A Big Data Analysis of PM$_{2.5}$ and PM$_{10}$ from Low Cost Air Quality Sensors near Traffic Areas

Shida Chen$^1$, Kangping Cui$^{1*}$, Tai-Yi Yu$^{2**}$, How-Ran Chao$^{3,4,5***}$, Yi-Chyun Hsu$^6$, I-Cheng Lu$^7$, Rachelle D. Arcega$^3$, Ming-Hsien Tsai$^7$, Sheng-Lun Lin$^{8,9,10}$, Wan-Chun Chao$^{11}$, Chunneng Chen$^{11}$, Kwong-Leung J. Yu$^{12}$

$^1$ School of Resources and Environmental Engineering, Hefei University of Technology, Hefei 246011, China
$^2$ Department of Risk Management and Insurance, Ming Chuan University, Taipei 11103, Taiwan
$^3$ Department of Environmental Science and Engineering, College of Engineering, National Pingtung University of Science and Technology, Pingtung 91201, Taiwan
$^4$ Institute of Food Safety Management, College of Agriculture, National Pingtung University of Science and Technology, Pingtung 91201, Taiwan
$^5$ Emerging Compounds Research Center, General Research Service Center, National Pingtung University of Science and Technology, Pingtung 91201, Taiwan
$^6$ Department of Environmental Engineering, Kun Shan University, Tainan 71003, Taiwan
$^7$ Department of Child Care, College of Humanities and Social Sciences, National Pingtung University of Science and Technology, Pingtung 91201, Taiwan
$^8$ Department of Civil Engineering and Geomatics, Cheng Shiu University, Kaohsiung 83347, Taiwan
$^9$ Center for Environmental Toxin and Emerging-Contaminant Research, Cheng Shiu University, Kaohsiung 83347, Taiwan
$^{10}$ Super Micro Mass Research and Technology Center, Cheng Shiu University, Kaohsiung 83347, Taiwan
$^{11}$ JS Environmental Technology and Energy Saving Co. Ltd., Kaohsiung 80661, Taiwan
$^{12}$ Superintendent Office, Pingtung Christian Hospital, Pingtung 90059, Taiwan

ABSTRACT

Particulate matter (PM) pollution (including PM$_{2.5}$ and PM$_{10}$), which is reportedly caused primarily by industrial and vehicular emissions, has become a major global health concern. In this study, we aimed to reveal spatiotemporal characteristics and diurnal patterns of PM$_{2.5}$ and PM$_{10}$ data obtained from 50 air quality sensors situated in public bike sites in Kaohsiung City on June and November 2018 using principal component analysis (PCA). Results showed that PM concentrations in the study were above the standard World Health Organization criteria and were found to be associated, although complicated, with relative humidity. Specifically, the relationship between PM concentrations and relative humidity suggest a clear association at lower PM concentrations. Temporal analysis revealed that PM$_{2.5}$ and PM$_{10}$ occurred at higher concentrations in winter than in summer, which could be explained by the long-range transport of pollutants brought about by the northeast monsoon during the winter season. Both PM fractions displayed similar spatial distribution, wherein PM$_{2.5}$ and PM$_{10}$ were found to be concentrated in the heavily industrialized areas of the city, such as near petrochemical factories in Nanzih and Zuoying districts in north Kaohsiung and near the shipbuilding and steel manufacturing factories in Xiaogang district in south Kaohsiung. A pronounced diurnal variation was found for PM$_{2.5}$, which generally displayed higher peaks during the daytime than in the nighttime. Peaks generally occurred at 7:00–9:00 a.m., noontime, and 5:00–7:00 p.m., while minima generally appeared at nighttime. The diurnal pattern of PM was greatly influenced by a greater number of industrial and human transportation activities during the day than at night. Overall, a

* Corresponding author.
E-mail address: cuikangping@163.com

** Corresponding author.
Tel.: 886-2-2882-4565 ext. 2619; Fax: 886-2-2880-9755
E-mail address: yti@mail.mcu.edu.tw

*** Corresponding author.
Tel.: 886-8-770-3202 ext. 7517; Fax: 886-8-774-0256
E-mail address: hrchao@mail.npust.edu.tw
number of factors such as relative humidity and type of season, transboundary pollution from neighboring countries, and human activities, such as industrial operations and vehicle use, affects the PM quality in Kaohsiung City, Taiwan.

