Research for Development of a Daejeon Fine Dust Prediction Model through Weather Data and Air Pollutants

Tae-Hyung Kim, Jae-Ou Lee

Abstract Background/Objectives: Fine dust is a first-tier carcinogen that is adversely affecting animals, plants, and people. Since 2005, the Ministry of Environment has been implementing fine dust forecasts to manage such fine dust. The forecast study conducted research on the causes of fine dust and how to cope with it. Methods/Statistical analysis: In this study, a model for predicting fine dust was developed through phase of multiple regression analysis. Independent variable can largely sort the forecast data and the air pollutants, and the dependent variable is fine dust (PM_{10}). It consists of a total of fifty variables and data has been collected on a monthly basis from 2006 to 2017. Findings: The result of regression from the model with the highest explanatory power shows that at the magnitude of the definition, a large total evaporation loss was the highest with .492 followed by CO with .264. Partial effect was 1.0m, average ground temperature was -.566, and highest temperature was -.325 in order. Therefore, fine dust (PM_{10}) increases when the large total evaporation loss and CO concentration increases. On the other hand, fine dust decreases when the average ground temperature and the highest temperature decreases. As a result, fine dust is influenced by diverse factors, and their effects show both increase and decrease. Improvements/Applications: Through this research, to identify factors that influence fine dust and use those in areas to manage fine dust in the future. In the final fine dust prediction model, the 1.0m average underground temperature and the maximum temperature were negatively affected, while the carbon monoxide and the large layer emissions were positively affected.

Index Terms: fine dust, phases of regression analysis, weather data, air pollution, multiple linear regression model.

I. INTRODUCTION

Clean Air Conservation Act defines air pollutant as 61 kinds of gas substances, such as sulfur dioxide, ozone, carbon monoxide etc. PM_{10} and $PM_{2.5}$ are the two types of fine dust. PM_{10} Has the diameter of 10$\mu g/m^3$ and $PM_{2.5}$, fine particulate matter, has the diameter of 2.5$\mu g/m^3$ [1]. In October 2013, the World Health Organization (WHO) has classified the fine particulate matter as the first-class carcinogen. Such fine dust gets placed on the surface of plants inhibiting metabolism and causes corrosion of buildings. When people inhale it in large amounts, it causes respiratory and inflammatory reactions, and increases cardiovascular diseases.

The research on the analysis of the causes of fine dust began to take place actively since 201. Lee analyzed the chemical composition, degree of contamination, and emission characteristics by collecting PM_{10} and $PM_{2.5}$ fine dust samples. The analysis shows that mass density of $PM_{2.5}$ accounts for about 61% of PM_{10}, and the fine dust formation is most affected by artificial contaminants depending on energy use, followed by soil or oceans [2]. Cho analyzed Seoul's annual PM_{10} and $PM_{2.5}$ density, and confirmed that the average density of PM_{10} is higher at 41 $\mu g/m^3$ and 23 $\mu g/m^3$, respectively [3]. Monthly average density of PM_{10} and $PM_{2.5}$ was the highest in January at 60 $\mu g/m^3$, 35 $\mu g/m^3$ and was the lowest in August at 23 $\mu g/m^3$, 14 $\mu g/m^3$. In addition, the density of fine dust varied from 10 P.M. (highest density) to 6 A.M. (lowest density) [4].

Since 2002, the Ministry of Environment has enacted a "Special Act for Improving the Atmospheric Environment in the Capital Region" to reduce the density of fine dust and developed a statistic forecast model for fine dust. Based on the statistic forecast model, Seoul city enacted the "Regulations on the Seoul dust forecast and Alert" and implemented the forecast since 2005. Presently, the Air Quality Integrated Administration Center predicts the level of fine dust pollution across the nation four times a day in four stages (good, normal, bad, very bad)[5].

