Revealing Nighttime Construction-related Activities from a Distributed Air Quality Sensor Network

A Thesis

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Master of Science

by

Jintao Gu

August, 2022

ABSTRACT

Coarse particulate matter (PM_c) refers to aerosol particles between 2.5 and 10 μm in diameter. Exposure to ambient PM_c has been associated with adverse health effects such as cardiovascular diseases and respiratory mortality. In this study, we analyzed the spatial and temporal patterns of PM_c levels in a 165-node PM monitoring network in Xi'an, China. We employed a technique called network analysis, focusing on peer-to-peer comparison within the network. The network analysis revealed that the highest PM_c concentrations in the city occurred during late night and early morning. Through further analysis using satellite-based aerial imagery and data mining of internet resources, we confirmed with high confidence that the construction-related emission sources, both at the construction sites and from traffic transporting construction materials and debris, are a key contributor. It could be found that both local policies and construction practices incentivized construction contractors to implement earthwork at nighttime, leading to distinct peak PM_c concentrations from late night to early morning, which often triggered both noise and air pollution complaints from residents. Our work demonstrated the potential of utilizing air quality monitoring networks for construction-related environmental monitoring and enforcement. Based on our findings, we also recommend that policymakers re-assess construction-related policies by considering the trade-offs among efficiency, safety, air quality, and noise.

BIOGRAPHICAL SKETCH

Born in Taigu, a beautiful small city in Shanxi, China, Jintao Gu went to HUST in Wuhan for college and then joined EERL as an engineering student at Cornell in 2019.

For Dad, Mom, and my Grandparents

ACKNOWLEDGMENTS

Foremost, I would like to express my deepest gratitude to my committee: my advisor, Professor K. Max. Zhang, for his endless support and guidance throughout my whole study life at Cornell, for his tremendous patience, infinite enthusiasm, immense knowledge, and incredible insight in the field we are working. Professor Shefford P. Baker, for his great kindness, super-warm suggestions, and many, many very valuable comments.

I am also grateful to the current or previous members of the EERL group, including Dr. Jiajun Gu, Bo Yuan, Dr. Bo Yang, Dr. Khaled N. Hashad, Dr. Jeffrey A. Sward, Dr. Zachary E. Lee, Alfredo A. Rodriguez, Ali Amadeh, Henry J. Williams, Haomiao Wang, Hongli Tao, Shiyan Yang; Prof. Shaojun Zhang from the School of Environment at Tsinghua University and many other people. They really helped me a lot both on research and life!

Thanks should also go to my friends for their friendship and emotional support. I was very lucky to know you all (too many to list here but you know who you are!).

Finally, I want to say thank you to my dad, mom, and grandparents. I could not have undertaken this journey without your love.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. Introduction

Coarse particles (PM_c), which refer to aerosol particles between 2.5 and 10 μm in aerodynamic diameter, are mainly produced by mechanical processes such as crushing, grinding, and abrasion of surfaces and are emitted from multiple sources, including natural sources (wind-blown soil, sea spray, pollens), non-exhaust vehicular emissions (tire wear, brake wear, road dust), and construction and quarrying activities [1]–[3]. Short-term and long-term PM_c exposure has been associated with multiple adverse health effects, including increased morbidity and hospital admissions for cardiovascular and respiratory diseases, and these impacts have been proved to be independent of $PM_{2.5}$ (aerosol particles less than 2.5 μm in aerodynamic diameter) in multiple studies[4]–[8].

Despite strong human health implications from exposure to PM_c, the understanding of spatiotemporal variations of PM_c in the urban areas, which is crucial to exposure assessment, is limited by several factors. First, as PM_c has relatively short residence time, its spatial and temporal variations are expected to show greater heterogeneity than $PM_{2.5}$, thus requiring denser monitoring networks to resolve the variations. However, air quality monitoring networks for PM_c (usually derived from the difference between PM_{10} and $PM_{2.5}$) are sparse [4] because many PM networks do not include PM_{10} for routine monitoring purposes. A number of research studies on PM_c have conducted detailed chemical speciation, but samples were usually connected on one or two sites in large metropolitan areas [9]–[14]. Second, characterizations of PM_c sources are inadequate.

For example, traffic-related, non-exhaust sources and construction activities are the two major PM_c local sources in urban areas. Traffic-related, non-exhaust sources [15]–[18] have received growing attention recently. Brake and tire dynamometers have been employed to characterize non-exhaust PM emissions in laboratory settings[19]. But real-world measurements are still lacking. For construction activities, it is challenging to directly track individual construction sites. As such, the National Emission Inventory (NEI) heavily relies on surrogate data and approximations to estimate fugitive PM emissions from construction sites at the county-level. Third, the current air quality modeling method is inadequate of capturing the spatiotemporal variations of PM_c effectively. The Community Multi-scale Air Quality (CMAQ) modeling system, widely adopted for regional-scale air quality modeling, underpredicted PM_c at 49 out of 51 monitoring sites in the western U.S. and failed to reproduce their spatial patterns[20]. Studies using land use regression (LUR) and other data-driven techniques to model PM_c are rare and generally show much lower performance compared to other pollutants [21]–[23].

To advance our understanding of spatiotemporal variations of PM_c in the urban areas, we applied innovative analytical techniques to a dense, distributed air quality monitoring network that employs Federal Equivalent Method (FEM) equipment to monitor both $PM_{2.5}$ and PM_{10}. Our main analytical tool is named Network Analysis, serving as a screening tool to process data from distributed air quality monitoring networks effectively and generate insights about related emission sources for PM_c. Specifically, Network Analysis integrates two complementary techniques, i.e., intra- / inter-ranking

and time series clustering. Intra- and inter-ranking, to be elaborated in Section 2.2, take advantage of peer-to-peer comparisons within the network to differentiate influences exerted by the regional events and local sources and identify the sites heavily influenced by local sources for further analysis. Time series clustering [24], instead of traditionally using mean diurnal pattern to provide an analysis of diurnal properties of target pollutants[25]–[27], captures day-to-day temporal-spatial differences for PM_c, providing the typical diurnal behavior for the whole region and specific sites for further analysis.

The distributed air quality monitoring network we utilized was deployed in Xi'an, China. Though a significant number of publications have focused on $PM_{2.5}$ and PM_{10} issues in Xi'an city due to its severe pollution and large population[28], [29], [38]–[41], [30]–[37], the PM_c issue in this city has never been explicitly studied. By analyzing the sub-city level hourly PM_c concentration data over two months in the urban area of Xi'an, China, where more than 9.3 million people reside, the Network Analysis reveals that construction-related emission sources led to exceedingly high nighttime PM_c concentrations in the city. Based on our findings, we investigated the policy drivers of the construction practices and recommended policymakers reassess the construction-related traffic policies by considering the trade-offs between safety, air quality, and noise.

2. Materials and methods

2.1. Study domain and dataset

As a major metropolitan area in Northwest China, the municipality of Xi'an, urban and rural combined, has a population of ~13 million, among which the urban population accounts for around 9 million[42], [43].

Our study utilized data from a PM monitoring network in Xi'an, shown in Figure 1. Beta attenuation monitoring (BAM)[44], a Federal Equivalent Method (FEM) approved by the U.S. Environmental Protection Agency (EPA) [45], employs the absorption of beta radiation by solid particles extracted from air flow to detect the PM_{10} (aerosol particles less than 10 μm in diameter) and $PM_{2.5}$. In this network, more than 165 BAM based air quality monitoring stations were installed and each of them was responsible for monitoring the $PM_{2.5}$ and PM_{10} concentration levels in a specific community (i.e., the smallest administrative unit) with a specific ID in format "0000C". Hourly air-quality data were collected in July 2019, December 2019, April 2020, and May 2020. PM_c concentration data were derived from PM_{10} and $PM_{2.5}$. For each month, more than 95% of the monitoring sites have more than 650 hourly records.