**Keywords:** Particulate matter; Public bike sites; Principal component analysis; Internet of things; Low-cost air sensor.

## INTRODUCTION

Air pollution has gained a global interest in various parts of the world, particularly in developing countries where there is a high demand for industrial development, urbanization, energy consumption, and agricultural burning (Fotourehchi, 2016). Of all the air pollutants, attention has been paid to particulate matter (PM), which is a complex mixture of extremely small particles and liquid droplets consisting of organic chemicals, acids, and soil and dust particles (U.S. EPA, 2018). PM includes inhalable PM (PM10, aerodynamic diameter \( \leq 10 \) \( \mu \)m) and fine PM (PM2.5, aerodynamic diameter \( \leq 2.5 \) \( \mu \)m) and their environmental levels as well as epidemiological and toxicological effects have been well documented. Most PM occurs naturally in the environment, but increasing anthropogenic interferences in the environment have significantly increased the PM burden (Zhang et al., 2015; Cha et al., 2019). For example, vehicular emissions and human activities such as biomass burning, industrial processing, agricultural operations, and construction activities, to name a few, all contribute to the anthropogenic sources of PM (Wang et al., 2013; Widiana et al., 2017; Guan et al., 2019; Hao et al., 2019; Shahid et al., 2019). Rapid industrialization plays a major role in PM emissions. For example, developing countries such as India and China have displayed large amounts of increasing PM emissions in the last few years, which may be attributed to the rapid industrialization of both countries (Guo et al., 2017; Ye et al., 2018; Li et al., 2019b). The PM2.5 and PM10 values from the aforementioned studies have both exceeded the Air Quality Guideline from the World Health Organization, which stipulates that PM2.5 and PM10 levels should not exceed 10 \( \mu \)g m\(^{-3}\) annual mean (25 \( \mu \)g m\(^{-3}\) 24-hour mean) and 20 \( \mu \)g m\(^{-3}\) annual mean (50 \( \mu \)g m\(^{-3}\) 24-hour mean), respectively (WHO, 2005). In other parts of Asia, such as in the urban areas of Iran and Bangladesh, PM emissions still exceeded the WHO guideline (Khanabadi et al., 2017; Begum and Hopke, 2019; Miri et al., 2019). Even though air pollution has reduced substantially over the past decades in European countries, air pollutant concentrations are still too high. For example, PM emissions in Poland (Badya et al., 2017) and Greece (Terzi et al., 2010) are well above the WHO guideline. In Taiwan, there are currently a huge amount of data concerning PM levels, with diesel vehicles and manufacturing plants identified as the main PM emission sources in various Taiwanese cities (Chen et al., 2017; Lee et al., 2018; Lee et al., 2019; Lu et al., 2019). In terms of seasonal variation, summers in Taiwan were found to have the best air quality while winters have the worst, with PM10 as the typical primary pollutant during the winter (Lee et al., 2019; Lu et al., 2019). Although Lee et al. (2018) found decreasing annual average PM2.5 concentrations from 2013 to 2017, the PM2.5 levels throughout Taiwan still exceeded the standard WHO criteria. The alarming global levels of PM have pushed local authorities to implement appropriate measures for the reduction of PM emissions to decrease risk to humans.

