D. Thoma and R. W. Baldauf, "High-resolution mobile monitoring of carbon monoxide and ultrafine particle concentrations in a near-road environment," *Journal of the Air & Waste Management Association*, vol. 60, no. 3, pp. 328-336, 2010.
https://doi.org/10.3155/1047-3289.60.3.328

[19] S. H. Park and D. W. Ko, "Investigating the effects of the built environment on PM2. 5 and PM10: A case study of Seoul Metropolitan city, South Korea," *Sustainability*, vol. 10, no. 12, pp. 4552, 2018.
https://doi.org/10.3390/su10124552

[20] L. Guo, J. Luo, M. Yuan, Y. Huang, H. Shen *et al.*, "The influence of urban planning factors on PM2. 5 pollution exposure and implications: A case study in China based on remote sensing, LBS, and GIS data," *Science of The Total Environment*, vol. 659, pp. 1585-1596, 2019.
https://doi.org/10.1016/j.scitotenv.2018.12.448

[21] W. J. Bealey, A. G. McDonald, E. Nemitz, R. Donovan, U. Dragosits *et al.*, "Estimating the reduction of urban PM10 concentrations by trees within an environmental information system for planners," *Journal of Environmental Management*, vol. 85, no. 1, pp. 44-58, 2007.
https://doi.org/10.1016/j.jenvman.2006.07.007

[22] H. Feng, B. Zou and Y. Tang, "Scale-and region-dependence in landscape-PM2. 5 correlation: Implications for urban planning," *Remote Sensing*, vol. 9, no. 9, pp. 918, 2017.
https://doi.org/10.3390/rs9090918

[23] E. L. Glaeser and M. E. Kahn, "The greenness of cities: Carbon dioxide emissions and urban development," *Journal of urban economics*, vol. 67, no. 3, pp. 404-418, 2010.
https://doi.org/10.1016/j.jue.2009.11.006

[24] B. Bereitschaft and K. Debbage, "Urban form, air pollution, and CO2 emissions in large US metropolitan areas," *The Professional Geographer*, vol. 65, no.4, pp. 612-635, 2013.
https://doi.org/10.1080/00330124.2013.799991

[25] H. S. Cho and M. J. Choi, "Effects of compact urban development on air pollution: Empirical evidence from Korea," *Sustainability*, vol. 6, no. 9, pp. 5968-5982, 2014.
https://doi.org/10.3390/su6095968

[26] E. M. Salomons and M. B. Pont, "Urban traffic noise and the relation to urban density, form, and traffic elasticity," *Landscape and Urban Planning*, vol. 108, no. 1, pp.2-16, 2012.
https://doi.org/10.1016/j.landurbplan.2012.06.017

[27] H. J. Ryu, I. K. Park, S. I. Chang and B. S. Chun, "The Spatial Statistical Relationships between Road-traffic Noise and Urban Components Including Population, Building, Road-traffic and Land-use," *Transactions of the Korean Society for Noise and Vibration Engineering*, vol. 24, no. 4, pp. 348-356, 2014.
https://doi.org/10.5050/ksnve.2014.24.4.348

[28] L. T. Silva, M. Oliveira and J. F. Silva, "Urban form indicators as proxy on the noise exposure of buildings," *Applied acoustics*, vol. 76, pp. 366-376, 2014.
https://doi.org/10.1016/j.apacoust.2013.07.027

[29] M. Kawabata and J. Ferreira Jr, "GIS modeling of urban components to monitor and model urban respiration," *CSIS Discussion Paper*, no. 58, 2003.

[30] E. Gaja, A. Gimenez, S. Sancho and A. Reig, "Sampling techniques for the estimation of the annual equivalent noise level under urban traffic conditions," *Applied Acoustics*, vol. 64, no. 1, pp.43-53, 2003.
https://doi.org/10.1016/s0003-682x(02)00050-6

[31] H. J. Ryu, K. J. Hee, S. I. Chang and B. C. Lee, "A Study on Sampling Techniques to Assure the Representativeness of Short-term Equivalent Noise Level," *Transactions of the Korean Society for Noise and Vibration Engineering*, vol. 22, no. 12, pp. 1213-1219, 2012.
https://doi.org/10.5050/ksnve.2012.22.12.1213

[32] J. H. Jung and J. Nam, "Types and Characteristics Analysis of Human Dynamics in Seoul Using Location-Based Big Data," *Journal of Korea Planning Association*, vol. 54, no. 3, pp. 75-90, 2019.
https://doi.org/10.17208/jkpa.2019.06.54.3.75

[33] Q. Wang, X. Ouyang and J. Zhan, "A Classification Algorithm Based on Data Clustering and Data Reduction for Intrusion Detection System over Big

Data," *KSII Transactions on Internet and Information Systems(TIIS)*, vol. 13, no. 7, pp. 3714-3732, 2019. https://doi.org/10.3837/tiis.2019.07.021

[34] K. Han, H. Yang and H. Choo, "Deep Prediction of Stock Prices with K-Means Clustered Data Augmentation," *Journal of Internet Computing and Services*, vol. 24, no. 2, pp. 67-74, 2023. https://doi.org/10.7472/jksii.2023.24.2.67

[35] H. C. Lee, Y. J. Cho, B. H. Lim and S. B. Kim, "Study on the Association of Casualties and Classification of Heat Wave Weather Patterns in South Korea Using K-Means Clustering Analysis," *Journal of The Korean Society of Hazard Mitigation*, vol. 20, no.3, pp. 11-18, 2020. https://doi.org/10.9798/kosham.2020.20.3.11

[36] W. J. Wang, S. E. Kim, J. H. Song, J. H. Lee, K .T. Kim *et al.*, "Cluster and Trend Analysis of Rainfall Time Series in the Nakdong River Basin," *Journal of Korean Society of Hazard Mitigation*, vol. 22, no. 2, pp. 205-213, 2022. https://doi.org/10.9798/kosham.2022.22.2.205

◐ 저 자 소 개 ◑

김 대 환(Daehwan Kim)
2020년 서울대학교 건설환경공학부 공학석사
2020년~현재 서울대학교 건설환경공학부 박사과정
2020년~현재 차세대융합기술연구원 연구원
2022년 서울시 서울데이터펠로우쉽사업 펠로우 대표
관심분야 : 도시공학, 데이터사이언스, 어반컴퓨팅, 스마트시티
E-mail : dhbit@snu.ac.kr

남 우 민(Woomin Nam)
2020년 군산대학교 컴퓨터정보통신공학과 정보통신공학전공(공학사)
2022년 성균관대학교 대학원 인공지능융합학과(공학석사)
2022년~현재 차세대융합기술연구원 도시데이터혁신연구실 연구원
관심분야 : 스마트시티, 머신러닝, 딥러닝
E-mail : applewoods@snu.ac.kr

박 건 철(Keon Chul Park)
2011년 연세대학교 대학원 정보시스템학과 석사
2015년 연세대학교 대학원 정보시스템학과 박사
2015년~2016년 성균관대학교 인터렉션사이언스학과 박사후 연구원(Post-doc)
2016년~2021년 서울디지털재단 정책연구팀장, 데이터혁심팀장
2021년~현재 차세대융합기술연구원 도시데이터혁신연구실 실장(선임연구원)
관심분야 : 스마트시티, 공유경제, 디지털경제, 지능형정부
E-mail : parkkc07@snu.ac.kr