Received November 23, 2019, accepted December 21, 2019, date of publication December 31, 2019, date of current version January 21, 2020.

Digital Object Identifier 10.1109/ACCESS.2019.2963341

Why Is Short-Time PM2.5 Forecast Difficult? The Effects of Sudden Events

NAI-CIH LIOU[1], CYUAN-HENG LUO[1], SACHIT MAHAJAN[2], AND LING-JYH CHEN[1], (Senior Member, IEEE)
[1]Institute of Information Science, Academia Sinica, Taipei 11529, Taiwan
[2]Department of Sociology, University of Cambridge, Cambridge CB2 1SB, U.K.

Corresponding author: Nai-Cih Liou (imritaliou@gmail.com)

This work was supported in part by the Ministry of Science and Technology of Taiwan, and in part by the Academia Sinica under Grant MOST 105-2221-E-001-016-MY3, Grant MOST 108-2218-E-001-002, Grant MOST 107-2218-E-001-006, and Grant 5591.

ABSTRACT The existing forecast models for PM2.5 concentration can be classified into long term and short term models depending on whether the forecast is performed for the next few hours or days. However, short term forecast models feature narrow forecast time and are thus vulnerable in their sensitivity to soaring variations in air quality, defined as sudden events. The purpose of this work is to investigate the causes behind these sudden events. The PM2.5 data were obtained from monitoring devices deployed in Taichung as a part of the Airbox project. The data were fed into the current short-term forecast model to forecast air quality for the next hour. Event timing was detected by feeding the forecast result as an input to the sudden event detection model. We then combined the filtered timing with factors in environment and human activities. With the application of Hierarchy Clustering, the clustering result was analyzed to find the causes of sudden events. In the springtime and summertime, unexpected changes in rainfall and temperature were critical for forecast models. Moreover, unanticipated changes in the intensity of rainfall and wind are important in the autumn and winter. For human activities, crowds of commuters, tourists, and pilgrims also have influence on unusual air quality. By carefully considering the effects of sudden events, we believe that the response ability of short time forecast can improve significantly in the near future.

INDEX TERMS Low-cost sensors, Internet of Things, data analysis, PM2.5, hierarchical clustering.

I. INTRODUCTION

A variety of environment problems caused by air pollution have gained increasing attention over the past few years. Air pollution is a major environmental risk not only to cardiovascular and respiratory health, but also to reductions in urban visibility [1]. PM2.5, fine particulate matters with a diameter of 2.5 microns or less [2], is a critical factor for air quality indicators. It has been indicated that PM2.5 concentration is associated with mortality from cardiovascular disease, myocardial infarction, stroke, and respiratory disease [3], which has been calculated as leading to the premature deaths of more than 50,000 annually in the US [4], [5]. Moreover, according to research in Taiwan, a significantly higher mortality risk was found for people with higher exposure rates of PM2.5 [6]. Air quality also has an influence on human activities and health related decisions. As the air quality deteriorates, outdoor activities become less attractive; conversely, lower particulate matter concentration enables individuals and especially families to enjoy their leisure time outdoors in good conscience. These are good reasons to acquire a deeper and more precise understanding of poisonous substance pattern. As a result, getting accurate knowledge of PM2.5 concentration in advance allows people to make beneficial decision based on future air quality. Forecast models for PM2.5 level can therefore actualize a comprehensive understanding of air quality, thereby enabling a better capacity to make health-related decisions about outdoor activities.

To date, there have been numerous studies proposing forecast models for PM2.5 [7], [8]. We distinguished long-term forecast and short-term forecast by the length of forecasting time: forecasting within one day (such as five minutes or a few hours) are considered short-term, and those more than one day (such as one month) long-term. The parameters which these two types of forecasts are based on are usually different. Long-term forecasts usually references external factors like emission data and climatic conditions.

The associate editor coordinating the review of this manuscript and approving it for publication was Lu Liu.

AIRNow Website (https://airnow.gov), for example, publicly provides national air quality forecasts for major U.S. areas of the current and next day. In the work [9], the service of AIRNow was improved with satellite aerosol observations, which identified important moderating weather patterns. In [10], Konstantinos et al. input weather data including temperature, wind speed, and air pressure, to forecast daily values of the European Regional Pollution Index (ERPI) 3 days in advance. However, most short-term forecast algorithms refer to past PM2.5 data directly, without moderating variables [11], [12].

Although forecast for short-time PM2.5 concentration has become more precise after the emergence of low-cost sensor research, sudden surrounding changes still limit forecast model performance. Forecast models for PM2.5 concentration based on short continuous data are sensitive to fluctuations in air quality resulting from sudden events. These unexpected events cause particulate-matter levels to soar abnormally in a short time, the effects of which are only later detected in forecast models. However, shortcomings of forecast models not only include the under-attention of the impact from sudden events, but also the over-attention of the rapid changes in air quality. Fluctuating PM2.5 concentrations can arise from numerous unexpected situations. Smoke from a passing smoker, factory emissions brought by a gust of wind, or burning incense from incense sticks because of traditional religious or pilgrimage activities at certain times can all make pollutant levels ascend near the sensors. Nonetheless, these sudden events with only a transitory impact can be overestimated as models apply the particulate matter concentrations into the forecast calculations for air quality. Late-response and over-attention to the influence of sudden events restrict forecast accuracy.

Adaptive Iterative Forecast is one of the short-term PM2.5 forecast models [13], which performs well most of the time but is overly sensitive to the sudden changes in nearby PM2.5 concentrations. This short-term model is thus limited in its short-term forecasts if sudden local changes in air quality are not taken into consideration.

In this paper, we have analyzed the performance of the PM2.5 short-time forecast model for the following few hours [13]. Data were collected from the low-cost, real-time sensors in the Airbox project [14], a program facilitating citizen science into the deployment and monitoring mechanisms. The PM2.5 monitoring of data in Taichung City has a 5-minute sampling rate. The main purpose of this research was to detect sudden events which would influence the performance of the short-time forecast model.

The contribution of this paper is three-fold:

1) Analysis of the forecast model to identify possible sudden events that moderate model performance.
2) Proposal of a method to detect the causes of sudden events.
3) Explanation for the relationship between sudden events in local air quality and meteorological and anthropogenic factors.

The rest of this paper is organized as follows. In Section 2, we discuss some related works and our motivation behind this research. For Section 3, we give a system overview concerning our proposed architecture and the Airbox Project. Section 4 explains the proposed method in more detail, including the feature transformation, clustering model selection, and analysis strategies. In Section 5, we scrutinize the causes of sudden events based on the meteorological and anthropogenic conditions. In the Section 6, we used the concept of a confusion matrix to validate our experiment and results. In the last section, we conclude this paper and give a short description of possible research directions in the future.

II. RELATED WORK AND MOTIVATION

An investigation into previous studies will necessarily be relatively brief because the problem we are trying to solve belongs to a novel domain. Most research efforts until now have applied modern day technologies to the chemical characteristics of irregular air pollution events. Kulkarni *et al.* [15] tracked industrial emission events with the chemical analysis of certain airborne metals, namely lanthanum and lanthanides. Gao *et al.* [16] also proposed that human behavior had an influence on air quality by analyzing the chemical variations of Beijing air pollution on hazy days. In addition to these features, natural sources such mineral dust and sea salt have been widely analyzed to detect events of high particulate matter concentration [17].

Large pollution events are well known threaten human health. Nowadays, several studies focusing on high pollutant concentration events have been conducted in various ways and domains. However, studies analyzing air quality change within a narrow time-frame are relatively sparse. This lack of investigation into short-term change in pollutant concentrations not only limits the understanding of air patterns, but also confines forecast accuracy. The motivation behind this exploration into the causes of sudden events was to raise public and researcher awareness about the significance of sudden events on air quality as well as to extend our understanding of air quality. With a better understanding of air pollutants events, short-term forecasting models will be more accurate in the future.

III. SYSTEM OVERVIEW

A. PROPOSED SYSTEM ARCHITECTURE

The proposed system architecture is shown in Figure 1. Sensors deployed in Taichung City provided real-time monitoring data, including PM2.5, temperature, and humidity levels. The data were stored in a database and are accessible to anyone at the Airbox project website (https://pm25.lass-net.org/). After acquiring hourly PM2.5 concentration forecasts from the short-term forecast model proposed by Luo *et al.* [13], it is possible to detect the timing of a sudden events. Data concerning the environment, such as temperature, rainfall, relative-humidity, wind speed and direction, and human activities, like traditional celebration, days-off, commuting, were mixed with the detected sudden event timing

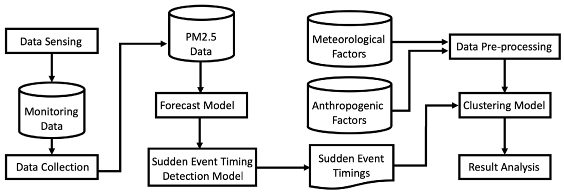


FIGURE 1. The framework of the system architecture for detecting sudden event causes.