The health effects of PM to humans range from allergic rhinitis and asthma to the development of more serious chronic diseases and premature death. Epidemiological studies have shown PM2.5 and PM10 to be closely associated with type 2 diabetes, chronic kidney disease, Alzheimer’s disease, respiratory diseases, rate of hospitalization, and mortality (Bell et al., 2009; Kloog et al., 2013; Jung et al., 2015; Hwang et al., 2017a, b; Guo et al., 2018; Kim et al., 2018; Li et al., 2018; Li et al., 2019a). In addition, PM10 was found to harm the nervous system and possibly aid in the development of mild cognitive impairment in elderly people (Ranft et al., 2009). Evidence from epidemiological studies have led researchers to conduct supporting cell and animal studies to confirm effects of PM on human health. In a recent review of the literature, PM exposure was found to be toxic to various cell types and causes changes in normal cellular functions that may lead to pulmonary diseases, cardiovascular dysfunctions, and immune inflammatory responses (Cho et al., 2018). For example, PM2.5 induced cell death and reduced cell viability in rat cells (Cao et al., 2016). In vivo studies also demonstrated that PM exposure leads to inflammatory response in the respiratory system, effects on cardiovascular responses, weakening of the pulmonary function, progression of atherosclerosis and diabetes mellitus, and induced oxidative stress and allergic inflammation (Cho et al., 2018; Maciejczyk et al., 2018). Due to these observed health effects, research on environmental levels of PM is growing. Outdoor PM pollution, however, is vastly under-monitored compared to indoor PM despite the threats to the environment and human health. Regardless, the government is making efforts to establish air quality monitoring sensors in public areas where people could be exposed.

Traditional methods of PM measurement and analysis can be costly and time-consuming. The concept of Internet of Things (IoT) is to connect numerous objects (such as appliances, vehicles, and sensors) to the Internet, which allows for data and information sharing. IoT improves measurement and analysis of air pollutants by equipping environmental sensors with IoT capabilities. This reduces the need for human labor and cost, allows frequent sampling, increases the scale of sampling and monitoring, allows sophisticated testing on-site, and binds response efforts to detection systems. At present, there is an increasing number of studies utilizing IoT for environmental monitoring of pollutants. For example, IoT system communication models have been created to measure and store real-time...
established iDust, which is a system based on IoT for real-time monitoring in various parts of the world are also available. Marques et al. (2018) established iDust, which is a system based on IoT for real-time monitoring of ambient PM$_{10}$, PM$_{2.5}$, and PM$_{1.0}$. PM exposure data were collected for two months inside two laboratories of a Portuguese university. In addition, iDust was equipped with an alert manager that notifies the user when the air quality is poor. Kumar and Jasuja (2017) also designed a monitoring system based on IoT that measures outdoor PM$_{2.5}$, CO, CO$_2$, temperature, humidity and air pressure in Delhi, India. The system proved to be highly accurate after comparing the results with data provided by the local environment control authority. In United Kingdom, AQ IoT devices have been deployed and the results obtained were similar compared with the data from government-operated AQ stations (Johnston et al., 2019). In order to interpret and give value to the data generated from IoT devices, scientists utilize a wide variety of statistical analyses.

Principal component analysis (PCA) is widely used in environmental pollution cases, especially in air and water pollution. The main purposes of PCA are to simplify the number of variables (Wang et al., 2008; Mari et al., 2016; Chang et al., 2019), analyse intercorrelations between pollutants’ concentration and other physical parameters (Tian et al., 2018), confirm the causal effects between pollutants and sources (Singh et al., 2005; Yongming et al., 2006), and evaluate statistical effectiveness of the monitoring data (Dominick et al., 2012). For regional classification, the use of PCA has three main advantages: (1) it provides geographical classification with statistical significance and physical meanings; (2) it identifies the distribution and characteristics of pollutant concentrations in the sub-region; and (3) it determines common features of most stations and explains the bottleneck faced by single station phenomenon (Eder et al., 1987).

With the emergence of IoT, researchers have attempted to obtain air quality monitoring data from AQ sensors/monitors and use statistical analyses for data interpretation instead of using the traditional monitoring methods. Given that concerns on PM pollution is growing, this study aims to address PM$_{2.5}$ and PM$_{10}$ in public bike sites near traffic roads in Kaohsiung City, Taiwan using big data analysis. Specifically, the objectives are to (1) reveal temporal and spatial characteristics of outdoor PM$_{2.5}$ and PM$_{10}$ in public bike sites, (2) analyse the time-varying characteristics of PM$_{2.5}$ and PM$_{10}$ in public bike sites, (3) map the distribution of PM$_{2.5}$ and PM$_{10}$ and reveal the relationship between them and with the demography of the region, and (4) determine associations of PM$_{2.5}$ and PM$_{10}$ emissions with temperature and relative humidity.