Xi'an's densely populated urban areas with a square of ~912 km^2, including 56 orange monitoring sites also shown in Figure 1, were selected as our research target area due to its high population density, high density of monitoring sites, and relatively high PM_c concentration level. As PM_c/PM_{10} ratios were much higher in April-May 2020 than in the other two months, indicating PM_c dominates the PM pollution in this period, we selected April-May 2020 as our target period. Furthermore, in winter, coal combustion

for residential heating and potential biomass burning dominates the PM emissions for the whole region[40], and local construction activities face ban under many situations, leading to two difference on main emission source compared with spring and making it not easy to extract information using peer-to-peer comparison. July's $PM_{2.5}$ and PM_{10} level was quite low, making uncertainties associated with deriving PM_c (i.e., PM_{10}-$PM_{2.5}$) high in July. Therefore, July 2019 and December 2019 are not discussed in this paper. For information on data statistical analysis and processing, see S1.

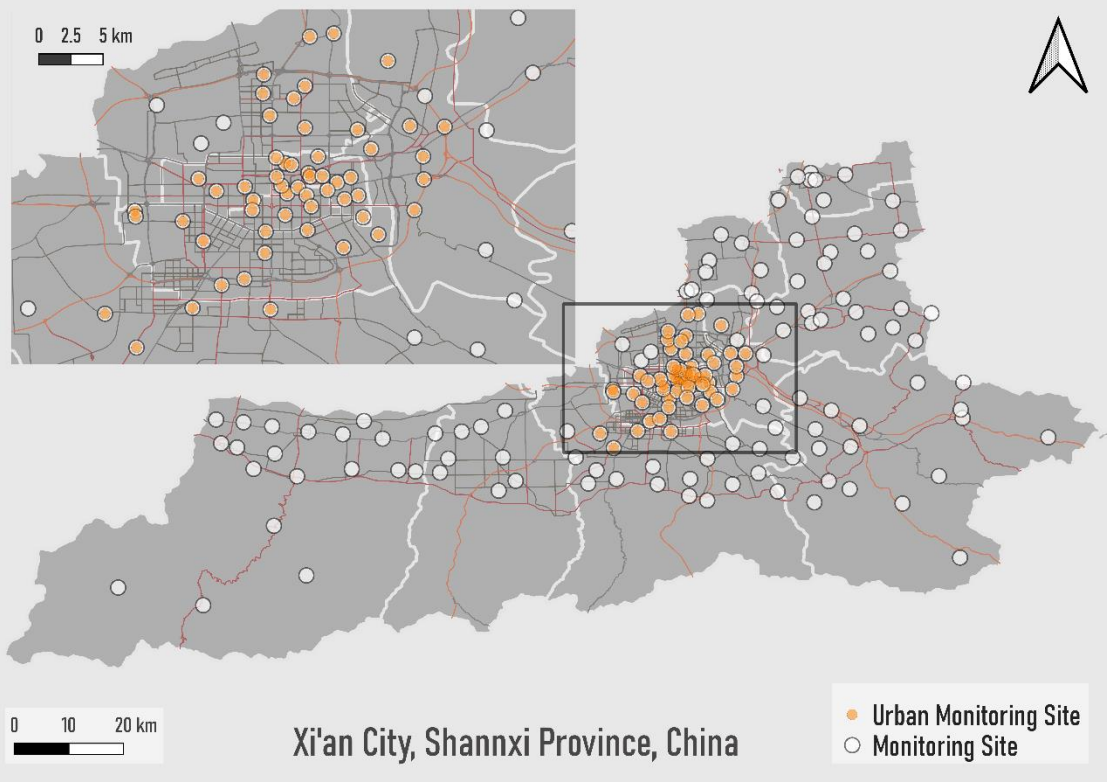


Figure 1: The air quality monitoring network in Xi'an city. Monitoring sites located in the main urban area are defined as urban monitoring sites and are selected as targets (orange ones).

2.2. **Network analysis**

The main goal of network analysis is to identify both areas and emission sources of interest within the network for further investigation. More specifically, the local anomalous concentrations and the associated local drivers are identified by comparing specific pollution trends with the general trend. In this study, we applied this tool multiple ways to explore potential local drivers and their impact on PM_c in Xi'an city's urban area.

2.2.1 Intra-ranking and inter-ranking

First, an intra-ranking system is created to identify the potential network-wide phenomena. The daily averaged PM_c concentrations at the same monitoring site ("intra") are ranked to identify periods when a pollution episode happened across the urban network regardless of specific locations.

Secondly, an inter-ranking system is created to identify urban sites influenced by continuous local drivers and provide hyperlocal insight into potential emission sources. We rank the daily averaged PM_c concentrations among all the urban monitoring sites ("inter") after excluding identified periods with regional events from the intra-ranking system. Those sites that always rank among the highest are called *hotspots* and used for further analysis.

Thirdly, the representative diurnal patterns for urban sites are identified by applying time series clustering method to the dataset. The representative clusters for hotspots are generated and compared with the common diurnal patterns for further insight. Combined

with the local environment survey described in Section 2.3, an analysis regarding potential local drivers and their impacts will be conducted.

2.2.2 Time series clustering

Each sensor records a PM_c measurement each hour and the 24 measurements from one sensor from one day are put into a single 24-dimentional time-series data point, $\boldsymbol{x_i}$. We want to identify the different types of behavior possible. To do this we sort the datapoints into cluster in which each of the diurnal time series have similar behavior. We use the K-means algorithm[46], one of the most popular clustering algorithms, which has been widely applied for time series clustering[24], [47]. Assume we have a set of *n* datapoints $\chi(\boldsymbol{x_1}, \boldsymbol{x_2}, \boldsymbol{x_3} \ldots \boldsymbol{x_n})$. The procedure is as follows:

Step 1. We first assume that there are *k* different types of behavior and we generate *k* datapoints which represent the centers $C\ (\boldsymbol{c_1}, \boldsymbol{c_2}, \boldsymbol{c_3} \ldots \boldsymbol{c_k})$ of k different clusters. Each center is a 24-value time series datapoint which we initially populate with random values.

Step 2. We now divide the *n* datapoints into *k* clusters $\mathbb{C}(S_1, S_2, S_3 \ldots S_k)$ based on the centers: We calculate the distance between each datapoint and each of the centers using the Euclidean distance shown in (2-1) and place each datapoint into the cluster represented by the center to which it is the closest.

For two observations $\boldsymbol{c}\ \{a_1, a_2, a_3, \ldots a_n\}$ and $\boldsymbol{x}\ \{b_1, b_2, b_3, \ldots b_n\}$, the Euclidean distance between x and c is defined as:

$$d(\boldsymbol{c}, \boldsymbol{x}) = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \cdots + (a_n + b_n)^2} \quad (2\text{-}1)$$

Step 3. Next, we calculate the actual center $\boldsymbol{c_i}$ of each of the *k* clusters using

$$\boldsymbol{c_i} = \frac{1}{|S_i|} \sum_{\boldsymbol{x} \in S_i} x \quad (2\text{-}2)$$

Step 4: Repeat step 2 and 3 until no there are no more changes for any of the clusters.

We now have a provisional set of *k* possible typical PM_c diurnal variations. To find the actual number of cluster (typical types of behavior) for our data we use the elbow method [48] to do this. We repeat the calculations above for each of a series of k values. For each one, the Within-Cluster-Sum of Squared Errors (WSS) is calculated. Generally, the higher the k is, the lower the WSS is, indicating a better fit for each cluster. Obviously the best fit would occur if the number of clusters *k* was equal to the number of datapoints *n*. However, to the extent that the data fall into a number of distinct types $k < n$, a plot of *k vs.* WSS will show *k* dropping rapidly at first, then leveling off, taking the form of an "elbow," and indicating that, at some point, an increase in *k* won't improve the fit results significantly. The *k* clusters around this point will be selected to represent the typical diurnal behaviors.

In our study, after removing diurnal patterns with missing data and excluding days affected by regional events using intra-ranking analysis, 2041 diurnal patterns are finally selected for clustering using the above algorithm.