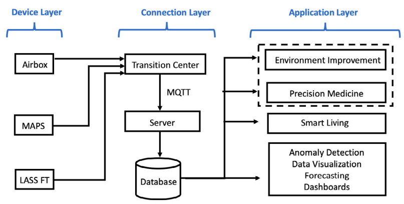
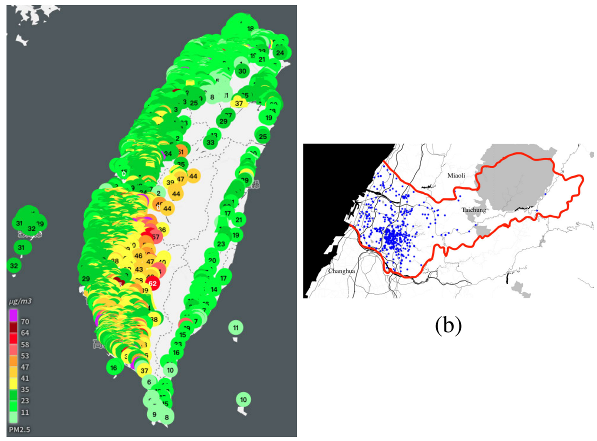


FIGURE 2. Flowchart for PM2.5 sensing in the Airbox project.



(a) (b)

FIGURE 3. Location for sensors in the Airbox project. (a) Sensors deployed in Taiwan. Color shows the PM2.5 Air Quality Index (AQI) level. (b) Sensors in Taichung used in this research.

and then input into the clustering model. With the analysis of the clustering result, we were able to identify the causes of sudden events.

B. AIRBOX PROJECT

The Airbox project combines participatory urban sensing (PUS) and an Internet of Things (IoT) system into PM2.5 monitoring to crowd-source in PM2.5 sensing and data collection [12]. The critical feature of the Airbox project was the open system architecture with principles of open hardware, open source software, and open data [14]. Cooperating with the LASS (Location Aware Sensing System), Taipei City government, and industrial companies, the Airbox project developed monitoring devices to keep track of particulate matter. Figure 2 presents the flowchart for PM2.5 sensors transmitting monitoring data in the Airbox project.

Supported by different organizations and development boards, the three various sensor versions of LASS FT (Field Try), MAPS (Micro Air Pollution Sensing System), and Airbox compose the device layer in the Airbox project [12]. LASS FT with the DHT22 temperature/humidity sensor and G3 PM2.5/PM10 sensor was developed by the LASS community and Network Research Lab at the Institute of Information Science, Academia Sinica, Taiwan. The source code, guide, and 3DP models template of LASS FT were made public online to make it customizable and available to anyone. MAPS, designed by the Network Research Lab, was equipped with the BME 280 temperature/humidity sensor and PMS5003 PM2.5/PM10 sensor. The flexibility of MAPS enables users to reach their personal requirements. Airbox, the industrial product version for PM2.5 monitoring, was the result of a collaboration between Edimax Inc. and Realtek Inc. in Taiwan, and the Network Research Lab mentioned above. Based on the HTS221 for temperature/humidity sensing and the PMS5003 for particulate matter, Airbox was found to be resistant to both water and dust.

The monitoring data from the device layer went to the server in the Network Research Lab through transition center adopting Message Queuing Telemetry Transport (MQTT) data distribution protocol [18]. Featuring low communication overhead and simple implementation, MQTT has been studied [19], [20] and utilized in IoT research for analysis in various applications [21]–[24]. After being stored in the database, the monitoring data were made public and available for everyone interested in real-time air quality.

PM2.5 monitoring data from these three sensors was accessible on the open data website (https://pm25.lass-net.org/en/). As illustrated in Figure 3a, over 7,000 low-cost, real-time sensors have been deployed in urban cities in Taiwan since March 2016 with five-minute frequency sampling rate. Besides that, PM2.5 monitoring devices had been deployed in 39 countries in the world until December, 2018.

In a previous project [25], the authors built a chatbot as an intelligent conversation interface for sending real-time PM2.5 information to users. Users could subscribe to the device and receive real-time air quality for certain sensors. Another feature lets the users to request information through the chatbot for any region users want and for any time. This system also provides an air pollution alarm service that enables users to make healthful daily plans based on air quality.

IV. METHODS

A. SAMPLING DURATION AND LOCATION

To have a better understanding of the relationship between sudden events in PM2.5 concentrations and forecast model accuracy, Taichung City, a special municipality and the second most populous city since July 2017, was selected as the analysis target. Taichung has suffered from air pollution problems caused by heavy industry factories, such as the Taichung Thermal Power Plant, Cheng Loong paper mill, and Dragon Steel Corporation, as well as factories in nearby industrial parks. More than 450 sensors,

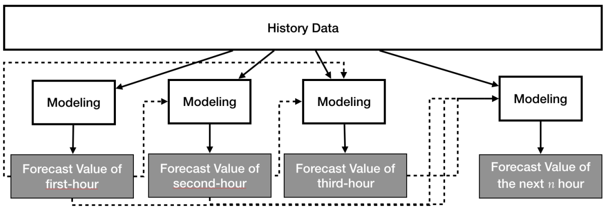


FIGURE 4. Workflow of the Adaptive Iterative Forecast (AIF).

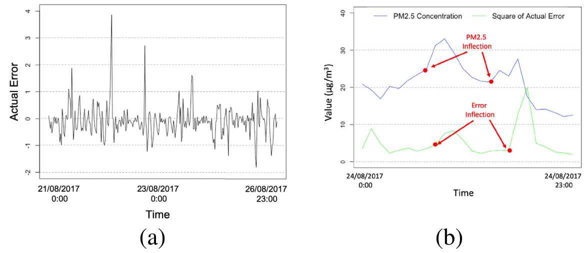


FIGURE 5. Inflection points (sudden spiking values) for actual error of the forecast model and PM2.5 concentrations. (a) Hourly forecast of actual error from 21/08/2017 to 27/08/2017. The inflection points show the sudden increase in accuracy error of the AIF model. (b) Hourly values for PM2.5 concentration and square of AIF actual error on 24/08/2017. As the PM2.5 concentrations increased, error was forecast to increase, which indicates the impact of unexpected change in air quality on forecast results.

from the Airbox project were deployed in Taichung urban zones covering a 492-km2 area and continuously monitored air quality and acquired PM2.5 concentrations data. The serious air pollution problems and high sensor density makes Taichung City a good choice for analysis. Figure 3b shows the Taichung sensor locations in the Airbox project used for PM2.5 concentration data collection during the study period from 01/01/2017 to 31/12/2017.

B. SHORT-TERM FORECAST MODEL

Hourly forecast results for PM2.5 concentration derived from the input of Taichung's monitoring data with a 30-minute sampling rate into the short-term forecast model called Adaptive Iterative Forecast (AIF) [13]. AIF is a PM2.5 concentration predictive model with the advantages of a short processing time, simple input data, and low error rate. The core idea of AIF's forecast model is to find out the relationship matrix that can make forecast error approach zero. More specifically, "A" for adaptive means that this model updates the parameters of its forecast model based on input data for each forecast process; "I" refers to the iterating forecast method which will continuously update the relation matrix though the latest forecasting values and historical data. The AIF workflow is shown in Figure 4.

Although AIF performs well for PM2.5 forecast in most cases, the unexpected variation in the particulate matters alludes to the impact of unstable environment factors, like meteorological features, human behavior, traffic situation, terrain type. Figure 5a shows hourly mean of actual error during August 21 to 26 in 2017. The inflection points, which are sudden increase in actual error, represent the abnormal change in forecast accuracy. The sharp raises in actual error intimate the abrupt change in the surrounding. The figure 5b is composed by the PM2.5 concentration and square of actual error on August 24 in 2017. The error inflection points behind the PM2.5 ones illustrate the delay response of the forecast model to the sudden raise in PM2.5 levels. To find out the causes of sudden changes in air quality, we applied the PM2.5 forecast result of AIF model in five hours into the sudden events timing detection model for further analysis.