**MATERIALS AND METHODS**

**Study Area and Data Collection**

The terrain of Kaohsiung City and locations of the monitoring stations in this study are shown in Fig. 1. The 6-in-1 air sensor (NPL-AirSensor), which is serviced for monitoring of temperature, humidity, PM$_{2.5}$, PM$_{10}$, carbon monoxide (CO), and total volatile organic compounds (TVOCs), from JS Environmental Technology and Energy Saving Co. Ltd (Kaohsiung, Taiwan) was established in a public bike site (CityBike system, Kaohsiung Public Bike). A total of 50 air sensors were set up in 50 sites that were selected from the 300 public bike sites in Kaohsiung City. All of the selected public bike rental sites are located nearby the heavily trafficked roads in the urban areas of Nanzih, Zuoying, Sanmin, Xinxing, Qianjin, Yancheng, Lingya, Qianzhen, and Xiaogang Districts in Kaohsiung City. The districts of Nanzih, Zuoying, and Xiaogang have nearby industrial areas as well. The air sensors were set up beside the rental-system automatic machine at a height of 2 m above the ground. Each air sensor was connected with wifi and internet system to transfer the data in the cloud system per minute. The data, which include temperature, humidity, PM$_{2.5}$, and PM$_{10}$, were gathered daily from June 6–9 and November in 2018. For the whole duration of sampling, 1.52% of loss data was found, which could be probably due to the unreasonable real-time levels of PM$_{2.5}$ and PM$_{10}$ from the air sensors. Totally, 9,792,000 data were obtained for further statistical analysis. In the first stage, the data in the cloud system were transferred to Microsoft Excel or Statistics Analysis system (SAS) before subjecting to big data analysis. Four items (temperature, humidity, PM$_{2.5}$, and PM$_{10}$) were listed on each site based on time series. Finally, 50 site data were combined for the statistical analysis.

**Principal Component Analysis**

This study applied the PCA method to perform spatial delineation, simplify the dimensions of variables, and figure out the possible sources of PM$_{2.5}$ and PM$_{10}$ pollution over specific regions with the massive monitoring data. The unrotated PCA is the best method to obtain the initial conditions and replace the original complex variables with linear combinations of major components. However, despite the fact that unrotated PCA can simplify the numbers of complicated variables, it may not be able to explain reasonably the physical phenomena of air quality problems, including PM$_{2.5}$ and PM$_{10}$. The unrotated PCA maximizes sums of root mean square (RMS) of correlation coefficient, whereas, Varimax method (Kaiser, 1958) maximizes the variance of squared correlation coefficients in rotated principal components (RPCs). The unrotated PCA maximized the first-order moment of correlation coefficients and formed poor distribution between unrotated principal components and measurable variables, which resulted to no significant differences in correlation coefficients of unrotated principal components. The unrotated principal components was then difficult to identify. The rotational method maximized the second-order moment of correlation coefficients and formed wide distribution between RPCs and measurable variables. Therefore, among any of the RPCs, few measurable variables have high factor loadings with most of them being zero, which made it easy to explain correlation between measurable...
variables and the RPCs. First, the time series values of air quality values were normalized (Eq. (1)).

$$Z_{ik} = \frac{C_{ik} - \mu_i}{S_i}$$

where $Z_{ik}$ represents the Z score of kth time series value from station i; $C_{ik}$ stands for the kth time series value from station i; $\mu_i$ stands for the mean value at station i; and $S_i$ represents the standard deviation at station i. The correlation between the RPCs and Z scores could be depicted as in Eq. (2).

$$Z_{ik} = \sum_{j=1}^{n} L_{ij} P_{jk}$$

where $L_{ij}$ represents the factor loadings of the jth RPCs from station i; and $P_{jk}$ represents the kth observation value of the jth RPC.