2.2.3 Data mining for emission-related features

We evaluate the micro-environment of interest around each monitoring site to assist local driver identification in the network analysis process.

Dust from the road and construction are supposed to be the main contributor to local urban PM_c emissions in China. According to information released by the website of the Ministry of Ecology and Environment of the People's Republic of China[49], roads and construction are the primary sources of dust in Beijing-Tianjin-Hebei and surrounding "2+26" cities in China, accounting for more than 80% of the total dust emissions together, of which 5%-10% are $PM_{2.5}$, suggesting that the roads and construction mainly contribute to local PM_c. Meanwhile, dust is the primary source of PM_c in China: from a large-scale research project[50], it accounts for 54% of PM_{10} emissions and a higher number of PM_c emissions is inferred from its much lower contribution to $PM_{2.5}$ (21%). Compared with construction, the road dust, which is also influenced by construction-related heavy vehicles traffic, is more studied: The contribution rate of road dust to PM_{10} is 36% in Beijing[51], 44.6% (2010) and 19.2% (2015) in Hangzhou, 60% in the Sichuan Province[52] and 79% in Delhi[53], indicating a higher contribution rate to PM_c in those areas. In a recent study conducted in Xiong'an new area, China, 98.49% of fugitive dust originated from road dust and construction dust and accounts for 29.38% of PM_{10} emissions[54]. Thus, we put construction and traffic as the focus of our micro-environment survey due to their potentially significant contribution rate to PM_c.

As shown in Table 1, construction and traffic-related sources are considered as focus, and their related features are generated and used in Network Analysis to help us understand the relationship between diurnal behavior and emission source. GIS tools are combined with Google, Baidu, OSM and other GIS-based documents to collect the data for each site. Hourly traffic is sampled from the selected period onweekdays and weekends for roads in Xi'an, the detailed information is shown in S2. The construction site-related information requires more data mining, and its detailed process is shown in Section 3.4.

Table 1. Traffic and construction PM_c emission source and related features. In comparable northern cities in China, traffic and construction account for more than 80% of dust emissions in urban areas, while dust is the primary contributor to PM_c.

Potential Emission-source	Examples	Related features
Non-exhaust vehicle emissions	Tire wear, brake wear and road dust generated by light vehicles and public transportation	Roadside(Y/N*), distance to nearby road, traffic flow count, unpaved road (Y/N)
Construction	Dust from construction and demolition	Construction nearby (Y/N); the type, scale, stage, and status of related construction sites; unpaved construction area (Y/N)
	Excess non-exhaust vehicle emissions generated by construction-related heavy	

	vehicles
Soil or other construction-related materials tracked onto roads by trucks related to construction sites	

* Y/N means yes or no

3. Results and discussion

3.1. Intra-ranking system analysis: network-wide episode driven by regional events

We first look at the intra-ranking of daily average PM_c concentrations to explore the network-wide trend of concentrations. As shown in Figure 2, multiple episodes, regardless of locations, were observed during the target period, indicating the existence of PM_c regional drivers. For example, dust storms, which often result from the natural dust transported from outside the city under strong wind, happened during May 17^{th}-18^{th}, when an apparent episode could be observed. Regional events like rain could lead to a low PM_c concentration across the network. This is evident in the case of April 18^{th}-19^{th}. After a rain, all the urban sites dropped to a relatively lower PM_c concentration level

compared with themselves in the selected two months, which format a deep blue (ranking<20%) area in Figure 2.

Since the main goal of network analysis is to identify the local drivers for PM_c, we need to exclude the days influenced by regional events. Days with extreme low (80% of sites are below 25% in a single day) or high PM_c concentration across the whole network (80% of sites are above 75% in a single day) was identified by the intra-ranking system and are excluded for further inter-ranking analysis and time series clustering if a record about regional events such as a dust storm or rain could be found. Fifteen days were excluded.

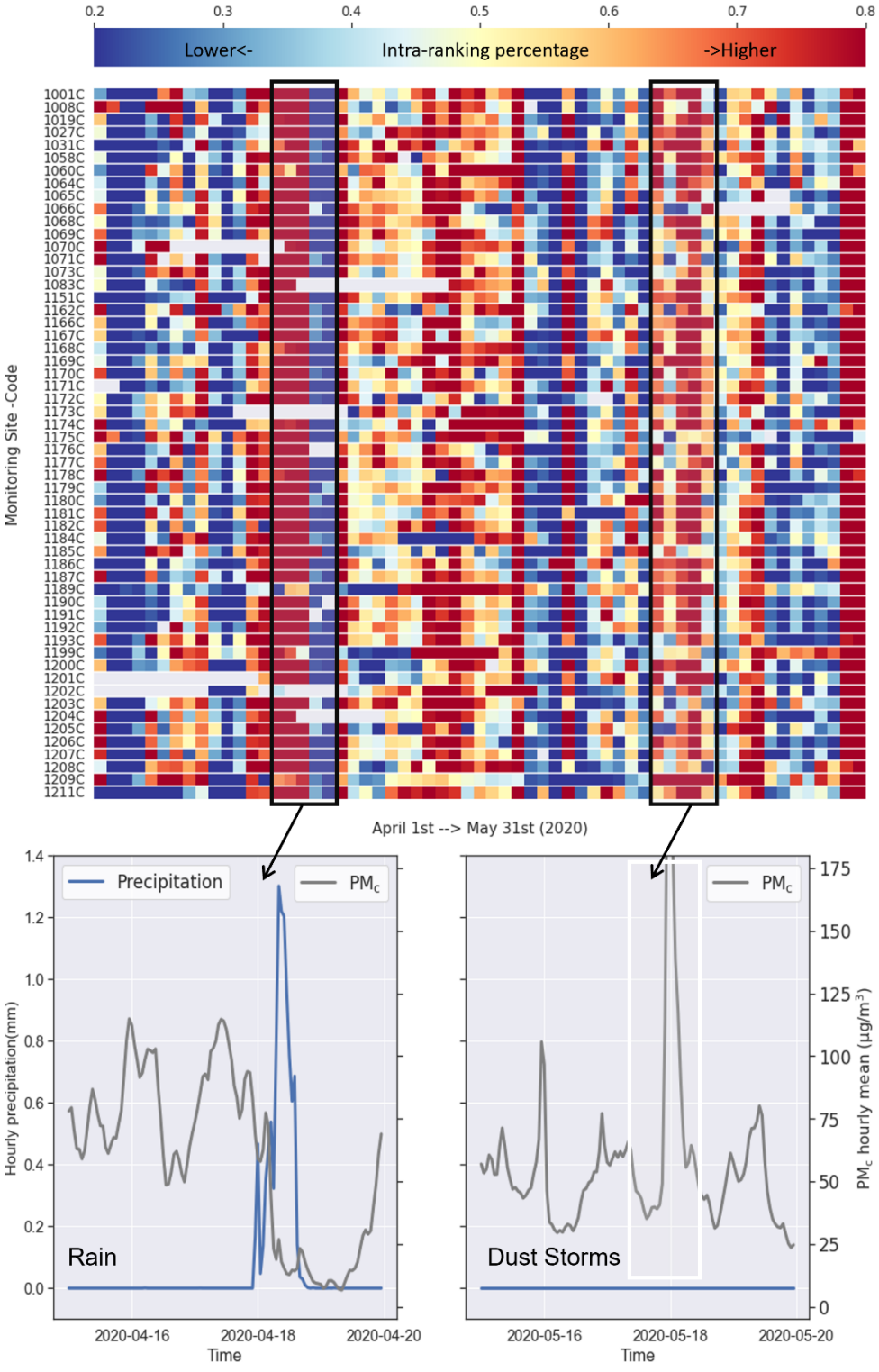

0.2
0.3
0.4
0.5
0.6
0.7
0.8
Lower<-
Intra-ranking percentage
->Higher
Monitoring Site -Code
1001C
1008C
1019C
1027C
1031C
1058C
1060C
1064C
1065C
1066C
1068C
1069C
1070C
1071C
1073C
1083C
1151C
1162C
1166C
1167C
1168C
1169C
1170C
1171C
1172C
1173C
1174C
1175C
1176C
1177C
1178C
1179C
1180C
1181C
1182C
1184C
1185C
1186C
1187C
1189C
1190C
1191C
1192C
1193C
1199C
1200C
1201C
1202C
1203C
1204C
1205C
1206C
1207C
1208C
1209C
1211C
April 1st --> May 31st (2020)
Precipitation
PM$_c$
Hourly precipitation(mm)
Rain
2020-04-16
2020-04-18
2020-04-20
Time
PM$_c$
PM$_c$ hourly mean (µg/m^3)
Dust Storms
2020-05-16
2020-05-18
2020-05-20
Time

Figure 2: Heatmap of intra-ranking system results. Episodes, regardless of locations, could be observed, indicating the existence of regional events such as rain and dust storms shown in two cases[55]. Days with regional event records are excluded for further analysis.