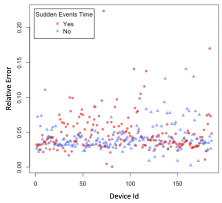
C. SUDDEN EVENTS TIMING DETECTION MODEL (SETD MODEL)

The hourly forecast result for PM2.5 concentration was applied in the SETD model to acquire the timings when sudden events happened. Air quality follows certain patterns sensor-by-sensor. Every time a sudden event occurs, an unexpected and dramatic change in air quality as well as pollutant concentration happens. Three approaches, namely the Time-series-based approach, the confidence interval-based (CI) approach, and the Boxplot-based approach, were considered to locate sudden events happening time by tracking down the abnormal change in PM2.5 concentration. The first method, the Time-series-based approach, finds peaks in PM2.5 concentration based on the air quality historical pattern in time series. The Boxplot-based approach filters timings with irregular high levels of particulate matters by employing the third quartile of sensor-based PM2.5 concentration as thresholds. The last approach was the CI-based approach, which selects the time when PM2.5 concentration covers 97.5 percent confidence intervals under a normal distribution. The judgment benchmark for detecting performance is the discrepancy between actual error medians in sudden-event and non-sudden-event time. With the assumption that forecast error increases when sudden events happen, the optimal method to detect these sudden events should have larger benchmark values.

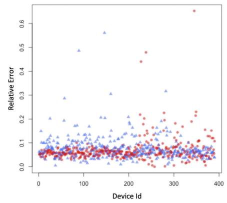
TABLE 1. Comparison between different approaches to detect sudden-event timings. Values in cells for each season is the mean difference of the relative error for each sensor when sudden events happened or not.

Approach	Spring	Summer	Autumn	Winter
Time-series-based	0.665	0.749	0.762	1.750
Boxplot-based	0.155	0.205	0.167	0.474
CI-based	0.142	0.218	0.150	0.475

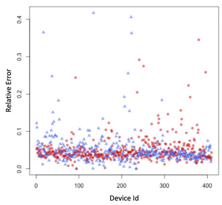
The mean discrepancy is presented in Table 1 and shows that the time-series-based approach has the best ability to detect the timings when sudden events happen with the characteristic of higher PM2.5 concentration compared to normal timing. Based on these results, the first approach for the SETD model was finally selected to detect sudden-event timing and was realized by utilizing the *pracma* package in R language to locate peaks in PM2.5 concentration. Figure 6 illustrates the relative error in normal timing and sudden-event timing in different seasons and indicates the impressive detection ability of the SETD model.



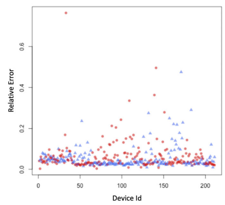
(a) Scatter plot in the Spring (February to April, 2017).



(b) Scatter plot in the Summer (May to July, 2017).



(c) Scatter plot in the Autumn (August to October, 2017).



(d) Scatter plot in the Winter (November to January, 2017).

FIGURE 6. The results of the time-series-based approach to detect sudden events timings in different seasons. The red and blue circles indicate the sudden events timings and normal timings, respectively. Each point is the relative error in a period of time.

D. ADDITIONAL FACTORS

In the data pre-processing section, meteorological and anthropogenic features were both taken into consideration as potential causes of sudden events. Environmental factors have been proven to be critical to air quality change and have already been discussed at length in the literature [26], [27].

Anthropogenic factors have been the focus of fewer studies. We took into account this category of features, which include air quality change resulting from cars in rush hours, celebrations, and religious activities. One important anthropogenic factor that aggravates air pollution involves traffic: exhaust emissions from vehicles like automobiles, buses, and motorcycles, especially during peak traffic hours. Commuting rush hours were defined as 7 to 9 am and 5 to 7 pm on normal weekdays from Monday to Friday. Moreover, crowds of weekend travelers are also a hidden factor that can affect air quality in a short time. Traffic on Saturday and Sunday was therefore seen as a weekend factor. Moreover, traditional events for religious and cultural purposes sometimes trigger air pollution. In Taiwan, for example, many people burn "spirit money" as an offering to the ancestors as well as Gods, and also burn incense sticks to convey the prayers of the faithful to heaven. Unfortunately, the incense creates increase in particulate matter concentrations. To have more comprehensive consideration in the human activities which cause sudden events, religious events, such as the celebration for a god's birthday, were included in this study. Information of temples in Taichung was acquired from the open data website [28]. There are many various gods widely worshiped in Taiwan, but the top 20 most worshipped gods with temples in Taichung City were chosen to represent all the gods in Taichung. The gods' birth dates in the lunar calendar were converted into solar calendar dates and taken as one anthropogenic feature. Moreover, it is important to consider the pollution-causing activities of certain festivals, like the Taiwanese customs of having barbecues outside around the Moon Festival and setting off firecrackers to celebrate the Lunar New Year. The special-day factor was classified as 1 if the day is a god's birthday or festival celebration day. In total, we identified five meteorological factors concerning environmental situations and three anthropogenic features. These additional factors above were considered before implementing the clustering process.

E. CLUSTERING MODEL

In the clustering process, the clustering model utilized the hourly combined data of sudden event timing chosen from the SETD model as well as data of the additional environment and human factors. The clustering model applied in this research was the hierarchical clustering algorithm. Based on its strengths of flexible argument choice, relaxed requirement for cluster number, and visualized results, the hierarchical clustering algorithm has been applied to solve a variety of problems in multiple industries such as biology, pharmacology, economics, and earth sciences [29]–[33]. Hierarchical clustering builds a hierarchy of clusters using bottom-up or top-down merging-splitting strategies. To reach optimal clustering results, it is important to choose the appropriate metric to compute the distance between observations and to select the proper linkage for the measure of the dissimilarity between sets of observations. In this paper, the chosen clustering strategy was divisive, and the decided metric and linkage were Euclidean distance and Ward's criterion, respectively. The Euclidean distance is the straight-line distance between two points in Euclidean space. The Euclidean distance for point p, represented as $(p_1, p_2, p_3, .., p_n)$ and point q, represented as $(q_1, q_2, q_3, .., q_n)$, is calculated with formula 1. Ward's criterion is a criterion to decide clusters in each step, which depends on the optimal value of an assignable objective function [34]. As a special case of Ward's criterion, Ward's minimum variance method takes the error sum of squares as the objective function and applies squared Euclidean distance as the initial distance of sets. The criterion in this hierarchical clustering model is the improved version known as the Lance-Williams algorithm, which implements Ward's minimum variance method recursively at each step to minimize the dissimilarity between sets [35].

After obtaining the clustering results based on the meteorological and anthropogenic factors, the outcome in the groups was embedded in the analysis procedure to scrutinize the causes of sudden events with case studies.

$$d(p, q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \ldots + (p_n - q_n)^2} \tag{1}$$

F. RESULT ANALYSIS

Besides the case study for each cluster, the results were also applied into the Kullback-Leibler Divergence (KL Divergence) algorithm for further analysis. KL Divergence has been used by researches in multiple fields to find how much two possibility distributions are different from each other. For two discrete probability distributions in the same probability space P and Q, the KL Divergence formula is demonstrated in the equation 2. A KL divergence of 0 suggests that the two compared distributions are identical. The higher the score of the KL divergence, the larger the dissimilarity of the distributions. In this paper, the KL divergence helped make it feasible to compare the difference between factors in normal time and sudden-event time. The evaluation of the features' differences is practically achieved by ranking the KL divergence of each environment factor in the two time periods.

$$D_{KL}(P\|Q) = -\sum_{x\in X} P(x)log(\frac{Q(x)}{P(x)}) \tag{2}$$

V. RESULTS

This section scrutinizes the results of both the KL divergence and clustering in order to interpret the impact of sudden events. The analysis of KL divergence for meteorological factors reveals the relative importance of features in normal and sudden-event time. The case study revealed the causes of sudden events in terms of a season-base for environment factors and in cause-base for human-activity features.

A. KL DIVERGENCE RESULTS

This part of the paper discusses and interprets the impact of changing features by comparing the KL divergence of environmental factors in different seasons. What follows is an account of the specific findings from the perspective of factor.

Among all the features, the rainfall factor demonstrates the highest KL divergence. This finding is reasonable based on the fact that rainfall always plays a direct role in purifying air in instant time. The change of the relative humidity factor in the previous hour ranks second. This phenomenon reveals the potential joint relationship between the values of rainfall and humidity. Relative humidity represents the current amount of water vapor in the air given its maximum possible value. This ratio is determined by temperature and air pressure. High relative humidity means a high level of water vapor in the air and an increase in the probability of precipitation in most instances. The relative humidity reaches one hundred percent on rainy days. The attribute of relative humidity accounts for the high ranking in the KL divergence results. The factor following the relative humidity in the rank of KL divergence was wind speed. Wind conditions affect air quality in direct and immediate ways. Gusts of wind disperse air pollutants towards other places and drives down both pollutant and PM2.5 concentrations.

