ICCEPM 2022
The 9th International Conference on Construction Engineering and Project Management
Jun. 20-23, 2022, Las Vegas, NV, USA

Analysis of Relationship between Construction Accidents and Particulate Matter using Big Data

Minsu Lee[1], Jaewook Jeong[2*], Jaemin Jeong[3], Jaehyun Lee[4]

[1] *Department of Safety Engineering, Seoul National University of Science and Technology, 232, Gongneung-ro, Nowon-gu, Seoul, Republic of Korea,* E-mail address: alstn9881@seoultech.ac.kr
[2] *Department of Safety Engineering, Seoul National University of Science and Technology, 232, Gongneung-ro, Nowon-gu, Seoul, Republic of Korea,* E-mail address: jaewook.jeong@seoultech.ac.kr
[3] *Department of Safety Engineering, Seoul National University of Science and Technology, 232, Gongneung-ro, Nowon-gu, Seoul, Republic of Korea,* E-mail address: ss96011@seoultech.ac.kr
[4] *Department of Safety Engineering, Seoul National University of Science and Technology, 232, Gongneung-ro, Nowon-gu, Seoul, Republic of Korea,* E-mail address: archi0528@seoultech.ac.kr

Abstract: Because construction work is conducted outdoors, construction workers are affected by harmful environmental factor. Especially, Particulate Matter (PM_{10}) is one of the harmful environmental factors with a diameter of 10$\mu g/m^3$ or less. When PM_{10} is inhaled by human, it can cause fatal impact on the human. Contrary to the various analyses of health impact on PM_{10}, the research on the relationship between construction accidents and PM_{10} are few. Therefore, this study aims to conduct the relative frequency analysis which find out the correlation between construction accidents and PM_{10}, and the modified PM_{10} grade is suggested to expect accidents probability caused by PM_{10} in the construction industry. This study is conducted by four steps. i) Establishment of the database; ii) Classification of data; iii) Analysis of the Relative Frequency of accidents in the construction industry by PM_{10} concentration; iv) Modified PM_{10} groups to classify the impact of PM_{10} on accident. In terms of frequency analysis, the most accidents were occurred in the average concentration of PM_{10} (32$\mu g/m^3$). However, we found that the relative frequency of accident was increased as the concentration of PM_{10} increased. This means the higher PM_{10} concentration can cause more accidents during construction. In addition, PM_{10} concentration was divided as 6 groups by the WHO, but the modified PM_{10} grade by the relative frequency on accident was suggested as 3 groups.

Key words: PM_{10}, Relative Frequency, Modified PM_{10} groups, K-means clustering, construction accident

1. INTRODUCTION

Contrary to the manufacturing industry, most works in construction have to be conducted outdoors having various environmental factors. These environmental factors can affect the construction worker in various aspects such as health, productivity and constructability. The environmental issue about air pollutants have been highlighted in most countries, and it leads to a number of regulations and policies to control the harmful environmental factors. Especially,

Particulate Matter (PM_{10}) is known as the representative factor which have significant impact on the human health. PM_{10} under $10\mu g/m^3$ particulate matter is classified as Group 1 carcinogen by International Agency for Research on Cancer under World Health Organization (WHO) [1].

There are various previous studies related to PM_{10}, some of them focused on the health impact to the human body [2,3]. And there are some studies which provided that PM_{10} can cause productivity loss and increase stress of workers [4,5]. The others suggested the health management methods for construction workers considering PM_{10} concentration such as wearing personal protective equipment [6,7].

However, it was found that the previous studies analyzing the effect of PM_{10} on accidents of construction workers were insufficient. And, it can be judged that PM_{10} concentration can affect the accident occurrence if PM_{10} concentration have significant impact on the worker's productivity. Thus, it is reasonable to analyze the relationship between PM_{10} concentration and the accident occurrence in construction.

In this regard, this study aims to analyze the relationship between PM_{10} concentration and construction accidents in terms of both the frequency and the probability aspects. And the modified PM_{10} grade system is suggested to identify the impact level of construction accident by PM_{10}, contrary to the conventional PM_{10} concentration established for measuring health impact by WHO.