3.2.**Inter-ranking system analysis: hotspots identification**

After excluding regional event days, we then obtain the inter-ranking system results to find the monitoring sites that consistently ranked high in concentrations in the network. Sites sorting by high-ranking (i.e., top 10) days fraction (defined as the number of high-ranking days/the number of total days with records) are shown in Figure 3. Thirteen sites have obviously more high-ranking days (13^{th} is 15 days while 14^{th} is 12 days) and are described as *hotspots* in our research.

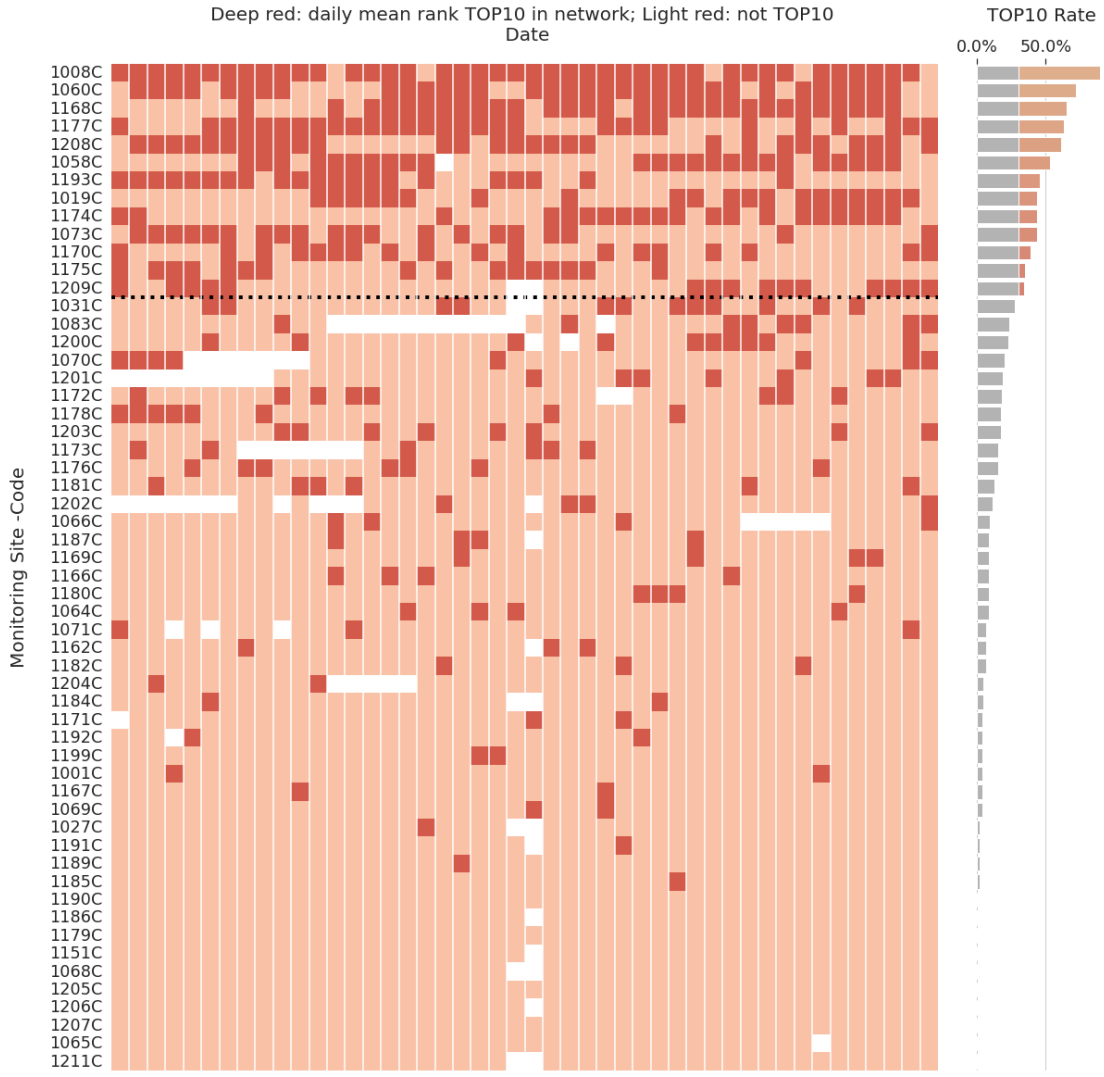


Figure 3: The color of one block in the heatmap represents whether a monitoring site (y-axis) was TOP10 in the inter-ranking system within a single day (x-axis) (red: TOP 10, orange: others, white: missing). Thirteen monitoring sites with consistently high PM_c concentrations are identified as hotspots for further analysis.

3.3. **Time series clustering: results**

It's important to find and understand the representative diurnal behavior for Xi'an's urban area. To achieve this, K-means time series clustering was applied to the urban dataset after excluding regional event days with records. The appropriate cluster number k = 30 was selected based on the elbow method result shown in Figure S2. The clustering results for each cluster are shown in Figure S3. Their temporal-spatial distributions and sizes are shown in Figure S4 and Figure S5, respectively.

Part of the clusters is shown in Figure 4. For a specific cluster, each grey line represents a single day's diurnal patterns from a single site in this cluster, and the red line represents the mean for all the grey lines. We could find that different clusters differ in mean pattern, group size, and temporal-spatial distributions. Information could be extracted from those properties. For clusters composed of diurnal patterns mainly from a single day, such as clusters 24, some unique and undetected regional events might happen in corresponding periods and result in unique PM_c diurnal behavior for a single day across the network. Clusters composed of diurnal patterns mainly from specific sites, such as cluster 15, may indicate common similar local issues for those sites. Clusters of large sizes such as 1 may show the repetitive diurnal behavior caused by common daily events in this region and are worthy of further analysis.