The KL divergence in each season suggests additional characteristics of meteorological factors relating to sudden events. Comparing the results in autumn and winter, changes of temperature and wind speed were more highly ranked but acted relatively differently. This finding suggests the influence of urban heat island on sudden changes in air quality.

TABLE 2. Number of sudden events in each group based on environmental factors.

Cluster id	Spring	Summer	Autumn	Winter
1	340	412	362	82
2	347	450	794	227
3	166	370	2	129
4	1	25		2

The phenomenon was first described by Luke Howard [36] who noticed the significantly higher temperatures caused by human activities in metropolitan areas compared to the surrounding rural ones. The possible causes of this urban heating phenomenon include the evapotranspiration due to the lack of plants and trees, the urban canyon effect caused by the sunlight reflection from skyscrapers, or the high solar radiation absorption of dark road surfaces and building surfaces. Increased temperatures and windless conditions arise as this effect happens. Another outcome is the urban dust dome, which not only makes the wind be drawn to city centers and descend on the surrounding areas but also traps pollution in the air above urban spaces. The existence of urban heat island provides the high ranking for KL divergence values for temperature and wind seed factors in the Spring and Summer.

As for the KL divergence rankings in the Autumn and Winter, the ranking results show the relatively higher difference of rainfall and wind speed factors when sudden weather events happen in these two seasons. Seasonally dry riverbeds, aeolian dust, and monsoons might be responsible to these findings. Most rivers in western Taiwan flow intermittently and have decreasing water levels in Autumn and Winter compared to the plentiful waters from the "plum rain" (or "Mei-Yu") season in Spring and typhoons and heavy rainfall in Summer. During the dry periods, the dry and dusty riverbeds are exposed. Dust blown up by monsoons not only coagulates with pollution particles, but can also deteriorate both air quality and visibility in a short time. The characteristic of aeolian dust caused by monsoons and dry riverbeds explains the high KL divergence of rainfall and wind speed features.

B. CLUSTERING RESULT OF ENVIRONMENT FACTORS

Table 2 shows the group number after the hierarchical clustering algorithm in each season. A detailed investigation of the characteristics of each cluster was conducted and described in the following.

$$Mean\ Change = \frac{\sum_{i=1}^{n} V_{i(t-h)} - V_{i(t)}}{n} \tag{3}$$

1) SPRING

The first group in the spring clustering results demonstrates the capacity of the spring rain to induce changes in air quality. Figure 7a demonstrates the environmental feature details with the mean change of each factors during the previous five hours with the comparison of each hour. In the equation 3, n means there are n AirBox devices, $V_{i(t)}$ means the value in current hour and $V_{i(t-h)}$ means the value of previous h hours. With a slight change in most environmental attributes, except

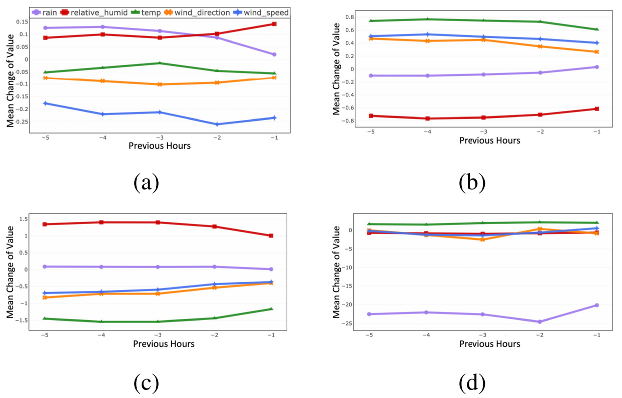


FIGURE 7. **Environmental features details of each cluster in the springtime. "Previous Hours" means the amount of hours ahead of forecasting. (a) Cluster 1: spring rain effect. (b) Cluster 2: northeast monsoon influence. (c) Cluster 3: prefrontal warm advection power. (d) Cluster 4: unknown human causes impact.**

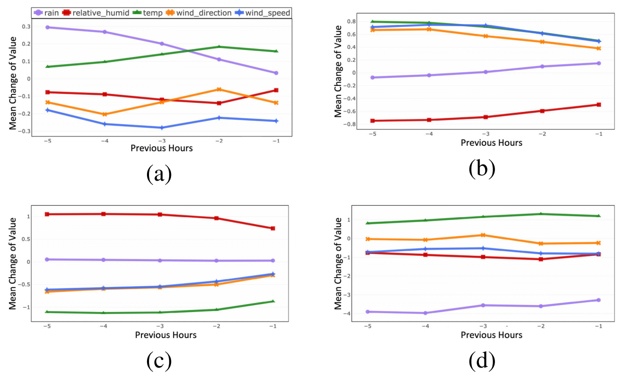


FIGURE 8. **Environmental feature details of each cluster in the summertime. "Previous Hours" means the amount of hours ahead of forecasting. (a) Cluster 1: plum rain front effect. (b) Cluster 2: photochemical reaction influence. (c) Cluster 3: subsidence inversion power. (d) Cluster 4: thermal thunder shower impact.**

for stronger wind and diminishing rainfall, the first cluster reflects the atmospheric effect of spring rain. Spring rain forms when two weather fronts collide and push against each other: the Siberian High, which originates in northeastern Eurasia, brings dry, cold, and strong cool fronts to Taiwan; at the same time, there is the warm and powerful west wind caused by the Aleutian Low. The meeting of these two air masses gives rise to shear lines, the lift of air, and sprinkling of spring rain, which typically falls from February to April. The scattered showers, chilly weather condition, and unstable wind persists during the spring rain periods. The rainfall and wind have purification effects that lower air particulate matter concentrations. As a result, the spring rains cause sudden events in air quality in the springtime.

The second cluster depicts the interplay of the influences of the northeast monsoon and the southern China rainy system. As illustrated in Figure 7b, the major properties of the environmental features are falling temperatures, calm winds shifting to the east, and heavy rainfall. Cold weather is also expected when rains from southern China accompanies most of these situations. As the wet weather system from southern China moves to east and approaches Taiwan, unstable atmosphere conditions, thick clouds, and prolonged precipitation are found. Even though the rainfall brought from the wet system from southern China performs purification effects similar to that of the spring rain, particulate matter is brought from China at the same time. On the top of that, the northeast monsoon also acts as one of the major sources of air pollution from overseas. Based on the discussion above, the complicated effect of both the northeastern monsoon and the rain system from southern China creates unexpected changes in pollutant concentrations and results in sudden events in air quality.

The third cluster in the springtime shows the ability of the prefrontal warm advection to change of air quality. In Figure 7c, the major characteristics of this cluster are present. The meteorological features in the time data of the this cluster are the dropping temperatures, stronger winds shifting away from the east, and diminishing rainfall. The prefrontal warm advection arises when the stationary front hovers the East China Sea and approaches to Taiwan. The decreased pressure gradient around Taiwan leads to weaker winds, slower movement of air currents, and diminishing air effects. What acts as a consequence of the prefrontal warm advection are winds blowing from the directions around the southwest, scattered showers,and sultry weather. Based on these observations, the prefrontal warm advection has been shown to have a significant effect on the pollution concentration and the occurrence of sudden events of air quality.

The final springtime cluster in the springtime indicates that random and unknown human behaviors have the power to make air quality unstable in a short period of time. Figure 7d shows the detailed features of this cluster. Considering the influential capability of anthropogenic features, limited amount of time data in this cluster, and the various environment factors have already been described to, the ensuing discussion focuses on the relationship between human actions and air pollution. Increased particulate matters concentration has been closely linked to human behaviors. Air pollution is caused by various human actions, such as the exhaust gas from automobiles, smoke from cigarettes, and incense from firecrackers. Furthermore, these acts are often unpredictable and irregular, which means that unknown anthropogenic features can cause unforeseen changes in airborne particulate matters and lead to sudden events in air quality.