All statistical analyses were performed using SAS JMP (NC, USA). The figures were drawn using the SigmaPlot 14.0 software (Systat Software, Inc, CA, USA).

**RESULTS AND DISCUSSION**

**Data Presentation**

Table 1 shows the main rotated principal components (RPCs) (with eigenvalues > 1.00) after subjecting the 50 monitoring air sensor stations to PCA test. The resulting eigenvalues and explained variances for the distinct RPCs for PM$_{2.5}$ and PM$_{10}$ were quite consistent, suggesting that the pollution characteristics of PM$_{2.5}$ and PM$_{10}$ could have the same spatial features. PCA results show that the 5 rotated

<table>
<thead>
<tr>
<th>RPCs</th>
<th>PM$_{10}$</th>
<th>PM$_{2.5}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\lambda$</td>
<td>$\sigma(%)$</td>
</tr>
<tr>
<td>1</td>
<td>18.83</td>
<td>37.67</td>
</tr>
<tr>
<td>2</td>
<td>15.89</td>
<td>31.77</td>
</tr>
<tr>
<td>3</td>
<td>4.25</td>
<td>8.50</td>
</tr>
<tr>
<td>4</td>
<td>3.88</td>
<td>7.75</td>
</tr>
<tr>
<td>5</td>
<td>1.39</td>
<td>2.78</td>
</tr>
<tr>
<td>Sum  (%)</td>
<td>88.48</td>
<td>88.64</td>
</tr>
</tbody>
</table>

$\lambda$: eigenvalues, $\sigma$: explained variances (%).
Principal Component Analysis

In order to study the variations between summer and winter PM concentrations, the sampling periods chosen in the present study were from June 6–9, 2018 and in November 2018. June and November sampling data are herein referred to as summer and winter data, respectively. The association between relative humidity and PM$_{2.5}$ and PM$_{10}$ concentrations are shown in Fig. 3. The summer data (Fig. 3(b)) showed a very similar pattern between PM$_{2.5}$ and PM$_{10}$, wherein a clear correlation between PM$_{2.5}$ and PM$_{10}$ concentration and relative humidity could be observed at relative humidity lower than 90%. However, for relative humidity greater than 90%, discrete patterns could be observed between PM$_{2.5}$ and PM$_{10}$ concentration and relative humidity, suggesting poor correlation between PM concentration and relative humidity.

Table 2. Quartile deviation values of PM$_{2.5}$ and PM$_{10}$ concentrations at respective spatial delineations (µg m$^{-3}$).

<table>
<thead>
<tr>
<th>RPCs</th>
<th>PM$_{10}$</th>
<th>PM$_{2.5}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Q1&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Q2&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>1</td>
<td>47</td>
<td>68</td>
</tr>
<tr>
<td>2</td>
<td>39</td>
<td>59</td>
</tr>
<tr>
<td>3</td>
<td>40</td>
<td>58</td>
</tr>
<tr>
<td>4</td>
<td>42</td>
<td>58</td>
</tr>
<tr>
<td>5</td>
<td>45</td>
<td>70</td>
</tr>
</tbody>
</table>