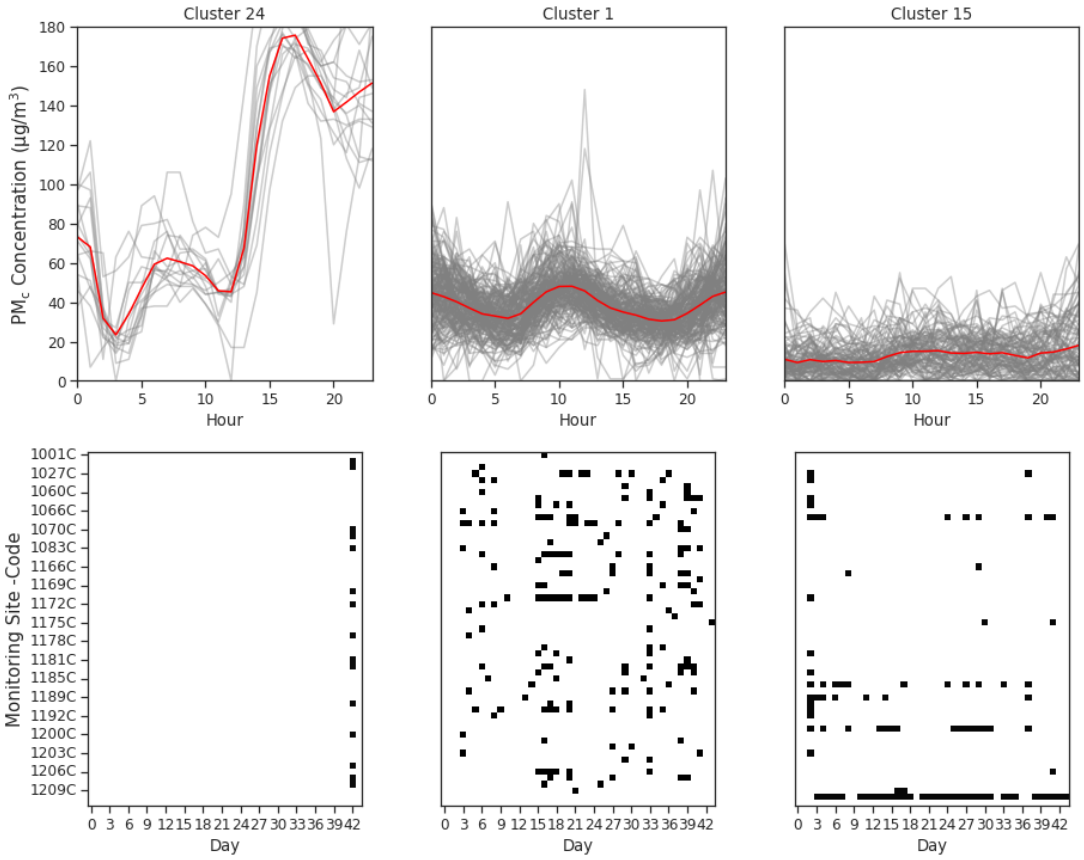


Figure 4: The examples of PM_c diurnal pattern clustering results and their corresponding tempo-spatial distributions. Time series clustering captures the tempo-spatial difference among diurnal patterns: days potentially influenced by regional events will format a single cluster like cluster 24, sites with similar issues can lead to a cluster like 15, and clusters of large size such as cluster 1 could show the general and repetitive diurnal patterns.

3.4.**Time series clustering: traffic led to the general diurnal pattern**

To find what make the hotspots different, it's important to understand what the repetitive general pattern for most sites is, and this could be achieved by focusing on

clusters of large size first. The six largest clusters 25, 1, 13, 8, 2, and 9, with much more data points (one day for one site) than any of the other clusters, occupy 46.3% of diurnal patterns participated in time series clustering, is shown in Figure 5.

It could be found the selected clusters of large size exhibit some similarities in diurnal patterns of variation. Cluster 1, 13 and 25 (25.4% of total counts) shares a very similar mean diurnal pattern: peak at midnight (23:00-0:00), decline to a "valley" in the early morning (4:00-5:00), increase to another peak in the morning (8:00-10:00), decline until late afternoon at 18:00 and then increase to the night peak. Clusters 8, 2, and 9 (20.9% of total counts) differ in some parts: cluster 8's evening peak is much higher than the morning peak. Cluster 2's night peak is delayed at 1:00 and is also much higher than the morning peak. Cluster 9's is an outlier compared with the other five clusters. Its morning PM_c concentration began to increase much earlier at 3:00 and peaked at 6:00. After further analysis shown in S4, the diurnal patterns with both morning and night peaks, represented by clusters 1, 13 and 25, are identified as the general diurnal pattern in Xi'an urban area.

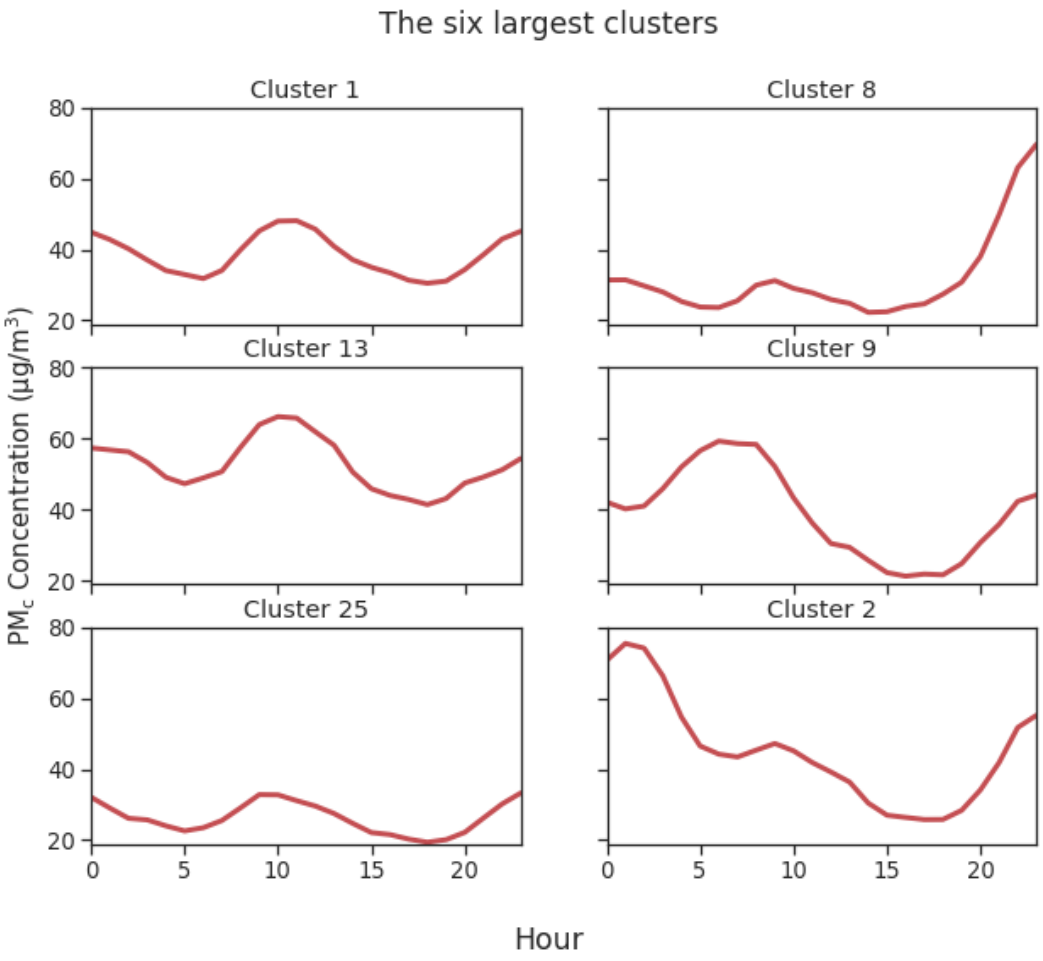


Figure 5: The mean of six largest clusters (number (N) > 125, all the other clusters' N <100). Clusters 1, 13, and 25 show diurnal patterns in a very similar shape with two close peaks at night and morning. Clusters 2 and 8 are different on peak intensity and peak time but still show night and morning peaks within one hour. We call such diurnal patterns with morning peaks at 8:00-10:00 and night peaks at 23:00-1:00 as general patterns due to their repeatability.

A likely explanation of the general pattern is the variance of mixing height and traffic. The traffic generates non-exhaust particle emissions (mainly about PM_c) from the wearing down of brakes, clutches, tires and road surfaces, as well as by the suspension of road dust[56], and the mixing height influences the PM_c concentration level by affecting its dispersion rate. As shown in Figure 6, during the night and early morning, the mixing height went to a very low value, even almost zero, accompanied by a potential surface temperature inversion, making the dispersion of PM_c minimal; during the daytime, especially in the afternoon, the mixing height was relatively higher, accelerating the dispersion of PM_c and making it more challenging to generate high PM_c concentrations even with relatively high traffic count. The explanation matches the general diurnal pattern well if traffic flow data is considered: there were few traffic counts after 0:00, resulting in a decline in PM_c concentration until 5:00. From 6:00, accompanied by a massive rise in traffic flow and the low mixing height, the PM_c concentration increased and peaked at 8:00-10:00. After 11:00, though the traffic remains not low, the PM_c concentration has dropped sharply due to the high mixing height. After 18:00, the mixing height becomes low and the traffic keeps a not low count until midnight, resulting in an increase in PM_c concentrations and a peak at around 0:00. Our explanation is also coherent to previous studies: a very strong correlation between high PM and low mixing height has been reported in multiple urban cases[57]–[59], and this correlation might be even more significant on traffic station[58].