2. MATERIALS AND METHODS

This research is conducted by four steps. 1) Establishment of the database, 2) Classification of data, 3) Analysis of the Relative Frequency of accidents in the construction industry by PM_{10} concentration, 4) Development of modified PM_{10} grade to classify the impact of PM_{10} on accidents

2.1. Establishment of the database

Accident cases in construction were collected based on national occupational accident compensation database operated by Korea Occupational Safety & Health Agency, and the cases by worker's illness were excluded to analyze on-site accident only. Overall 214,538 accident cases including 6,736 fatalities and 207,802 injuries from 2007 to 2019 were collected excluding missing data cases [8]. PM_{10} concentration including its location and date were collected from Korea meteorological administration (KMA).

2.2. Classification of data

Table 1 shows the classification of PM_{10} concentration by WHO to evaluate the health impact of PM_{10} [9]. PM_{10} concentration among the collected data set excluding missing data on 13-years have a range from 1 $\mu g/m^3$ to 123$\mu g/m^3$. Five-groups by WHO standard from "Good" to "Very unhealthy" were defined in this study by the collected data set. And this study used $1\mu g/m^3$ as unit to analyze PM_{10} concentration and relationship with accident cases.

Table 1. Groups of PM_{10} concentration by WHO

Concentration	Classification
Under 30$\mu g/m^3$	Good
30 to 49$\mu g/m^3$	Moderate
50 to 89$\mu g/m^3$	Unhealthy for sensitive people
90 to 119$\mu g/m^3$	Unhealthy
120 to 154$\mu g/m^3$	Very unhealthy
Over 155$\mu g/m^3$	Hazardous

2.3. Analysis of the Relative Frequency of accidents in the construction industry by PM_{10} concentration

This study aims to analyze the relative relationship among PM_{10} concentration and construction accidents. Thus, it is hard to find out whether PM_{10} concentration has a significant relationship with accident using the conventional frequency analysis. The authors have insisted that the accident analysis based on the probabilistic approach, not the conventional frequency analysis, should be conducted to evaluate the risk [10]. Thus, the relative frequency analysis is proposed in this study to identify the impact of PM_{10} concentration on construction accident occurrence quantificationally.

The proposed relative frequency analysis can be explained below with equations (1) to (3). This approach can evaluate the probability of accident occurrence depends on each PM_{10} concentration. Overall data set of 13-years about accident cases and related PM_{10} concentration was calculated using these equations.

$$Period\ fraction = \frac{Number\ of\ days\ by\ PM_{10}\ concentration}{Total\ number\ of\ days\ in\ 13\ years} \quad (1)$$

$$Accident\ fraction = \frac{Number\ of\ accidents\ by\ PM_{10}\ concentration}{Total\ number\ of\ accidents} \quad (2)$$

$$Relative\ Frequency = \frac{Accident\ fraction}{Period\ fraction} \quad (3)$$

Period fraction means the ratio of the period of a certain PM_{10} concentration among the total number of days for 13-years. Accident fraction means the ratio of the accident upon a certain PM_{10} concentration per total accidents for 13-years. Relative frequency means the ratio of the accident fraction and period fraction.

If the relative frequency is calculated as "1" using the above equations, it means that the probability of accidents is identical with the average level of overall duration. And, if the relative frequency is lower than "1", it means that the probability of accident is lower than the average level. On the other hand, the relative frequency is higher than "1", the probability of accident on a certain PM_{10} concentration is higher than average level.

2.4. Modified PM_{10} groups to classify the impact of PM_{10} on accidents

As mentioned above, the current group about PM_{10} concentration defined by WHO was classified based on the health impact of PM_{10}. This study aims to analyze the impact level of PM_{10} concentration on the accident occurrence in the construction industry, thus it is required to provide the modified group of PM_{10} concentration considering the relationship of PM_{10} and accident.