<sup>a</sup> Q1: 25<sup>th</sup> percentile.
<sup>b</sup> Q2: 50<sup>th</sup> percentile.
<sup>c</sup> Q3: 75<sup>th</sup> percentile.
humidity at higher relative humidity values. On the contrary, winter data (Fig. 3(a)) showed different patterns between PM$_{2.5}$ and PM$_{10}$. For example, PM$_{2.5}$ concentration showed good correlation with relative humidity at relative humidity lower than 90%, while it showed discrete pattern at relative humidity greater than 90%. For PM$_{10}$, on the other hand, a good correlation between PM$_{10}$ concentration and relative humidity could be observed at relative humidity less than 65%. However, at relative humidity greater than 65%, associations between concentration and relative humidity were not clear. It was also observed that a good association between PM and relative humidity can be seen at lower PM concentrations, suggesting that PM concentration is a major parameter on the association between PM and relative humidity. Overall, relative humidity affects the pollutant pattern of PM$_{2.5}$ and PM$_{10}$ during summer and winter seasons, with a more distinct increase in pollutant concentration observed during the winter than in summer. Currently, there are large gaps in understanding correlations between PM pollution and relative humidity due to their non-linear relationship (Wang, 2015). When high relative humidity occurs, concentration levels of large particles suspended in the air are easily reduced. In a study from the United States, the high PM$_{10}$ concentrations were likely to occur at relative humidity of 40–60% (Chu, 1997). This study revealed that the high values of PM$_{2.5}$ and PM$_{10}$ concentrated at relative humidity of 75–85% in summer and concentrated at relative humidity of 85–93% in winter. However, the current results are only preliminary results. More data are still needed to clarify the correlation between concentration levels of aerosols and relative humidity. In fact, big data analysis can provide more detailed results compared in the past. Therefore, the comparisons made between the individual 50 air sensors’ PM concentrations and relative humidity data in the present study cannot conclusively determine PM’s relationship with relative humidity. Nevertheless, present studies suggest positive correlations between PM$_{2.5}$ and relative humidity, which indicate that an increase in relative humidity could lead to higher PM concentrations (Xu et al., 2015; Zhu et al., 2016). In addition, higher concentrations of PM$_{2.5}$ and PM$_{10}$ were observed during the winter as compared to the summer. PM concentrations in the summer were all below 50 μg m$^{-3}$ while PM concentrations ranged from approximately 20 to 200 μg m$^{-3}$ in the winter. Similar results were observed in a Chinese study wherein relative humidity had a more intense accumulation effect on PM$_{2.5}$ and PM$_{10}$ concentrations in winter compared to summer (Lou et al., 2017). Previous studies have demonstrated that the local air quality in Taiwan deteriorates during the winter season (Chen and Lin, 2015; Hsu et al., 2016). For example, Hsu et al. (2016) reported that there were higher concentrations of PM$_{2.5}$ and PM$_{10}$ during the winter as compared to spring and summer seasons. This could be due to the northeast monsoon during the winter season carrying air pollutants and dust from mainland China to Taiwan, thereby resulting in an increased pollutant concentration.
during this period (Chi et al., 2014; Chuang et al., 2016). Furthermore, the study area in the present study, Kaohsiung City, is situated in a unique location that accepts long-range transported air pollutants to Taiwan (Tsai et al., 2012), resulting to higher measured PM concentrations in the present study. Aside from the long-range transport of air pollutants to Taiwan mainly due to the winter monsoon, coal combustion during the winter season could also be one contributor to higher PM concentrations during winter compared to summer (Zheng et al., 2005). However, most of Taiwanese households no longer uses coal as source of heating during winter. In addition, there is also an increased use of ventilation during summer due to the hot weather, which explains the lower pollutant levels during this season.