This explanation could also partly explain the formation of untypical general clusters 8 and 2. During weekends, compared with weekdays, the traffic volume remains

relatively high even at 1:00, and the early morning traffic peak becomes smoother, thus resulting in a delayed or/and more substantial elevation at night and a less evident morning peak (see om Figure S9). Since it could be found that clusters 1, 13, and 25 contain more weekdays and clusters 2 and 8 have more weekends (see in Table S7), the time series clustering results match our explanation.

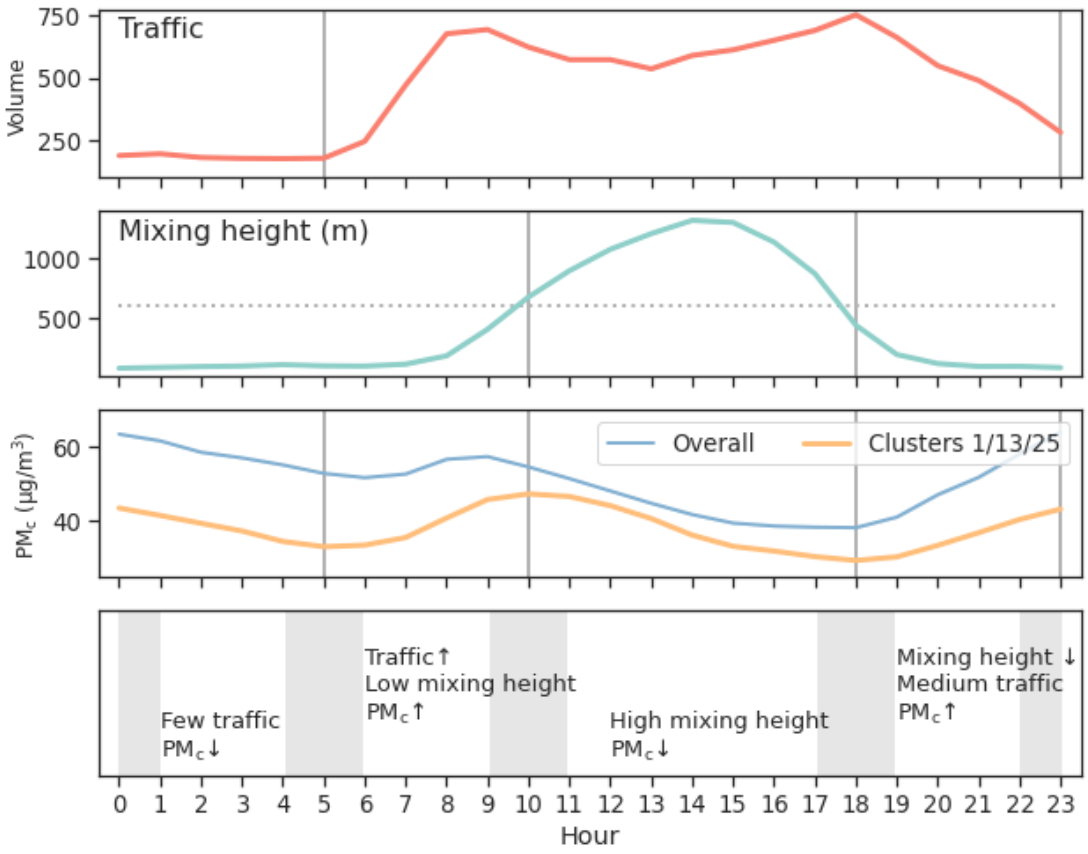


Figure 6: Figure 6: The traffic and mixing height variations lead to the general PM_c diurnal pattern (represented by all sites mean and clusters of largest size 1/13/25 mean). Higher traffic volume means more emission sources, and the lower the mixing height,

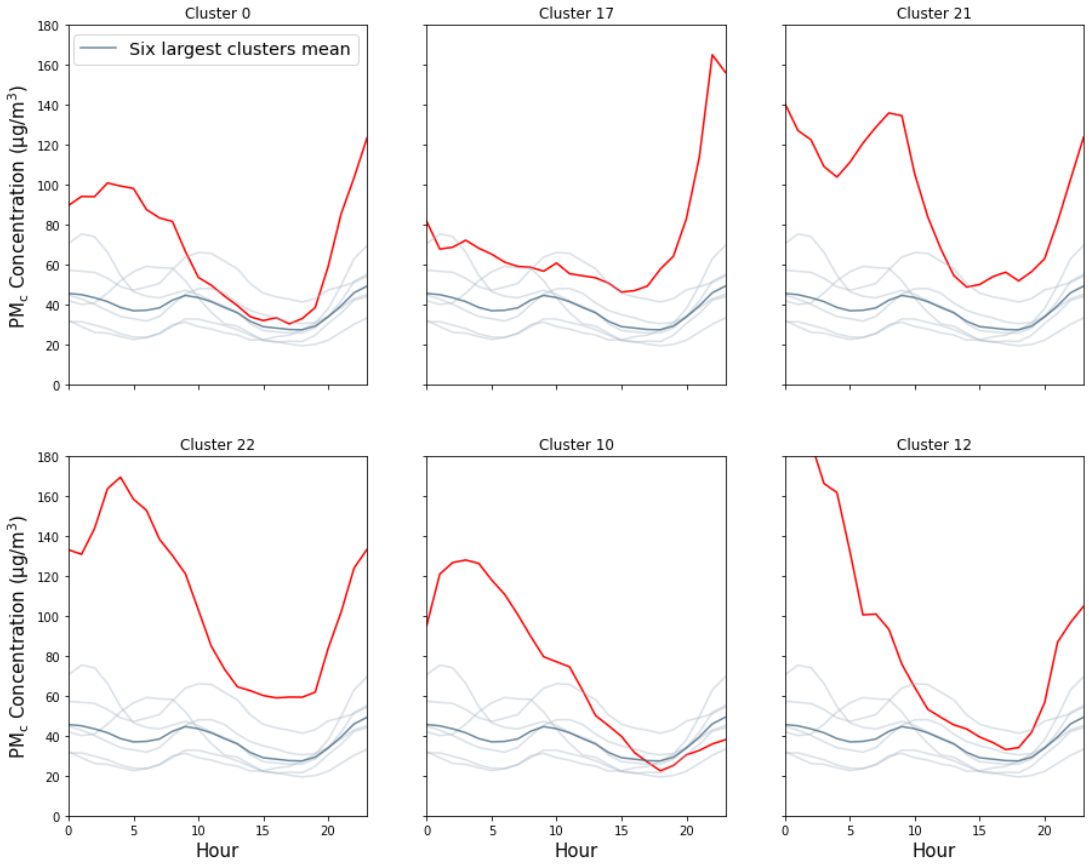
the longer the emissions could stay in the air. The shaded area represents range for peak/" valley" time caused by different traffic situations.