First, the hierarchical clustering method was conducted using the collected data set of PM_{10} concentration from 1μg/m^3 to 123μg/m^3. And then the elbow point was selected to define the optimal number of groups from the result of the hierarchical clustering [11]. Second, K-means clustering was applied to calculate the boundary of each group based on the number of groups using the elbow point. K-means clustering was used to calculate the median value of each group through minimizing the sum of squared errors [12]. Third, analysis of variance (ANOVA) and post-hoc test were applied to identify whether the developed groups are mutually independent of each other.

ANOVA and post-hoc test are a widely used methods to compare the average of groups which have 3 or more [13].

The modified group of PM_{10} concentration based on the accident occurrence can be developed using the above procedure. This modified group of PM_{10} concentration can show the range of PM_{10} concentration which affect the accident impact, not the range of health impact.

3. RESULTS AND DISCUSSION

3.1. Results of the relative frequency analysis by PM_{10} concentration

According to the research procedure mentioned in Chapter 2, accident cases and PM_{10} concentration data for 13-years were analyzed. Figure 1 shows both the conventional frequency and the proposed relative frequency of accident occurrence upon PM_{10} concentration for 13-years. The grey colored line represents the amount of conventional frequency of accident, and the multi-colored line represents the relative frequency of accident occurrence based on PM_{10} concentration. The colors of the relative frequency line about "green" as "good", "blue" as "Moderate", "yellow" as "Unhealthy for sensitive people", "orange" as "Unhealthy", and "red" as "Very unhealthy" represents the PM_{10} groups by WHO standard, respectively. And the value "1" on left axis means the baseline of relative frequency explained in chapter 2.3.

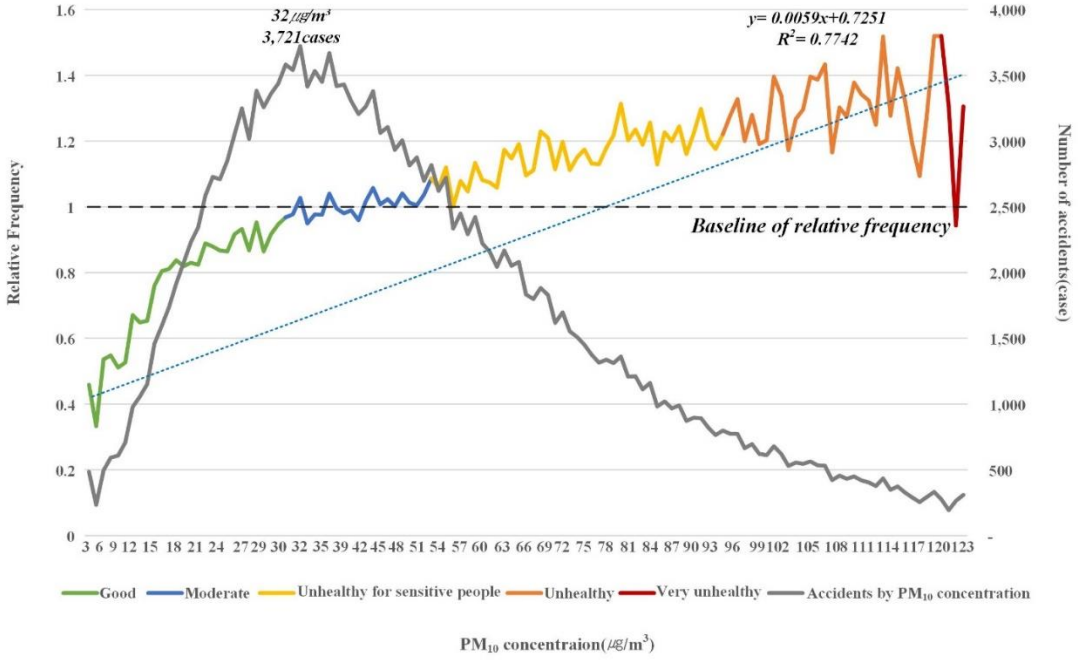

Figure 1. Accidents by PM_{10} concentration

As shown in Figure 1, the highest number of accidents showed at 32$\mu g/m^3$ (3,721 cases) of PM_{10} concentration on the conventional frequency graph. Because of the average PM_{10} concentration in South Korea was observed as 37.5$\mu g/m^3$ by KMA, it could be interpreted that most of the accidents had been occurred at the average condition based on the results of conventional frequency analysis [14].