2) SUMMER

The over four hundred time samples in the first summertime cluster shared the same attributions associated with the plum rain front. The overall attributes of the first cluster (i.e., the larger rainfall and plummeting temperatures following temperature increases), are common plum rain front phenomena in the rainy season (Figure 8a). According to the Taiwan EPA, the plum rain front is a stationary front forming with the equilibrium between the continental cold air mass and the Pacific warm air mass. When the warm air is strong enough to push the plum rain front north, the rainy season ends. This front brings precipitation to Taiwan starting from the late spring and ending in the early summer. During the plum rain period, the prolonged precipitation, heavy thundershowers, high humidity, and low temperatures often take place. The ability of precipitation to remove air pollution has been proved. Therefore, it is reasonable to take plum rain as the key property for the first cluster.

The second cluster in the summer implies the effect of the photochemical reaction on sudden changes in air quality. The number of time samples which were detected when sudden events happened and which were classified into this cluster was 450. The environmental characteristics for this clustering were the falling temperatures, soaring relative humidity, decreasing rainfall, infrequent wind shifting to the east (Figure 8b). This phenomenon might be attributed to photochemical reactions, which generate secondary pollutants. The study of the photochemistry of air pollution has been explored for many years and it has been proved that there is a relationship between photo-chemical reaction with air pollution [37], [38]. The air pollution caused by photo-chemical reaction has not only caused problems in western countries like the United Sates [39], but has also created serious environmental issues in recent decades in East Asian nations such as Japan [40]. The chemical mechanism behind the photochemistry of air pollution is that the particles in exhaust gas emitted by automobiles or other resources differ greatly after absorbing the sunlight. Consequently, this cluster illustrates that the photo-chemical reactions in summer brings fluctuations in air quality and impacts short-term forecast model performance.

The third summer cluster presents the main possible cause of sudden changes in air quality, which is the effect from subsidence inversion. The atmospheric attributes of the timing in this group, as shown in Figure 8c, are the increased temperature, drop in relative humidity, stronger wind, and wind blowing away from the east. Subsidence inversion might shed light on the primary shared characteristics. In normal atmospheric conditions, the air that is higher is cooler than the air close to the Earth's surface, where warmth is due to the absorption of solar radiation. However, air temperatures increase at higher altitudes when subsidence inversion occurs. This abnormal structure in vertical air temperature suppresses atmospheric convection and makes the air still. Subsidence inversion also traps pollution particles in the inversion layer and prevents the lifting of pollutants from the surface to higher altitudes. Subsidence inversion always happens on clear and sunny days with sunlight reactions, without interference from rainfall. The consequence of subsidence inversion is the deterioration of air quality, which validates the sudden event cause for the third cluster.

For the analysis of the fourth cluster, afternoon thunderstorms is the logical cause for sudden change in air quality. During most of these cases, temperatures drop, rainfall increases, and wind speed is unstable (Figure 8d). These indications are the features effects from the thunderstorm in the afternoon, which is one of the most frequently observed weather phenomena in Taiwan during the Summer, later in the afternoon. Triggered by the intense convections of warm air, the rapid cooling after the upward movement of air, and the falling of droplets when dew point is reached, the thunderstorm always happens after the surface of the Earth absorbs sunlight. With the down-draft from the falling of rain drops, in most situations, afternoon thunderstorms are often accompanied by strong winds. The heavy rainfall of thunderstorms purifies the air and lowers pollutant concentrations, which changes air quality at the same time. Based on the forming conditions and impacts mentioned above, afternoon thunderstorms account for the sudden events in air quality.

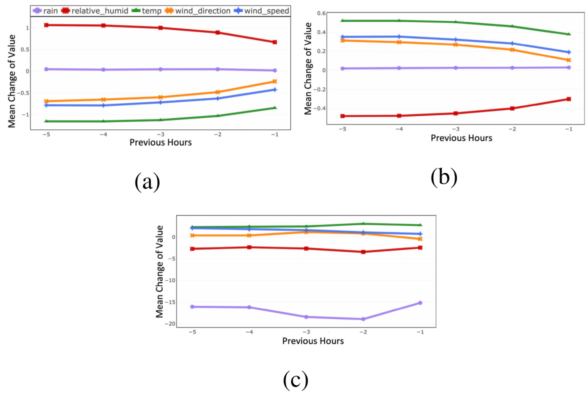


FIGURE 9. Environmental feature details of each cluster in the autumntime. "Previous Hours" means the amount of hours ahead of forecasting. (a) Cluster 1: Pacific high effect. (b) Cluster 2: weak northeast monsoon influence. (c) Cluster 3: unknown human causes power.

3) AUTUMN

The first cluster data is proof of the effect of the Pacific high on the change in air quality with special environmental features like skyrocketing temperatures and wind speeds, and reductions of rainfall (Figure 9a). The North Pacific High is a subtropical anticyclone located in the Pacific Ocean and is larger and more intense in warm seasons than in cool ones. The moist wind from the Pacific high brings heavy rainfall in Taiwan in the autumn particularly. As the center of the mass rises above Taiwan, clear and warm days occur, temperatures increase, and rainfall decreases. These steady atmospheric conditions pose problems for convection and the accumulation of air particulates, which can make the Pacific high responsible for the sudden change in air quality.

Moreover, the second cluster in the autumn time also indicates a weak northeast monsoon that can trigger sudden events in air quality. As presented in Figure 9b, the main environmental features in the timing samples of this cluster are decreasing temperature, calm and nearly wind-free conditions, and reduced rainfall, which are all consequences of a weak northeasterly monsoon typical in the autumn and winter. As the monsoon subsides, the slower wind speed results in the slower dispersion of pollutants. Further, the location of Taichung City is on the leeward side of Taiwan's Central Range, which worsens the spread of pollution. The lack of air

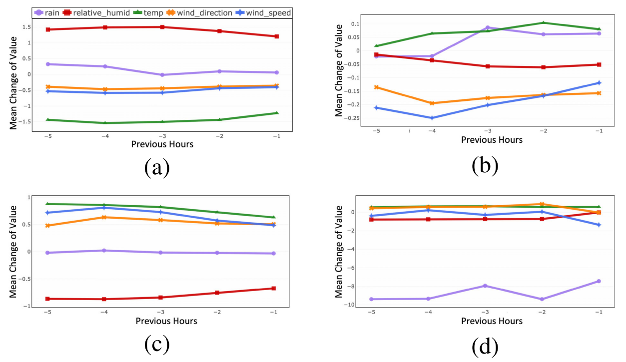


FIGURE 10. Environmental feature details of each cluster in the wintertime. "Previous Hours" means the amount of hours ahead of forecasting. (a) Cluster 1: Philippine Sea anticyclone effect. (b) Cluster 2: unknown human causes influence. (c) Cluster 3: descending current power. (d) Cluster 4: cold current impact.

pollution dispersion, caused by the weak northeast monsoon, can unexpectedly change air quality and lead to sudden events in air.

The final cluster in the autumn clustering results shows the influence of unknown human activities on sudden changes in air quality. As can be seen in Figure 9c, the major meteorological attributions in the timing samples for this cluster data involve relatively unstable changes in all factors compared to the other clusters' data. Taking together the minor alteration, environment factors in the other groups, and the paucity of data in this cluster, the irregular causes of sudden events may be associated with unknown human actions. Random human behaviors, such as gas from firecrackers or cigarettes, can initiate unanticipated air quality changes and thereby trigger sudden events.

4) WINTER

The first winter cluster illustrates the power of the Philippine Sea anticyclone (PSAC). Figure 10a shows that soaring temperatures, strong winds shifting away from the east, and unstable rainfall suggest the effects of the PSAC. The evolution and maintenance of the Philippine Sea anticyclone is the result of multiple particular meteorological conditions, such as the proper sea surface temperature, sea currents, and wind. The extreme period of the El Niño-Southern Oscillation (ENSO) generates in the Pacific Ocean an anomalous anticyclone, which induces a weak monsoon in East Asia, high air surface pressure in the tropical western Pacific, and below average sea temperatures along the eastern Pacific coast [41]. The critical role played by the East Asian winter monsoon in the PSAC system has been demonstrated by the research conducted by Lau and Nath [42]. The powerful south wind weakens the northeast monsoon in the wintertime and introduces warm air from the tropical sea surface, which leads to above average winter temperatures. Elevated PM2.5 concentration during El Nino years is observed not only in southern China [43] but also in Taichung City, in the middle Taiwan [44]. As a consequence, the PSAC is one of the crucial causes of sudden events in air quality change in the winter.

In the second cluster result (Figure 10b), the relatively slight change in atmospheric conditions as compared to the other clusters indicate stable environment situations and implies the influence of unknown anthropogenic features on unexpected air quality change. In addition to the exhaust emissions from transportation vehicles, the open burning of biomass also has the potential to produce a lot of particulate matter. The period of harvest for the second cultivation of rice crops in Taiwan starts in the winter and ends the following spring. To deal with the considerable amount of straw after reaping the rice, farmers always burn this waste as the most convenient method. The particles and gas produced by the straw burning immediately triggers air pollution and makes the air quality unstable. As a result, human behavior is one of the major causes of sudden events in the air.