Through analysis of the fine dust forecast model, the mechanism on fine dust emissions and causes and the process of spreading pollutants in the air can be understood, providing data to reduce the amount of pollutants, and bringing technological advances in air pollution management, such as forecasts and measurements. In this study, the cause-and-effect between atmospheric environmental data and fine dust(PM_{10}) density data, provided by the Korea Meteorological Administration and the Korea Environment Management Corporation, was intended to determine the causes of change in fine dust density. The analysis data were abstracted monthly from 2006 to 2017.

Revised Manuscript Received on May 13, 2019.

Tae-Hyung Kim, Department of Fire and Disaster Prevention, Daejeon University, Daejeon city, Korea, rucachi@naver.com.
Jae-Ou Lee, Department of Fire and Disaster Prevention, Daejeon University, Daejeon city, Korea, krokasig@dju.kr, +82-10-4691-3456

II. THEORETICAL BACKGROUND AND RESEARCH METHODOLOGY

A. MULTIPLE LINEAR REGRESSION MODEL

National Institute of Environmental Research studied the regression model, neural network model, and the decision making method model to develop forecast model. Each model has its own strengths and weaknesses. Amongst the models, the regression model, utilized in this study, weather measurement value and air quality measurement density are used as variables. To predict the past trends to easily develop and use, it is easy to develop and use as equation for past trends and has a good predictive rate for average values. However, it has the weakness of not being able to predict rapidly changing high density cases.

Multiple linear regression model is a model that uses regression analysis and predicts by mathematical methods to lessen the error between dependent variable (PM_{10}) and several independent variables (e.g. meteorological data, density of air pollutants, etc.). In regression analysis, a mathematical formula indicating the relationship between variables, such as Y=f(X), is called the regression equation. The regression model can describe the relationship between an independent variable and a dependent variable, or predict the value of the dependent variable with the value of the independent variable. For example, if you estimate the mathematical model Y=k(X) for the relationship between fine dust density (Y) and average precipitation (X), you could not only describe the relationship between fine dust density and precipitation, but also predict the fine dust density of that day by entering the rainfall on a specific day[6].

In the multiple linear regression model, the relationship between the dependent variable Y and K with the independent variable $X_1, \cdots, X_K$ is hypothesized in the following.

$$Y_i = a_0 + a_0 X_{1i} + \cdots + a_k X_{ki} + e_i \cdot \cdot \cdot \cdot (1)$$

Therefore, dependent variable is represented as the primary function of each independent variable, which adds probability variable e that shows the following clause. In equation (1), $a_0$is the slope of the y-axis, and a_i is the slope between Y andX_i, indicating the extent to which X_i affects Y when other independent variables are fixed. I refers to the sequence of observation data. Therefore, the equation for an observation points can be expressed in vectors in the following [7].

$$Y_i = a \cdot X + e \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot (2)$$

Here, it is defined.

$$Y = \begin{vmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{vmatrix}, X = \begin{vmatrix} X_{11} X_{21} X_{31} & \cdots & X_K \\ X_{12} X_{21} X_{31} & \cdots & X_K \\ \vdots \quad \vdots \quad \vdots & \vdots & \vdots \\ X_{1n} X_{21} X_{31} & \cdots & X_K \end{vmatrix}, a = \begin{vmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{vmatrix}, e = \begin{vmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{vmatrix}$$

At this point, multiple linear regression analysis is performed as follows.

First, when estimated regression factor, b = ($b_0, b_1, b_2, \cdots, b_k$), is defined, the residual using the sample above can be used as

$$e_i = Y_i - Y_i = Y_i - (b_0 + b_{1X} + \cdots + b_k X_{ki}) \cdot \cdot \cdot \cdot (3)$$

And also using vector,

$$e = Y - X_b \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot (4)$$

Can be defined.