However, if the period around the average PM_{10} concentration have a large portion for overall period, it is hard to judge that the largest number of accidents at the average PM_{10} concentration is the largest probability of accident occurrence. Thus, it might lead to a misunderstanding that PM_{10} concentration has no impact to the accident occurrence.

The developed relative frequency was provided to solve this problem in this study. Contrary to the result of the conventional frequency analysis, the relative frequency analysis showed that when the PM_{10} concentration is increased, the accident occurrence ratio is increased continuously. And its r^2 value based on linear regression of the relative frequency graph was calculated 0.7742. Additionally, the highest relative frequency of accidents on PM_{10} concentration was found at 123$\mu g/m^3$ (1.61RF), and this means the probability of accident occurrence at 123$\mu g/m^3$ PM_{10} concentration is 1.61 times higher than the average level during same duration.

3.2. Result of modified PM_{10} concentration groups development

3.2.1. Hierarchy clustering for selecting the optimal number of groups

As mentioned in Chapter 2.4, this study developed the modified PM_{10} groups to classify the PM_{10} concentration focused on the impact of accident occurrence in construction, not the health impact.

As shown in Figure 2, the elbow point was calculated as 3-groups through the hierarchical clustering for the collected data set of PM_{10} (1$\mu g/m^3$ to 123$\mu g/m^3$) and accident cases.