Descending air currents on leeward mountain slopes as the east wind blows through the Central Range is another reason of air quality change in a short time, which is depicted in the third cluster with main characteristics of decreasing temperatures, calm winds shifting near to the east, and unpredictable rainfall (Figure 10c). Descending air currents obstruct the diffusion of pollutants in Taichung City, which is located in the central region of western Taiwan and in a basin landform that blocks the air currents, which in turn makes pollutant dispersion difficult and allows the accumulation of pollutants. The capacity of descending currents to increase pollution concentrations is a cause of sudden events in air quality.

In the final winter cluster, the meteorological features reveal the influence of cold air currents on unstable particulate matter concentrations. As illustrated in Figure 10d, the characteristics in the majority of these cases are diminishing temperatures, unsteady wind shifts close to the east, and heavy downpours, which all indicate the effects of cold currents. Cold spells from China push southward and blast Taiwan in the winter, causing plummeting temperatures. What accompanies the cloud mass is the particulate matter originating from northern China (Beijing, Tianjin, and Hebei), which acts as the major conduit for transnational pollutants in Taiwan. According to this 10-year research in Taiwan, the haze particles transported to Northern Taiwan were associated to the relatively high PM2.5 levels [45]. With the wind of cold spells, the PM2.5 would move to central Taiwan. Because of this overseas air pollution carried by cold currents from China, concentrations of five types of air pollutants surge in short periods of time and often cause sudden events in air quality.

C. CLUSTERING RESULT OF ANTHROPOGENIC FACTORS

This section will not only analyze the clustering results related to the meteorological features on sudden event causes, but will also discuss the causes of sudden events from the perspective of anthropogenic actions. Factors concerning human behaviors not only play essential roles on the levels of particulate matter concentrations, but also lead to unexpected changes in air quality. Table 3 lists the clustering results of the timing data number in each group based on

TABLE 3. Number of sudden events in each group based on anthropogenic features.

Cluster id	Spring	Summer	Autumn	Winter
Commuting	106	270	249	94
Festivals during Rush Hours	189	69	74	42
Weekends	159	322	304	97
Festivals on Weekends	105	58	46	33
Unknown Reasons	295	550	485	174

anthropogenic features. Five major causes of unexpected air quality changes are found and displayed in the Cluster ID column. The causes of the sudden changes in air quality are also elaborated in this section. Finally, the section finishes with a discussion on the relationship between air quality conditions and human behavior.

1) COMMUTING CLUSTER

Commuting behavior plays a crucial role in the sudden changes to air conditions. In Taiwan, local pollution is the major source of particulate matter. In terms of the source of domestic pollution, exhaust fumes from vehicles are one of major contributors to PM2.5 concentrations. Large crowds of people commute to work or school every day and this leads to heavy congestion on roads during rush hours in the morning (7-9 AM CST) and evening (5-7 PM CST). The gases generated from heavy traffic lower air quality levels and trigger sudden events in air quality.

2) FESTIVALS DURING RUSH HOURS CLUSTER

When festival times co-occur with morning or evening rush hours, air quality tends to change quickly. Pollution is exacerbated when the effects of transportation are combined with those of festivals. For example, the Mazu international festival exemplifies the synergistic pollution effects of worshippers and commuters. The pilgrimage of the Taoist sea-goddess Mazu covers four counties in Western Taiwan and is more than a local religious folklore practice; it is one of the biggest cultural celebrations in Taiwan with thousands of devout followers from all over the world not only burning incense and spirit money, but also setting off firecrackers, which release hazardous and malignant chemicals such as benzene and aldehyde. During this time, real-time government monitoring recorded harmful levels of PM2.5 concentrations [46]. Such festival worshipping behavior intensifies air pollution, which adds to the pollution caused by rush hour traffic, which in turn leads to sudden increases in airborne pollutants.

3) WEEKENDS CLUSTER

This cluster based on anthropogenic factors demonstrates the effect of travel and tourism on unstable air quality. Transportation-related pollution rises with the increase in the number of weekend excursionists, which further contributes to the deterioration of air quality. Previous research has identified the positive relationship between air pollution and tourism by showing that an increase of the percent of tourists number makes 0.45 percent higher in particulate matter concentrations [47]. Poor air quality and thick haze makes temporal distribution of tourist, which leads to seasonally high pollution levels [48]. As a result, the effect of weekend tourists on divergent patterns of pollution concentrations and sudden events of air quality is critical for defining the major features shared by the time data in this cluster.

4) FESTIVALS ON WEEKENDS CLUSTER

Based on the discussion above, this cluster illustrates the interplay effect when festivals coincide with weekends. Air quality worsens when excursionists increase on weekends. Moreover, festival activities also have a similar negative effect on particulate matter production. Festivals held on weekends lead to the aggravation of air quality and volatility of pollution levels. The Tomb-Sweeping Day, also known as the Qing Ming Festival, epitomizes the effect of weekend excursionists and religious or festival activities. At this time, families show their respect for as well as memory of the ancestors by visiting the graves, sweeping the tombstones, and making ritual offerings of traditional food dishes; the burning of incense sticks and silver-leafed spirit money is another common practice on the Tomb-Sweeping Day, which inevitably raises pollution levels. The combination of harmful effects resulting from festival activities and weekend traffic congestion is linked to the increase in the air pollution. The interplay effect of these two pollution sources multiply the particulate matter concentrations and the numbers of sudden events in air quality.

5) UNKNOWN REASONS CLUSTER

Besides the causes of sudden events in quality described above, the final cluster illustrates the effects of various man-made sources of air pollution that escape the above analyses. These include gases and smoke from fumigation, cooking, and smoking, which all contribute to air pollution without regular patterns. The influence of these unquantified anthropogenic sources on particulate matter concentrations are also causes of sudden events in air quality.

VI. VALIDATIONS

The problems we are trying to solve here are unsupervised problems, which means that there are no clear answers to those. However, to validate our findings, we compare our results with the official reports, which nonetheless have their own limitations. We applied daily reports from the Central Weather Bureau (CWB) [49] to validate our experiment results. The CWB is the authoritative and major information weather resource in Taiwan, which can make the CWB report data a proxy for validation data. The report articles contain information about weather situations, air quality conditions, and major meteorological events, with sentences such as "Taiwan was surrounded by the low pressure front, which brought stable weather" (). To compare the CWB reports and our results, the daily articles were transformed into a boolean list. Each element represents whether or not certain meteorological condition occurred. The value 1 means that

TABLE 4. Confusion table for the third cluster in the summer.

	Sudden Events	Normal Times	Sum
Keywords Exist	TP=13	TN=0	13
Keywords not Exist	FP=73	FN=6	79
Sum	86	6	92

specific keywords appeared on that day's report, and the value 0 means the report did not contain the keywords. Our clustering results were also transformed into a boolean list with the same mechanism. The value of 1 was used if there were more than one time interval that detected a sudden event timing, and the value of 0 if all time intervals were normal in air quality. After transforming the data, we used the concept of a confusion matrix (used in classic classification problems) as the validation approach. The traditional confusion matrix seeks to identify the relationship between a model's forecast results and reality. Normally, there are four numbers that serve to evaluate model performance. To fit the situation here, the definitions of these four terms are modified in the following.

1) **True Positives (TP)**: The days on which we predicted the occurrence of sudden events, and the CWB also reported the related meteorological conditions.
2) **True Negatives (TN)**: The days on which we predicted that certain sudden events did not occur, but the CWB nonetheless reported meteorological conditions.
3) **False Positives (FP)**: The days on which we predicted the occurrence of certain sudden events, but the CWB reported the related meteorological conditions did not happen.
4) **False Negatives (FN)**: The days on which we predicted certain sudden events did not occur, and the CWB also reported no relevant meteorological conditions.