Here, the minimum square method obtains

$$\sum e_i^2 = e'e = (Y - X_b)'(Y - X_b) \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot (5)$$

Which minimizes the sum of the squares of the residuals, and obtains the following regression equation [8].

$$Y = b_0 + b_1 X_1 + \cdots + b_k X_k \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot (1)$$

Table 1: Air Pollution Data and Weather Data

Ground, Ground Temperature	The average temperature of the ground, average minimum temperature, minimum temperature, 0.05m average underground temperature, 0.1m average underground temperature, 0.2m average underground temperature, 0.3m average underground temperature, 0.5m average underground temperature, 1.0m average underground temperature, 1.5m average underground temperature, 3.0m average underground temperature, 5.0m average underground temperature
Temperature	The average temperature, average high temperature, average low temperature, highest temperature, lowest temperature
Precipitation	Monthly total precipitation, daily maximum precipitation, hourly maximum precipitation, maximum precipitation per hour
Wind	Average wind speed, maximum wind speed, maximum instantaneous wind speed, maximum wind direction, maximum instantaneous wind speed direction
Humidity	Average relative humidity, minimum relative humidity

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Atmospheric pressure	Average local pressure, average water vapor pressure, average dew point temperature, average sea level pressure, highest sea level pressure, lowest sea level pressure, maximum water vapor pressure, minimum water vapor pressure
Solar ·Solar radiation	Sum of daylight hours, percentage of sunshine, sum of whole solar radiation
Quantity of evaporation	Small total increase, small maximum increase, large total increase, large scale increase
Snow	Average summer cloud amounts
Cloud	Average cloud amounts
Air Pollutant	Sulfur dioxide, ozone, carbon monoxide, nitrogen dioxide, fine dust (PM_{10})