3.5. Representative diurnal patterns for hotspots show abnormal high concentrations at night and early morning

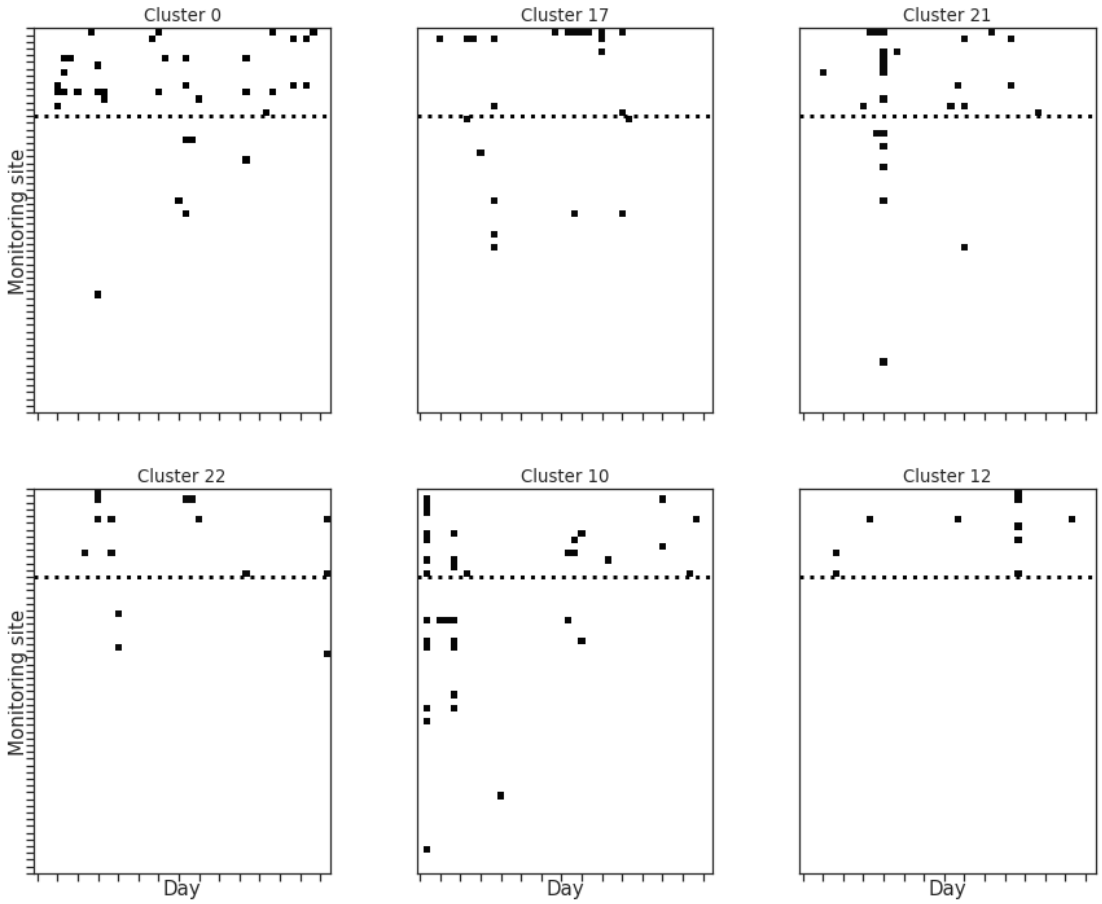
To reveal the mechanism of the highest PM_c concentrations in the urban area, we need to find and understand hotspots' unique and representative diurnal behavior. We used the relative count to find clusters that represent more about hotspots. For a single cluster, the relative count for hotspots is defined as the relative ratio of its number in hotspots and all urban sites. For instance, if 36 diurnal patterns are clustered in cluster 0 and 30 out of those 36 samples are from hotspots, 83.3% (30/36) will be the relative count of cluster 0. Obviously, clusters with a high relative count in hotspots showed diurnal patterns mainly in extremely high pollution areas with significant differences from other clusters and are worth further investigation.

Six clusters with a relative count higher than 50% (the 7^{th} is 45%) were detected as *hotspot feature clusters*. As shown in Figures 7a and 7b, all the hotspot feature clusters show a much higher PM_c mean concentration than the six largest clusters or other clusters shown in S3. An abnormal issue of the hotspots feature clusters can be perceived if we compare them with the general diurnal patterns: most of them, including

Clusters 0, 12, 22, 17, and 10, showed a high PM_c concentration during early morning or night without showing a morning peak. Traffic and mixing height can't explain this: super high PM_c concentration caused by general traffic emissions is not very likely to happen during the early morning because the traffic count is relatively low from 0:00 to 5:00. Furthermore, if the night or early morning peak in these clusters is caused by traffic, an apparent increase at morning (5:00 to 7:00) should happen due to the traffic morning peak. Regional events cannot explain this either: these clusters mainly contain hotspots and don't contain samples mainly from a single day. Thus, further investigation is needed to explain those hotspots feature clusters with super high concentrations.


Cluster 0
Six largest clusters mean
Cluster 17
Cluster 21
Cluster 22
Cluster 10
Cluster 12
PM$_c$ Concentration (µg/m^3)
Hour

(a)



(b)

Figure 7: (a) The feature clusters of hotspots and (b) corresponding temporal-spatial distribution (sites above the dotted line area are hotspots). The six largest clusters shown in Figure 6 areis used as reference in blue color. Except for cluster 21, all clusters are of abnormally high concentration during the night or early morning without showing an obvious morning peak, indicating the potential existence of a strong local driver other than traffic.

3.6. **Emission-related features data mining: Active construction sites detection**

The phenomenon that the representative cluster for hotspots showed an abnormally high concentration during early morning or night without an obvious morning peak is worthy of further exploration of their difference in emission source-related features. One of the essential bases of source-related features, construction site information, is lack of open source and needs further confirmation. Here all active construction sites around urban monitoring sites are identified based on specific criteria by utilizing data mining from multiple open sources: Google Satellite is used as a starting tool to find potential construction sites. Various methods are implemented to explore and confirm the construction project's status, stage, and type.

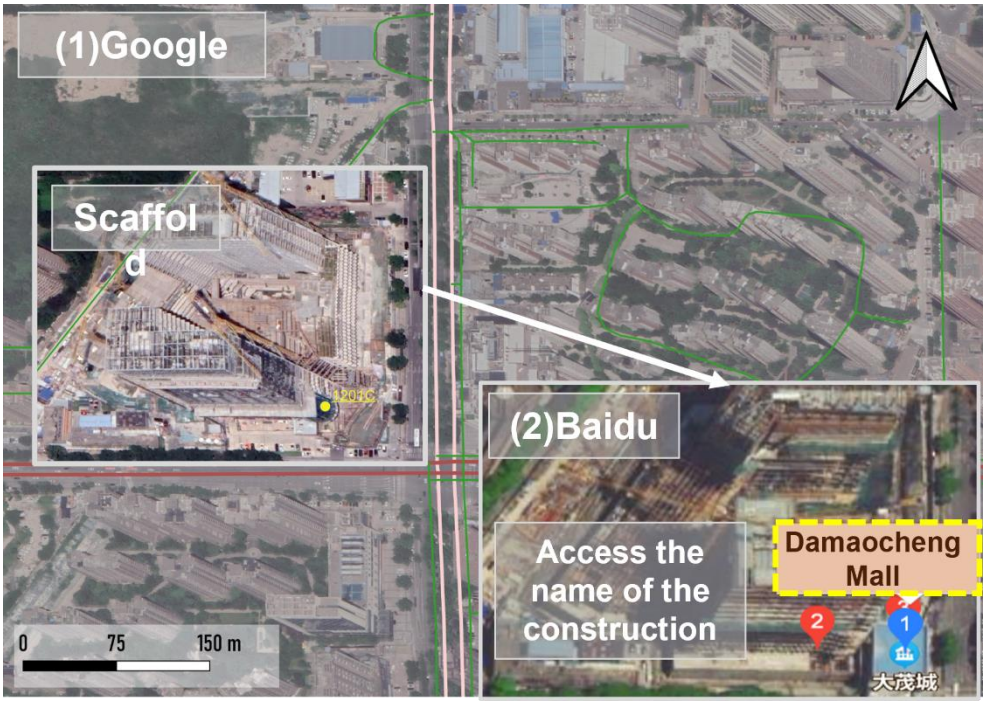
First, potential construction sites are searched and labeled by Google Satellite. A 0.003-degree (~330m) buffer is created for the general search scope working for urban monitoring sites, while a 0.009-degrees (~990m) buffer is used as supplemental search scope working for major-road side urban monitoring sites. Lands with listed features are labeled as potential construction sites waiting for further confirmation: Tower cranes and scaffoldings for unfinished buildings, earthwork traces with construction engineering vehicles, and large areas of unpaved land with demolition traces. It's estimated that the December 2021 version of the Xi'an Google Satellite was shot around January-March of 2021, while the January 2022 version of Xi'an Google Satellite was shot around August 2021.