By applying the concept of the confusion matrix, the validation of clustering results becomes feasible. The following are the validation results for the subsidence inversion cluster for the summer time. In the confusion matrix is shown in Table 4, the 13 days of the TP term represents the number of times when 1. we detected the sudden event called subsidence inversion, and 2. the CWB also reported similar findings. The value of FN term means that there were 6 days in which neither our model nor the CWB reports found the occurrence of the subsidence inversion. The TN term value is 0, which means that there was no day in which we detected normal conditions whereas the CWB reported subsidence inversion. The FP with value of 73 days means that on these days, our model detected the occurrence of sudden events but the CWB reports did not have the same findings. Although the FP value, which is known as Type 1 Error, seems high, this result is reasonable. The CWB reports contain the major meteorological events for the whole day in Taiwan, but the time scale of sudden events in our cluster is small, ranging from minutes to hours. As a result, it is reasonable that in discrepancies in our and CWB data are due to different time sampling of data: our model detected the happenings of sudden events across shorter time intervals which were undetected in the daily CWB data.

VII. CONCLUSION AND FUTURE WORK

To control for air pollution problems, forecast models have been used as an effective approach. However, the effect of sudden changes in air quality for short-time forecast models has not been fully studied. Therefore, this paper analyzes the causes of unexpected change factors in air quality using an analysis of the forecast results of AIF in Taichung during 2017.

There are three key conclusions drawn from this study.

1) In the experiment on the environment factors, the KL divergence results clearly show the power of purification capability of rainfall. From the perspective of seasons, humidity and wind direction accurately indicate the urban heat island effect in the summer. Moreover, the high ranking of rainfall and wind speeds effectively indicate influence of intermittent streams and aeolian dust from riverbeds in the autumn and winter.
2) According to the case study based on clusters of meteorological factors, seasonal and meteorological sudden events were detected.
3) The anthropogenic features such as commuting and festival activities were also shown to be potential triggers of sudden events in air quality.

In this work, we analyze the causes of sudden events by considering local environmental and atheological factors in Taichung. To achieve this, we apply the short-term forecast model and point out the existence of sudden events. These studies have many potential applications, such as to increase the efficacy of short-term forecast models and to detect the happening of sudden events. For the future work, we would like to take forecasting lags into consideration as integrating our current findings with the forecasting model. The uniqueness of different locations environment is also significant. We would like to extend this study by testing different areas in Taiwan and by considering more local factors.

ACKNOWLEDGMENT

The authors thank Edimax Inc. and the LASS community for their support, technical advice and admin-istrative assistance.

REFERENCES

[1] *Ambient (Outdoor) Air Quality and Health*. World Health Organization, Geneva, Switzerland, Sep. 2016.
[2] Y.-L. Zhang and F. Cao, "Fine particulate matter ($PM_{2.5}$) in China at a city level," *Sci. Rep.*, vol. 5, Oct. 2015, Art. no. 14884.
[3] L. Dai, A. Zanobetti, P. Koutrakis, and J. D. Schwartz, "Associations of fine particulate matter species with mortality in the United States: a multicity time-series analysis," *Environ. Health Perspect.*, vol. 122, no. 8, pp. 837–842, Aug. 2014.
[4] A. Kusiak, H. Zheng, and Z. Song, "Short-term prediction of wind farm power: A data mining approach," *IEEE Trans. Energy Convers.*, vol. 24, no. 1, pp. 125–136, Mar. 2009.
[5] X. Feng, Q. Li, Y. Zhu, J. Hou, L. Jin, and J. Wang, "Artificial neural networks forecasting of $PM_{2.5}$ pollution using air mass trajectory based geographic model and wavelet transformation," *Atmos. Environ.*, vol. 107, pp. 118–128, Apr. 2015.
[6] N. T. Hung, H. W. Ting, and K. H. Chi, "Evaluation of the relative health risk impact of atmospheric PCDD/Fs in $PM_{2.5}$ in Taiwan," *Aerosol Air Qual. Res.*, vol. 18, no. 10, pp. 2591–2599, 2018.

[7] X. Zhao, X. Zhang, X. Xu, J. Xu, W. Meng, and W. Pu, "Seasonal and diurnal variations of ambient $PM_{2.5}$ concentration in urban and rural environments in Beijing," *Atmos. Environ.*, vol. 43, no. 18, pp. 2893–2900, Jun. 2009.

[8] P. E. Saide, G. R. Carmichael, S. N. Spak, L. Gallardo, A. E. Osses, M. A. Mena-Carrasco, and M. Pagowski, "Forecasting urban PM_{10} and $PM_{2.5}$ pollution episodes in very stable nocturnal conditions and complex terrain using WRF–Chem CO tracer model," *Atmos. Environ.*, vol. 45, no. 16, pp. 2769–2780, May 2011.

[9] J. Al-Saadi, J. Szykman, R. B. Pierce, C. Kittaka, D. Neil, D. A. Chu, L. Remer, L. Gumley, E. Prins, L. Weinstock, C. Macdonald, R. Wayland, F. Dimmick, and J. Fishman, "Improving national air quality forecasts with satellite aerosol observations," *Bull. Amer. Meteor. Soc.*, vol. 86, no. 9, pp. 1249–1262, Sep. 2005.

[10] K. P. Moustris, I. C. Ziomas, and A. G. Paliatsos, "3-day-ahead forecasting of regional pollution index for the pollutants NO_2, CO, SO_2, and O_3 using artificial neural networks in Athens, Greece," *Water, Air, Soil Pollut.*, vol. 209, nos. 1–4, pp. 29–43, Jun. 2010.

[11] M. Niu, Y. Wang, S. Sun, and Y. Li, "A novel hybrid decomposition-and-ensemble model based on CEEMD and GWO for short-term $PM_{2.5}$ concentration forecasting," *Atmos. Environ.*, vol. 134, pp. 168–180, Jun. 2016.

[12] S. Mahajan, L.-J. Chen, and T.-C. Tsai, "Short-term $PM_{2.5}$ forecasting using exponential smoothing method: A comparative analysis," *Sensors*, vol. 18, no. 10, p. 3223, Sep. 2018.

[13] C.-H. Luo, H. Yang, L.-P. Huang, S. Mahajan, and L.-J. Chen, "A fast $PM_{2.5}$ forecast approach based on time-series data analysis, regression and regularization," in *Proc. Conf. Technol. Appl. Artif. Intell. (TAAI)*, Nov. 2018, pp. 78–81.

[14] L.-J. Chen, Y.-H. Ho, H.-C. Lee, H.-C. Wu, H.-M. Liu, H.-H. Hsieh, Y.-T. Huang, and S.-C. C. Lung, "An open framework for participatory $pm_{2.5}$ monitoring in smart cities," *IEEE Access*, vol. 5, pp. 14441–14454, 2017.

[15] P. Kulkarni, S. Chellam, and M. P. Fraser, "Tracking petroleum refinery emission events using lanthanum and lanthanides as elemental markers for $PM_{2.5}$," *Environ. Sci. Technol.*, vol. 41, no. 19, pp. 6748–6754, 2007.

[16] J. Gao, H. Tian, K. Cheng, L. Lu, M. Zheng, S. Wang, J. Hao, K. Wang, S. Hua, C. Zhu, and Y. Wang, "The variation of chemical characteristics of $PM_{2.5}$ and PM_{10} and formation causes during two haze pollution events in urban Beijing, China," *Atmos. Environ.*, vol. 107, pp. 1–8, Apr. 2015.

[17] M. Koçak, N. Mihalopoulos, and N. Kubilay, "Contributions of natural sources to high PM_{10} and $PM_{2.5}$ events in the eastern mediterranean," *Atmos. Environ.*, vol. 41, no. 18, pp. 3806–3818, Jun. 2007.

[18] OASIS. *Message Queuing Telemetry Transport (MQTT) Version 3.1.1*. Accessed: Dec. 30, 2018. [Online]. Available: http://docs.oasis-open.org/mqtt/mqtt/v3.1.1/mqtt-v3.1.1.html

[19] J.-H. Park, H.-S. Kim, and W.-T. Kim, "DM-MQTT: An efficient MQTT based on SDN multicast for massive IoT communications," *Sensors*, vol. 18, no. 9, p. 3071, Sep. 2018.

[20] A. Larmo, A. Ratilainen, and J. Saarinen, "Impact of CoAP and MQTT on NB-IoT system performance," *Sensors*, vol. 19, no. 1, p. 7, Dec. 2018.

[21] Ó. Belmonte-Fernández, A. Puertas-Cabedo, J. Torres-Sospedra, R. Montoliu-Colás, and S. Trilles-Oliver, "An indoor positioning system based on wearables for ambient-assisted living," *Sensors*, vol. 17, no. 12, p. 36, Dec. 2016.

[22] M. Alvarez-Campana, G. López, E. Vázquez, V. Villagrá, and J. Berrocal, "Smart CEI Moncloa: An IoT-based platform for people flow and environmental monitoring on a Smart University Campus," *Sensors*, vol. 17, no. 12, p. 2856, Dec. 2017.