In order to understand the spatial characteristics of the RPCs, this study provided factor loading contours over different RPCs. Figs. 4 and 5 illustrate the factor loadings of the four RPCs for PM$_{2.5}$ and PM$_{10}$, respectively. PC5 was not included due to low explained variance (< 5%) for both PM$_{2.5}$ and PM$_{10}$. It is important to note that Kaohsiung City is a heavily industrialized city, therefore, has pollution sources with large PM emissions, such as the Dashe and Renwu Industrial Zones in the north of Kaohsiung City and Xiaogang, Linyuan, and Fengshan Industrial Zones in the south. Figs. 4 and 5 show that the spatial distributions of PM$_{2.5}$ and PM$_{10}$ emissions in the study area were similar. According to the factor loading contours of PM$_{2.5}$ (Fig. 4), high factor loading values of PC1 were concentrated in the north district of Kaohsiung City, high factor loading values of PC2 were concentrated in the south district of Kaohsiung City, high factor loading values of PC3 were scattered in the southern district of Kaohsiung City, near the Kaohsiung harbor (factor loading values of the four stations were over 0.5), and lastly, high factor loadings of PC4 were concentrated near the Xiaogang Industrial Zone in the southern district of Kaohsiung City (factor loading of three stations were over 0.5). On the other hand, according to the factor loading contours of PM$_{10}$ (Fig. 5), high factor loading values of PC1 and PC2 were concentrated in the north and south districts of Kaohsiung City, respectively, high factor loading values of PC3 were concentrated near the Xiaogang Industry Zone (factor loading values of two stations were higher than 0.5), and lastly, high factor loading values of PC4 were scattered in the southern area of Kaohsiung City, close to the Kaohsiung harbor (the factor loading values of three stations were over 0.5). Overall, the distinct RPCs represent different PM pollution sources. The dominant sources of PM pollution in the northern district of Kaohsiung City were mainly associated with the petrochemical complex in Nanzih and Zuoying Districts, represented by PC1. On the contrary, the dominant sources of PM pollution in the southern district of Kaohsiung City were associated with the heavy industries in Xiaogang District, such as petrochemical steel, ship manufacturing, and electric power industries, represented by PC2. It is important to note that Xiaogang District houses China Steel Corporation and China Shipbuilding Corporation, which are the largest steel and shipbuilding companies in Taiwan, respectively. Other sources of PM pollution in southern Kaohsiung could be associated with PM originating from the Kaohsiung harbor and small-scale industries surrounding Xiaogang area as represented by PC3 and PC4, respectively. Previous studies have found coherent results. For example, Linhai Industrial Park in Xiaogang District exhibited high hourly PM$_{10}$ emissions, with sources originating from the heavy industries in the area, such as ship manufacturing, oil refinery, and steel industries (Liang et al., 2015). Lu et al. (2016) and Tseng et al. (2016) also reported petrochemical industries as one of the major sources of PM$_{2.5}$ emissions in Tainan, Taiwan and Chiayi, Taiwan. Overall, the numerous industrial zones located all throughout Kaohsiung City play an important role on the spatial distribution of PM$_{2.5}$ and PM$_{10}$ emissions.

The present study analyzed the diurnal pattern in time-series characteristics of the five RPCs for PM$_{2.5}$ (Fig. 6). The results showed that high component scores of PC1 appeared at three periods, namely the morning and afternoon peak traffic times (7:00–9:00 a.m. and 5:00–7:00 p.m.) and at noontime. The high component score of PC2 was also prone to occur in the aforementioned three periods, but the degree of change in the component score is slightly different from PC1. For example, the highest component scores of PC1 appeared at 8:00 a.m., noon, and 6:00 p.m. while the highest component scores of PC2 appeared at 9:00 a.m., 1:00 p.m., and 4:00 p.m. For PC1 and PC2, the peak component scores are possibly caused by human transportation activities, such as mobile emissions and dust entrainment (DeGaetano and Doherty, 2004; Rasheed et al., 2015). For example, people go and leave work during the morning and afternoon rush hours, therefore, there is an increase in vehicular use, which consequently results to an increased traffic air pollution. The same can be said for noontime, where there is also an increase in human transportation activities because it is set as lunch break for workers. Furthermore, at 9:00–11:00 a.m. and 1:00–4:00 p.m., a time generally for continued working, a decrease in the component scores can be observed for PC1 and PC2. This could be due to the fact that there is less human outdoor activities during these hours, therefore, fewer pollutants are emitted into the atmosphere. To sum up, PC1 and PC2 are more likely to occur in daytime than in nighttime and can be affected by both traffic and industrial pollution sources. For PC3, the high peaks were likely to appear after 6:00 p.m. to 8:00 a.m. from the night to the morning while the low ones were likely to occur from 9:00 a.m. to 6:00 p.m. in the daytime. The reason as to why the higher magnitudes were observed in the nighttime...
Fig. 4. Factor loadings of the four rotated principal components of PM$_{2.5}$.