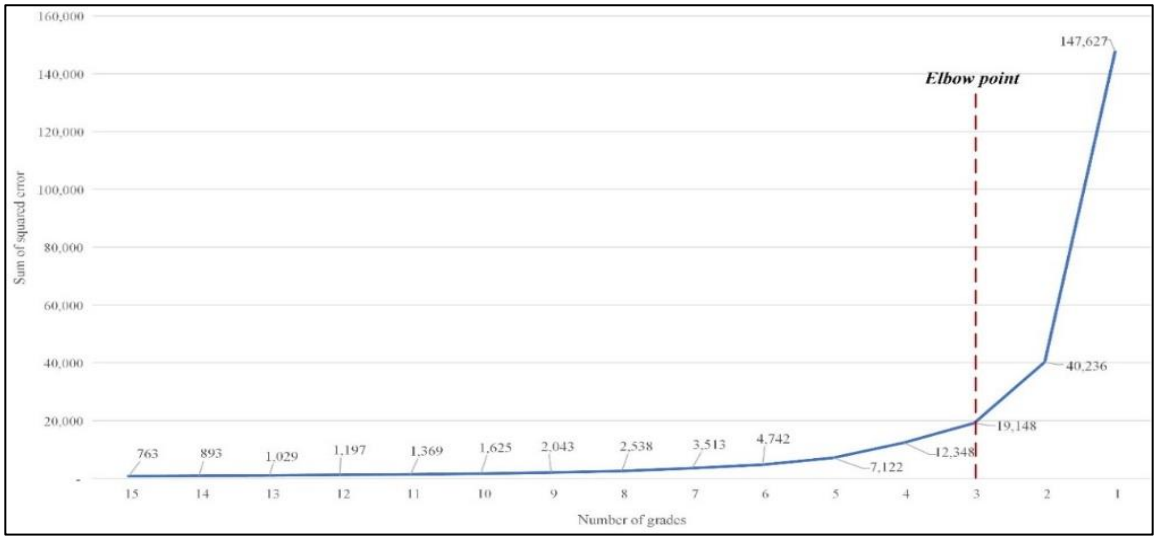

Figure 2. Hierarchical cluster analysis and elbow point results

3.2.2. K-means clustering for defining the range of each group

K-means clustering was conducted to define the range for the selected 3-groups calculated by the hierarchical clustering and the elbow point calculation in the previous chapter. Table 2 shows the results of K-means clustering for 3-groups based on PM_{10} concentration. Contrary to the WHO standard, the developed groups of PM_{10} concentration about the accident impact were divided by 3-groups, and Group 1,2,3 had the relative frequency of 0.31 to 1.02 under 41$\mu g/m^3$, 1.06 to 1.23 among 42 to 82$\mu g/m^3$, and 1.24 to 1.61 over 83$\mu g/m^3$ PM_{10}, respectively. It showed that Group 1 have lower probability of accident occurrence during a certain period, Group 2 has similar probability with average level, and Group 3 has a higher probability than other groups.

Table 2. Modified PM_{10} Groups of concentration

Classification	Concentration ($\mu g/m^3$)	Relative frequency
Group 1	Under 41	0.31 to 1.02

Group 2	42 to 82	1.06 to 1.23
Group 3	Over 83	1.24 to 1.61

3.2.3. Validation of the modified PM_{10} concentration group

ANOVA and post-hoc test were conducted to check whether each groups has significant differences among each other. Table 3 and 4 show the results of ANOVA and post-hoc test, respectively. It was identified that there were significant differences among groups through ANOVA as shown in Table 3. Additionally, the independence of each group was checked through post-hoc test, as shown in Table 4, each group has a significant difference for all combinations, and it showed that the modified groups of PM_{10} about accident occurrence are effective to judge the probability of accident upon PM_{10} concentration in construction.

Table 3. Modified PM_{10} group ANOVA results

Source	Levene's test	Levene's P	Welch's test	*P*-value
Grade	8.455	.000	84.265	.000

Table 4. Result of Post-hoc test of Modified PM_{10} group

Groups of PM_{10} (Games howell)	Group 1	Group 2	Group 3
Group 1	-	.000	.000
Group 2		-	.000
Group 3			-

4. CONCLUSION

Construction workers are generally exposed to the outdoor environments, and PM_{10} is a well-known hazardous air pollutant which can affect the worker's health and productivity throughout various previous studies. And there were a number of studies that focused on the health impact of PM_{10}, it was, however, hard to find researches to investigate the relationship of PM_{10} concentration and accident occurrence.

In this study, the relationship between PM_{10} concentration and accident occurrence was investigated through the relative frequency approach what we developed. And the modified PM_{10} concentration group in terms of accident impact, not the health impact, was also developed.

This study was conducted by four steps, i) Establishment of the database, ii) Classification of data, iii) Analysis of the Relative Frequency of accidents in the construction industry by PM_{10} concentration, iv) Modified PM_{10} groups to classify the impact of PM_{10} on accident.

The major results in this study are as follows. First, over 214,538 accident cases in construction and PM_{10} observations throughout 13-years in South Korea were analyzed. Second, contrary to the conventional frequency analysis, the relative frequency analysis showed that when the PM_{10} concentration is increased, the accident occurrence ratio is increased continuously. And its r^2 value based on linear regression of the relative frequency graph was calculated as 0.7742. Additionally, the highest relative frequency of accidents on PM_{10} concentration was found at 123μg/m^3 (1.61RF),

and this means the probability of accident occurrence at 123$\mu g/m^3$ PM_{10} concentration is 1.61 times higher than the average level during same duration. Third, the modified PM_{10} concentration group to judge the probability of accident occurrence upon PM_{10} concentration was developed and validated. And Group 1,2, and 3 had the relative frequency of 0.31 to 1.02 under 41$\mu g/m^3$, 1.06 to 1.23 from 42 to 82$\mu g/m^3$, and 1.24 to 1.61 over 83$\mu g/m^3$ PM_{10}, respectively. Contrary to the existing PM_{10} concentration groups by WHO standard, this modified group is focused on the accident impact of PM_{10} concentration, not the health impact.