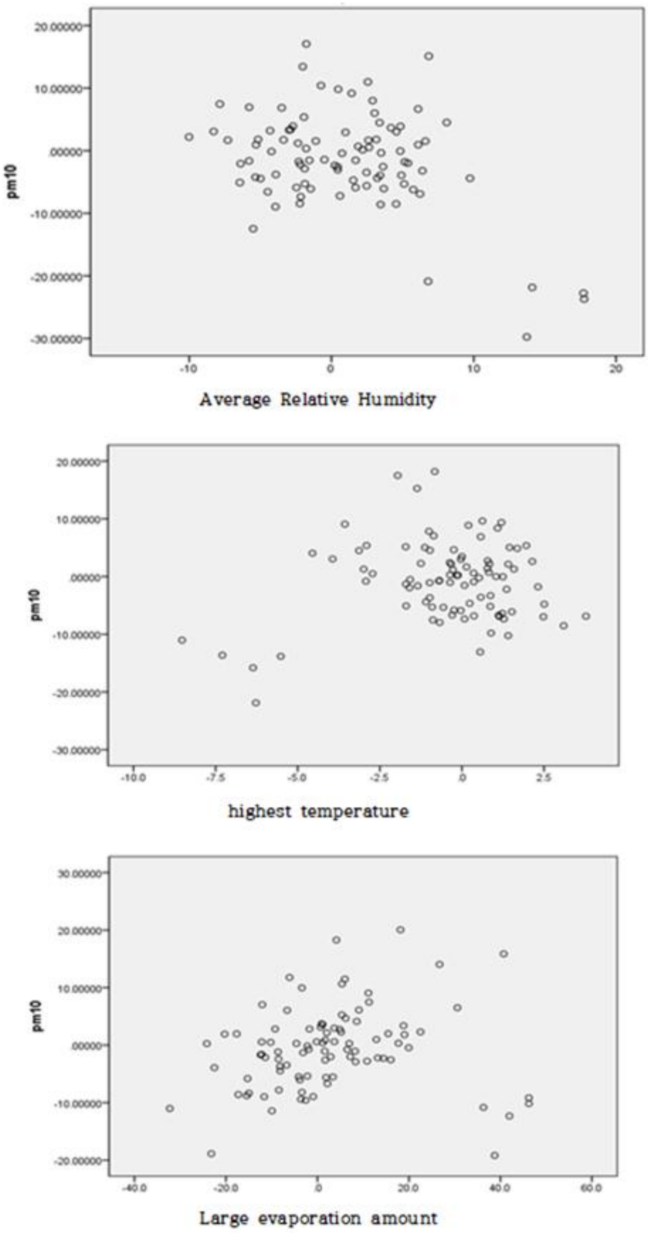

Figure 1. Relationship between fine dust (PM_{10}) and the independent variables

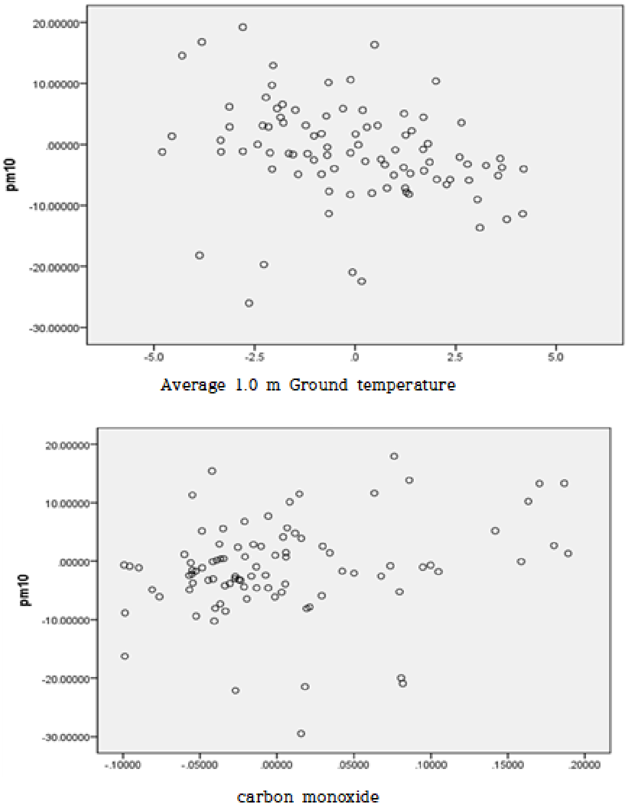

Figure 2. Relationship between fine dust (PM_{10}) and the independent variables

B. Study Variable

The key variables for fine dust forecast are the density data of air pollutant and weather data that reacts which represents the emission source and the chemical reaction. This study used sulfur dioxide (SO_2), ozone (O_3), carbon monoxide (CO), nitrogen dioxide (NO_2), and fine dust (PM_{10}) as air pollutants provided by the Environmental Management Corporation. The data of air pollutants provided by the Korea Environment Management Coporation is measured once every hour for 24 hours. Weather data was collected using the open weather data portal from the Meteorological Administration. The open weather data portal provides data in various formats, including time, day, month, quarter and year. The data provides ground, ground temperature, temperature, precipitation, wind, humidity, atmospheric pressure, solar and solar radiation, and evaporation volume as shown in Table 1.

C. Multiple Linear Regression Model

This study used stages of regression analysis method to select the optimal multiple linear regression model. Stages of regression analysis are there to select or eliminate to select the best regression model when there are many independent variables. The kinds of regression analysis consist of forward selection method, backward elimination method, and stepwise selection method.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Table 2: Multiple Regression Analysis for Fine Dust (PM_{10})