[23] P. Bellavista, C. Giannelli, and R. Zamagna, "The PeRvasive environment sensing and sharing solution," *Sustainability*, vol. 9, no. 4, p. 585, Apr. 2017.

[24] J. Alonso, C. Bayona, O. Rojas, M. Teran, J. Aranda, H. Carrillo, and C. Parra, "IoT solution for data sensing in a smart campus using smartphone sensors," in *Proc. IEEE Colombian Conf. Commun. Comput. (COLCOM)*, May 2018, pp. 1–6.

[25] S. Mahajan, W.-L. Wu, T.-C. Tsai, and L.-J. Chen, "Design and implementation of IoT-enabled personal air quality assistant on instant messenger," in *Proc. 10th Int. Conf. Manage. Digit. EcoSyst. (MEDES)*, 2018, pp. 165–170.

[26] G. Jin-Hui, K. Han-Qing, Z. Liang, W. Hong-Lei, Y. Peng, Z. Bin, and L. Yue-E, "A numerical simulate study of the pollution incident of the pm2.5 pollutant island in the summer of nanjing," *China Environ. Sci.* no. 2, pp. 321–330, 2016.

[27] X. Li, Y. J. Feng, and H. Y. Liang, "The impact of meteorological factors on $PM_{2.5}$ variations in Hong Kong," *IOP Conf. Ser., Earth Environ. Sci.*, vol. 78, Jul. 2017, Art. no. 012003.

[28] Taichung City Open Data. *List of Temples in Taichung City*. Accessed: Dec. 30, 2018. [Online]. Available: https://opendata.taichung.gov.tw/dataset/3ac4ca19-1a9f-11e8-8f43-00155d021202

[29] J. K. A. Macêdo, J. K. Joseph, J. Menon, T. Escalante, A. Rucavado, J. M. Gutiérrez, and J. W. Fox, "Proteomic analysis of human blister fluids following envenomation by three snake species in India: Differential markers for venom mechanisms of action," *Toxins*, vol. 11, no. 5, p. 246, Apr. 2019. [Online]. Available: http://www.mdpi.com/2072-6651/11/5/246

[30] Y. Gan, Y. Xiao, S. Wang, H. Guo, M. Liu, Z. Wang, and Y. Wang, "Protein-based fingerprint analysis for the identification of Ranae Oviductus Using RP-HPLC," *Molecules*, vol. 24, no. 9, p. 1687, Apr. 2019. [Online]. Available: http://www.mdpi.com/1420-3049/24/9/1687

[31] C. Santos, R. B. Neto, R. Da Silva, and S. Costa, "Cluster analysis applied to spatiotemporal variability of monthly precipitation over paraíba state using tropical rainfall measuring mission (TRMM) data," *Remote Sens.*, vol. 11, no. 6, p. 637, Mar. 2019. [Online]. Available: http://www.mdpi.com/2072-4292/11/6/637

[32] W. Hao, Z. Hao, F. Yuan, Q. Ju, and J. Hao, "Regional frequency analysis of precipitation extremes and its spatio-temporal patterns in the Hanjiang River Basin, China," *Atmosphere*, vol. 10, no. 3, p. 130, Mar. 2019. [Online]. Available: http://www.mdpi.com/2073-4433/10/3/130

[33] J. Iqbal, "Managerial self-attribution bias and banks' future performance: Evidence from emerging economies," *J. Risk Financial Manage.*, vol. 12, no. 2, p. 73, Apr. 2019. [Online]. Available: http://www.mdpi.com/1911-8074/12/2/73

[34] J. H. Ward, "Hierarchical grouping to optimize an objective function," *J. Amer. Stat. Assoc.*, vol. 58, no. 301, pp. 236–244, Mar. 1963.

[35] F. Murtagh and P. Legendre, "Ward's hierarchical agglomerative clustering method: Which algorithms implement Ward's criterion?" *J. Classification*, vol. 31, no. 3, pp. 274–295, Oct. 2014.

[36] L. Howard, *The Climate of London: Deduced From Meteorological Observations, Made at Different Places in the Neighbourhood of the Metropolis*, vol. 1, W. Phillips and J. A. Arch, Eds. London, U.K., 1818.

[37] A. C. Stern, Ed. *Air Pollution*, vols. 1–3. New York, NY, USA, 1968.

[38] P. Leighton, *Photochemistry of Air Pollution*. Amsterdam, The Netherlands: Elsevier, 2012.

[39] A. J. Haagen-Smit, "The air pollution problem in Los Angeles," *Eng. Sci.*, vol. 14, no. 3, pp. 7–13, 1950.

[40] N. Okawada, I. Mizoguchi, and T. Ishiguro, "Effects of photochemical air pollution on the human eye-concerning eye irritation, tear lysozyme and tear pH," *Nagoya J. Med. Sci.*, vol. 41, nos. 1–4, pp. 9–20, 1979.

[41] B. Wang, R. Wu, and X. Fu, "Pacific–East Asian teleconnection: How does ENSO affect East Asian climate?" *J. Climate*, vol. 13, no. 9, pp. 1517–1536, May 2000.

[42] N.-C. Lau and M. J. Nath, "ENSO modulation of the interannual and intraseasonal variability of the East Asian monsoon—A model study," *J. Climate*, vol. 19, no. 18, pp. 4508–4530, Sep. 2006.

[43] S. Zhao, H. Zhang, and B. Xie, "The effects of El Niño–Southern Oscillation on the winter haze pollution of China," *Atmos. Chem. Phys.*, vol. 18, no. 3, pp. 1863–1877, Feb. 2018.

[44] Z.-J. Li. (2015). *Variation of Ambient Environment in Association With el Nino/la Nina in Taiwan*. Accessed: May 19, 2019. [Online]. Available: https://hdl.handle.net/11296/mratfj

[45] S.-H. Wang, W.-T. Hung, S.-C. Chang, and M.-C. Yen, "Transport characteristics of Chinese haze over Northern Taiwan in winter, 2005–2014," *Atmos. Environ.*, vol. 126, pp. 76–86, Feb. 2016.

[46] Real-Time Air Monitoring for MAZU Festivals. *Taiwan Environmental Protection Administration Executive Yuan (EPA)*. Accessed: Dec. 30, 2018. [Online]. Available: http://www.mazu.airq.org.tw/

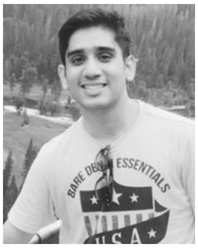
[47] O. Saenz-de-Miera and J. Rosselló, "Modeling tourism impacts on air pollution: The case study of PM_{10} in Mallorca," *Tourism Manage.*, vol. 40, pp. 273–281, Feb. 2014.

[48] A. Zhang, L. Zhong, Y. Xu, H. Wang, and L. Dang, "Tourists' perception of haze pollution and the potential impacts on travel: Reshaping the features of tourism seasonality in Beijing, China," *Sustainability*, vol. 7, no. 3, pp. 2397–2414, Feb. 2015.

[49] Taiwan Central Weather Bureau (CWB). *The Authoritativeand Major Information Weather Resources in Taiwan*. Accessed: Dec. 25, 2018. [Online]. Available: https://www.cwb.gov.tw/V8/C/

NAI-CIH LIOU received the B.S. degree in statistics from National Chengchi University, Taiwan, in 2018. She is currently a full-time Research Assistant with the Institute of Information Science, Academia, Taipei, Taiwan. Her research interests include machine learning, data science, statistics, and the Internet of Things.

CYUAN-HENG LUO received the B.Tech. degree in computer science from National Chengchi University, Taiwan, in 2016. She is currently a full-time Research Assistant with the Network Research Laboratory, Academia, Sinica, Taiwan. Her research interests include user experience, data science, data visualization, and the Internet of Things.



SACHIT MAHAJAN received the B.Tech. degree in ECE from Punjab Technical University, India, in 2012, the M.S. degree in communication engineering from The University of Manchester, U.K., in 2013, and the Ph.D. degree in social networks and human centered computing from National Chengchi University, Taiwan, in 2019. He is currently working as a Research Associate with the University of Cambridge, U.K. His research interests include machine learning, the Internet of Things, and Citizen Science.

LING-JYH CHEN (Senior Member, IEEE) received the B.Ed. degree in information and computer education from National Taiwan Normal University, in 1998, and the M.S. and Ph.D. degrees in computer science from the University of California at Los Angeles, Los Angeles, in 2002 and 2005, respectively. His research interests include wireless networks, mobile computing, network measurements, and social computing.