compared to daytime for PC3 is still unknown and needs further investigation in the future. However, since PC3 is mainly concentrated near the Kaohsiung harbor, major port activities that mainly happen in the night could possibly explain the increased peaks of PC3. For PC4, the high values were likely to occur at 3:00–4:00 a.m. and at 9:00–10:00 a.m. while low values were likely to appear from 12 noon-12 midnight. Similarly, the high component scores for PC5 peaked at 10:00 a.m. and 4:00 p.m. while low component scores were prone to appear during nighttime from 5:00 p.m. to 7:00 a.m. PC4 and PC5 were also likely affected by human activities in daytime similar to PC1 and PC2. Overall, all PCs (except for PC3) showed clear diurnal variations with higher values generally occurring during daytime than in nighttime. Higher PM$_{2.5}$ concentrations during daytime than in nighttime is likely influenced by more industrial and transportation activities happening during the day than at night. Srimuruganandam and Shiva Nagendra (2010) reported similar results where the highest PM concentrations were observed in the day during peak traffic hour times compared to non-traffic times at night. Ye et al. (2017) also reported that there is a 10–20% increase in PM$_{2.5}$ concentrations during daytime. Other studies reported contradicting data from the present study, wherein peak PM concentrations occurred both in the daytime and nighttime (Zhao et al., 2009; Liu et al., 2014) while others reported that PM$_{2.5}$ was observed with higher concentration at night than in day (Yao et al., 2015).

The present study, however, is subject to a limitation, which is described as follows. In Kaohsiung City and in Taiwan in general, higher PM$_{2.5}$ and PM$_{10}$ concentrations are observed during the winter season than in summer.
season due to the prevalence of winter northeast monsoon carrying PM particles from mainland China all the way to Taiwan. In our study, 4-day data were obtained for the summer season while a month-whole data were obtained for the winter season to reveal associations of PM concentrations with relative humidity and type of season. Due to the varying sampling durations of the 2 seasons, we cannot clearly show the differences in PM characteristics between winter and summer seasons in the present study. Even though our study poses this limitation, nevertheless, our study still provided a novel finding regarding the correlation of relative humidity and PM (including PM$_{2.5}$ and PM$_{10}$) concentrations showing different patterns between the summer and winter seasons, which may have significant contributions to the global data on PM$_{2.5}$ and PM$_{10}$. Our study may have an important role in the mitigation measures against PM emissions from industrial operations and transportation and vehicle usage as well.

CONCLUSIONS

This study combined low-cost air sensor, IoT, principal component analysis and big data analytical techniques to analyze the monitoring results of air pollutants in Kaohsiung City. To sum up, traffic-related emissions, ship and vessel emissions, human transportation activities, and heavy industries are important sources of PM$_{2.5}$ and PM$_{10}$. PM and relative humidity are associated with each other, wherein pollutant concentration is affected by relative humidity. Higher PM concentrations are likely to occur during the winter season as compared to summer season due to the long-range transport of pollutants from mainland China.
Variation in the diurnal pattern of PM$_{2.5}$ is mainly affected by human activities and traffic-related emissions. To the best of our knowledge, our results provided one of the first big datasets including PM$_{2.5}$ and PM$_{10}$ levels in 50 air quality monitoring sites in Kaohsiung, Taiwan. These data could be of use in the mitigation measures against the serious air pollution that Taiwan is currently facing.

ACKNOWLEDGMENTS

This study was supported by the grants from the Ministry of Science and Technology (MOST) (MOST 106-2221-E-020-001-MY3) and Environmental Protection Bureau, Kaohsiung City Government (H07115). The authors give thanks to the team members in JS Environmental Technology, Energy Saving Co. Ltd., and Environmental Health and Risk Assessment laboratory, National Pingtung University of Science and Technology.

DISCLAIMER

The authors declare no conflicts of interest.

REFERENCES


Widiana, D.R., You, S.J., Yang, H.H., Tsai, J.H. and ...


*Received for review, June 27, 2019*

*Revised, July 16, 2019*

*Accepted, July 17, 2019*