Model		Nonstandard coefficient		Standard coefficient	t	Level of significance
		B	Standard error	beta		
1	(Constant)	85.861	3.848		22.314	.000***
	1.0m Average underground temperature	-2.185	.179	-.803	-12.199	.000***
2	(Constant)	108.483	6.452		16.815	.000***
	1.0m Average underground temperature	-1.265	.274	-.465	-4.617	.000***
	Average relative humidity	-.595	.142	-.421	-4.180	.000***
3	(Constant)	93.643	8.474		11.050	.000***
	1.0m Average underground temperature	-1.371	.268	-.504	-5.116	.000***
	Average relative humidity	-.499	.143	-.353	-3.495	.001***
	CO	23.489	9.081	.156	2.587	.012*
4	(Constant)	51.130	12.212		4.187	.000***
	1.0m Average underground temperature	-1.961	.275	-.721	-7.129	.000***
	Average relative humidity	-.079	.159	-.056	-.494	.623
	CO	46.689	9.687	.311	4.820	.000***
	Large total increase rate	.176	.040	.308	4.456	.000***
5	(Constant)	69.420	13.185		5.265	.000***
	1.0m Average underground temperature	-1.540	.298	-.566	-5.163	.000***
	Average relative humidity	.024	.156	.017	.152	.880
	CO	39.579	9.549	.264	4.145	.000***
	Large total increase rate	.282	.052	.492	5.432	.000***
	The highest temperature	-1.310	.442	-.325	-2.964	.004**
1	R^2=.645, Adjusted R^2=.640, F=148.814, p=.000***					
2	R^2=.708, Adjusted R^2=.701, F=98.086, p=.000***					
3	R^2=.730, Adjusted R^2=.720, F=72.215, p=.000***					
4	R^2=.784, Adjusted R^2=.774, F=71.893, p=.000***					
5	R^2=.806, Adjusted R^2=.794, F=64.938, p=.000***					

Forward selection method starts with the independent variable that has the greatest correlation with the dependent variable, and selects the independent variable one by one. Backward elimination method is a method of removing independent variables one by one in a complete model with all independent variables [9]. If a variable is selected when using the forward selection method, it may become an unimportant variable among the selected variables. To compensate this kind of drawback, importance of selected variables is re-examined at each step of the forward selection method to eliminate non-critical variables. This is called stepwise selection method. In this study, stepwise

selection method was used to conduct multiple regression analysis [10].

According to the result of multiple regression analysis on fine dust (PM_{10}), the underground temperature of 1.0m was applied into the Model 1, and the explanatory power (R^2) was 0.645. In Model 2, the underground temperature of 1.0m and the average relative humidity were applied, with an explanatory power of .708. In Model 3, the underground temperature of 1.0m, the average relative humidity, and CO were applied. The explanatory power was shown at .730. In Model 4, the

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

underground temperature of 1.0m, the average relative humidity, CO, and large total increase were applied. The explanatory power was shown at .784. In Model 5, 1.0m underground temperature, average relative humidity, CO, large total increase, and highest temperature were applied. The explanatory power was shown at .806. Interestingly, the average relative humidity loses its influence in model 4.

In Module 5, with the highest explanatory power, large total evaporation came out to be .492, and CO came out to be .264. 1.0m average underground temperature was -.566, and the highest temperature was -.325. Therefore, when the density of large total evaporation and CO increased, fine dust (PM_{10}) also increased as shown in table 2. On the other hand, when the average underground temperature and the highest temperature decreased, fine dust (PM_{10}) also decreased. Figure 1 shows a graph that represents its relevance of composite.

Figure 1, 2 shows a graph of the relationship between fine dust and each independent variable. In figure 1, ideal data from standardized residuals, average relative humidity, highest temperature, and large total evaporation showed one pattern. Also, large total evaporation shows that the X-axis and Y-axis appear in oblique form, indicating more linearity than other factors.

III. Conclusion

Fine dusts are 1st group carcinogens that suppress the metabolism of plants and cause corrosion by depositing in buildings. When people inhale fine dusts in large amounts, it causes respiratory and inflammatory reactions, and increases cardiovascular diseases. The Ministry of Environment has been implementing find dust forecasts since 2005 to manage fine dust. Through the analysis process of fine dust forecast models, information on pollutant emissions and techniques on the process of spreading air pollutants can be understood.