The contributions of this study are as follows. First, this study showed that the higher PM_{10} concentration has a higher probability of accident in construction. It means that PM_{10} concentration can be utilized as an important factor, not only the existing health management, but as safety management to prevent accident. If the accident impact as PM_{10} concentration can be calculated, then accident loss and risk value can be expected [15]. Second, the modified PM_{10} concentration groups considering the accident impact were developed. Contrary to the conventional 6-groups by WHO were developed for worker's health management, the modified 3-groups can be used to manage and control the safety plan for accident prevention upon PM_{10} concentration of a certain period.

The limitation of this study is that the accident data was not subdivided by the severity level such as the fatalities, the unrecoverable injuries, and the recoverable injuries. And the various types of construction such as facility, work, and task were not considered.

ACKNOWLEGEMENTS

This work is supported by the Korea Agency for Infrastructure Technology Advancement(KAIA) grant funded by the Ministry of Land, Infrastructure and Transport (Grant 22CTAP-C163805).

REFERENCES

[1] International Agency for Research on Cancer, "https://iarc.who.int/adsdasd", Retrieved on 12.20.2021.

[2] M. Bentayeb, M. Simoni, N. Baiz, D. Norback, S. Baldacci, S. Maio, G. Viegi, I. Annesi-Maesano, Geriatric Study in Europe on Health Effects of Air Quality in Nursing Homes (GERIE) Group. "Adverse respiratory effects of outdoor air pollution in the elderly", International Journal of Tuberculosis and Lung Disease, vol. 16, pp. 1149-1161, 2012

[3] G. Polichetti, S. cocco, A. pinali, V. Trimarco, A. Nunziata. "Effects of particulate matter (PM_{10}, PM2.5 and PM1) on the cardiovascular system", Toxicology, vol. 261, pp. 1-8, 2009.

[4] J. He, H. Liu, A. Salvo. "Severe air pollution and labor productivity: Evidence from industrial towns in China", American Economics, vol. 11, no. 1, pp. 173-201, 2019.

[5] C. Shuai, Z. Dandan. "Impact of air pollution on labor productivity: Evidence from prison factory data", China Economic Quarterly International, vol. 1, no. 2 pp. 148-159, 2021.

[6] C. Daniel, J. H. Choi. "Data on different sized particulate matter concentration produced from a construction activity", Data in Brief, vol. 33, pp. 106467, 2020.

[7] F. W. Gary, G. David. "The impact of local fugitive PM_{10} from building works and road works on the assessment of the European Union Limit Value", Atmospheric Environment, vol. 38, pp. 4993-5002, 2004.

[8] Korean Statistical Information Service. “Available online: http://kosis.kr” Retrieved on 12.20.2021.

[9] Statistics Korea, “https://index.go.kr”, Retrieved on 12.20.2021.

[10] J. Jeong, J. Jeong, “Quantitative Risk Evaluation of Fatal Incidents in Construction Based on Frequency and Probability Analysis”, vol. 38, no. 2, pp. 04021089, 2022

[11] M. Lee, T. Hong, J. Jeong, K. Jeong. “Development of a rooftop solar photovoltaic rating system considering the technical and economic suitability criteria at the building level”, Energy, vol. 160, pp. 213-224, 2018

[12] J. Jeong, T. Hong, C. Ji, J. Kim, C. Koo, “Improvements of the operational rating system for existing residential buildings”, Applied energy, vol. 193, pp. 112-124, 2017.

[13] C. Reimann, P. Filzmoser, K. Hron, P. Kynčlová, R. G. Garrett. “A new method for correlation analysis of compositional (environmental) data – a worked example”, Science of the Total Environment. Vol. 607, pp. 965-971. 2017

[14] Korea Environment Corporation, “https://keco.or.kr”, Retrieved on 12.20.2021.

[15] J. Lee, J. Jeong, J. Soh, J. Jeong, “Development of Framework for Estimating Fatality-Related Losses in the Korean Construction Industry”, International journal of environmental research and public health, vol. 18, no. 16, pp. 8787, 2021