In this study, a total of 50 variables, including surface, underground temperature, temperature, precipitation, wind, humidity, atmospheric pressure, solar and solar radiation, evaporation amount, snow, cloud, and air pollution were used to develop a multiple linear regression model. Stepwise regression analysis was also performed to select the optimal multiple linear regression model. The analysis result shows that the multiple linear regression model in Table 3 is derived. According to the model 5, occurrence of fine dust decreases when the average underground temperature of 1.0m and the highest temperature increases, and occurrence of find dust increases when large total evaporation increases. The underground temperature refers to the temperature of soil which can differ depending on the soil and moisture conditions. However, the underground temperature is close to the ground, it is highly affected by the weather and changes daily. There are no daily changes at 1m underground, and no annual changes at 6~7m. The amount of evaporation refers to the amount of water, lost by evaporation from the surface or from the surface of the unit area within a period of time. The amount of evaporation is expressed as the depth of water in mm like precipitation, and the time is indicated as 1 hour or 1 day. In Korea, a large scale evaporator (120cm) is used as the standard, but the amount of evaporation in the lake is about 70 percent of the measurement taken by this evaporator.

Table 3: Multiple Linear Regression Model for Fine Dust

Model 1 : PM_{10} = 85.861-2.185*A
Model 2 : PM_{10} = 108.483-1.265*A-.595*B
Model 3 : PM_{10} = 93.643-1.371*A-.499B+23.489*C
Model 4 : PM_{10} = 51.130-1.961*A+46.689*C+.176*D
Model 5 : PM_{10} = 69.420-1.540*A+39.579*C+.282*D-1.310*E

A: 1.0m average underground temperature
B: average relative humidity
C: CO
D: large total evaporation
E: highest temperature

ACKNOWLEDGMENT

This research was done through [New Business Model Job Creation Project] in Daejeon Techno park (2019).

REFERENCES

1. J.Y. Moon, Y.B. Lee, J.Y. Jeong, K.H. Jeong. “Distribution Characteristics of heavy Metals in the Ambient Air of Ulsan Area,” [Pittsburgh (PA)]: Busan University; 2001
2. http://www.dbpia.co.kr/Journal/ArticleDetail/NODE00637758?TotalCount=0&Seq=11&isIdentifyAuthor=1&Collection=0&isFullText=0&specificParam=0&SearchMethod=0&Page=1&PageSize=20
3. S.B. Lee, D.S. Jeong, E.K. Jo, H.A. Kim, E.Y. Hwang, C.H. Kang, *PM10, PM2.5 The Formation and Pollution Characteristics of Fine Dust: Year 2008 Measured on high mountain in Jeju Island*. Analytical Science Society Vol. 24 No. 4 in 2011:85-94.
4. http://www.dbpia.co.kr/Journal/ArticleDetail/NODE01670150?TotalCount=821&Seq=21&q=%5B%EB%AF%B8%EC%84%B8%EB%A8%BC%EC%A7%80%C2%A7coldb%C2%A72%C2%A751%C2%A73%5D&searchWord=%EC%A0%84%EC%B2%B4%3D%5E%24%EB%AF%B8%EC%84%B8%EB%A8%BC%EC%A7%80%5E*&Multimedia=0&isIdentifyAuthor=0&Collection=0&SearchAll=%EB%AF%B8%EC%84%B8%EB%A8%BC%EC%A7%80&isFullText=0&specificParam=0&SearchMethod=0&Sort=1&SortType=desc&Page=2&PageSize=20
5. W.S. Cho, *Statistical Characteristics of Meteorological Factors and Particulate Matters(PM10, PM2.5) in Seoul*. Seoul National University of Science and Technology : Environmental Engineering 2014. 2:120-121.
6. http://m.riss.kr/search/detail/DetailView.do?p_mat_type=be54d9b8bc7cdb09&control_no=9d1abf99a178370dffe0bdc3ef48d419
7. Y.S. Koo, H.Y. Yoon, H.Y. Kwon, S.H. Yoo, *Fine Dust Forecast System Development.* Journal of Korean Society for Atmospheric Environment, 26, 6, 2010.12, 666-682.
8. http://www.dbpia.co.kr/Journal/ArticleDetail/NODE01572664?TotalCount=4&Seq=1&q=%5B%EB%AF%B8%EC%84%B8%EB%A8%BC%EC%A7%80%20%EC%98%88%EB%B3%B4%EC%8B%9C%EC%8A%A4%ED%85%9C%20%EA%B0%9C%EB%B0%9C%C2%A7coldb%C2%A72%C2%A751%C2%A

73%5D&searchWord=%EC%A0%84%EC%B2%B4%3D%5E%24%EB%AF%B8%EC%84%B8%EB%A8%BC%EC%A7%80%20%EC%98%88%EB%B3%B4%EC%8B%9C%EC%8A%A4%ED%85%9C%20%EA%B0%9C%EB%B0%9C%5E*&Multimedia=0&isIdentifyAuthor=0&Collection=0&SearchAll=%EB%AF%B8%EC%84%B8%EB%A8%BC%EC%A7%80%20%EC%98%88%EB%B3%B4%EC%8B%9C%EC%8A%A4%ED%85%9C%20%EA%B0%9C%EB%B0%9C&isFullText=0&specificParam=0&SearchMethod=0&Sort=1&SortType=desc&Page=1&PageSize=20#

9. H.J. Chae, “A Study on Improving of PM10 Forecast in Seoul.” [Pittsburgh (PA)]: : GwangJu : Chosun University, 2010
10. http://www.riss.kr/search/detail/DetailView.do?p_mat_type=be54d9b8bc7cdb09&control_no=f49ea5b8b9c26636ffe0bdc3ef48d419
11. S.W. Kang, D.H. Lee, G.B. Jang, D.Y. Jin, *Environmental Big Data Analysis and Service Development*. Institute for Environmental Policy Evaluation 2017 Vol. No. 0, 147-154.
12. http://m.konetic.or.kr/include/EUN_download.asp?str=TBL_ENV_THEMA&str2=6838
13. H.Y. Yoon, “A Study on the Development of the Integrated Forecasting Model for Fine Dust Data.” [Pittsburgh (PA)]: Anyang University Graduate School, 2017.2
14. http://www.riss.kr/search/detail/DetailView.do?p_mat_type=be54d9b8bc7cdb09&control_no=40f4f4e977eb3586ffe0bdc3ef48d419
15. H.W. Chung, T.W. Park, *A study on detached housing price decision factors : the case of Ilsandonggu Goyang-si*. Kwangwoon University Graduate School of Business : Real estate and urban planning department real estate and urban planning major 2016 Vol 2. No. 1, 101-110.
16. http://scholar.dkyobobook.co.kr/searchDetail.laf?barcode=4010024925911
17. J.S. Choi, “Analysis on the Relations hip between Factors of Urban Creativity and Urban Regeneration.” [Pittsburgh (PA)]: Yonsei University Graduate School : Department of Public Administration 2016. 8
18. https://academic.naver.com/article.naver?doc_id=189098441
19. Korea Power Exchange. “A Study on the Improvement of Construction Intention Evaluation Criteria for Construction Performance” [Electronic Data]. Seoul : Korea Power Exchange, 2014.
20. http://www.dbpia.co.kr/SKnowledge/ArticleDetail/NODE06347826

AUTHORS PROFILE

Tae Hyung Kim Student, Department of Fire and Prevention, Daejeon university. Korea Water Environment Co., Ltd. CEO. Big Data Analysis, Statistical analysis, disaster planning and management.

Jae Ou Lee Professor, Department of Fire and Disaster Prevention, Daejeon University, Ph.d. Fire Safety Engineering, Fire protection PE(KR,US), Disaster prevention, Fire protection, Piping seismic analysis