Secondly, those potential construction sites are confirmed on the active and working status by comparing Baidu Satellites with Google to acquire the Chinese name and data

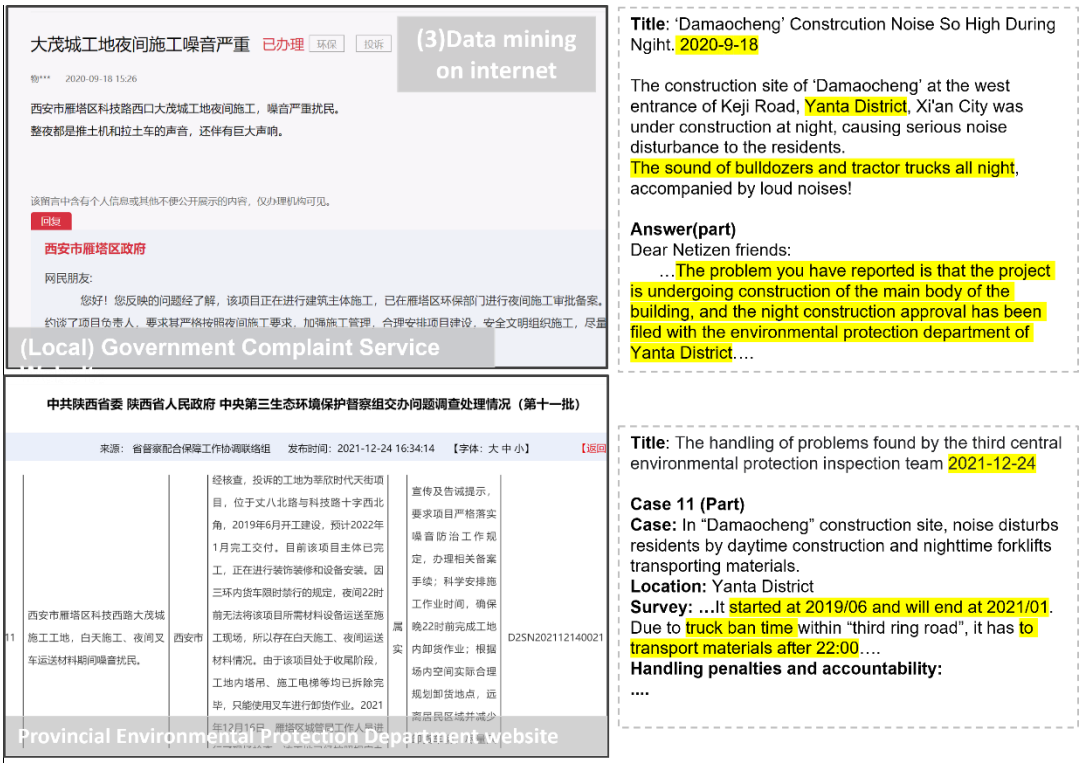
mining utilizing such open information on the internet as real estate websites, local forums, and government websites. 1201C, as shown in Figure 8a, is used as an example to help understand this process. Initially, the construction site with tower cranes and scaffoldings is identified with Google Satellite. Based on image similarity, the same address is then opened on Baidu Satellite. It should be noted the general WGS-84 doesn't work for Baidu and Google maps (not satellites) in China due to the existence of a unique encryption algorithm with hundred meters biases required by law. After acquiring the Chinese name of this potential construction site, search engines will present multiple information regarding the stage and status.

As shown in Figure 8b, on a website operated by the government, people complained about the overnight noise disturbance to the local government in September 2020 and received an official reply regarding that the construction was in the main body construction stage with a commitment to improvement. A case also reported in December 2021 from the provincial environment protection department, mentioned that construction materials had to be transported at night due to a traffic ban. A real estate website showed that a residential building was available for sale starting November 2019. According to local law, the residential building for commodities is permitted for sale after the completion of the sixth floor, which is set to prevent real estate developers from committing fraud in residential building transactions. Thus, we could conclude that this is an active construction site under the main body construction stage with potential overnight construction activities during the target period.

Following process shown in context, nineteen urban sites are identified with active construction sites nearby.



(a)



(b)

Figure 8: Identification of construction site: an example around monitoring site 1201C. Process: (1) Find a potential construction site around 1201C using Google Satellite. (2) Access the construction's name using Baidu Satellite. (3) Conduct data mining on the internet. In this case, construction was active during the target period with overnight construction activities, leading to many complaints from nearby residents.

3.7. **Explore the potential local driver of hotspots**

An analysis of the within-hotspots difference in their clustering results and local emission sources was conducted to explain why the representative clusters for hotspots are abnormal compared with the general pattern. For each hotspot, the abnormal days

rate (the percentage of days clustered into hotspot feature clusters) and emission-related features are shown in Table 2 and Table S5. The terms *construction station* and *non-construction station* are used to refer to the monitoring site with/without an active construction site nearby (~330m).

As Table 2 shown, the monitoring sites with largest feature clusters days rate are more likely to be a construction station (in red color), indicating that a hotspot identified as construction station is more likely to show a diurnal pattern other than general. In other words, the hotspots feature clusters are generated by hotspots identified as construction station rather than other hotspots. For example, the rank#1 station 1008C, shown in S10, is directly located on a large construction site. More satellite-based images are shown in S4.

Table 2. The feature clusters' days rate and emission-related features* for hotspots. Construction hotspots (in red color) are the main contributor to the hotspot feature clusters, indicating construction activities may lead to the abnormally high concentration in the night and early morning of these clusters.

Hotspots ID	Feature clusters' days rate (%) **	Construction site (within 300m)	Mean hourly traffic flow (If roadside)
1008C	37.78%	Y	No records
1060C	35.56%	Y	1569
1208C	31.11%	Y	786
1073C	28.89%	Y	544

1209C	22.22%	Y	
1174C	15.56%		
1193C	13.33%	Y	
1175C	13.33%	***	
1170C	13.33%		973
1177C	8.89%	Y	2579 (Top one) ****
1019C	6.67%		
1058C	6.67%		49
1168C	2.22%		

* For a detailed version, see Table S1

**For a single site, abnormal days rate=number of days clustered into hotspots feature clusters / 45 days

*** Only active during April

**** Among fifty-six sites and their nearest road / road within 100m

Among the common PM_c local drivers, recall that traffic emissions cannot adequately explain the formation of hotspot feature clusters; same as industrial sources based on the local environment analysis: no obvious industrial factories exist around those sites, and most of the factories also have no reason for an opening during the early morning. Thus, consistent with the intuitional results shown in Table 3, the construction site relevant activities, including associated unpaved areas found in many sites and related heavy vehicles emissions, are considered a very likely candidate for the emission source of super high PM_c concentration in the night or early morning found in hotspots feature clusters.

To verify this result intuitively, we plotted the PM_c bi-monthly mean diurnal patterns for hotspots in Figure 9. Consistent with the time series clustering analysis, it could be observed that the mean diurnal patterns for monitoring sites other than hotspots showed a general pattern, while hotspots identified as construction stations showed a more abnormal one: no obvious morning peak with high concentration during the night. Construction station 1177C, a roadside hotspot with an extremely high traffic count shown in Table 3, is the only exception among hotspots identified as construction stations: it showed a morning peak in Figure 8d and ranked low on abnormal days rate in Table 3. As traffic leads to the morning peak based on our analysis structure, this exception is qualitatively consistent with our explanation.

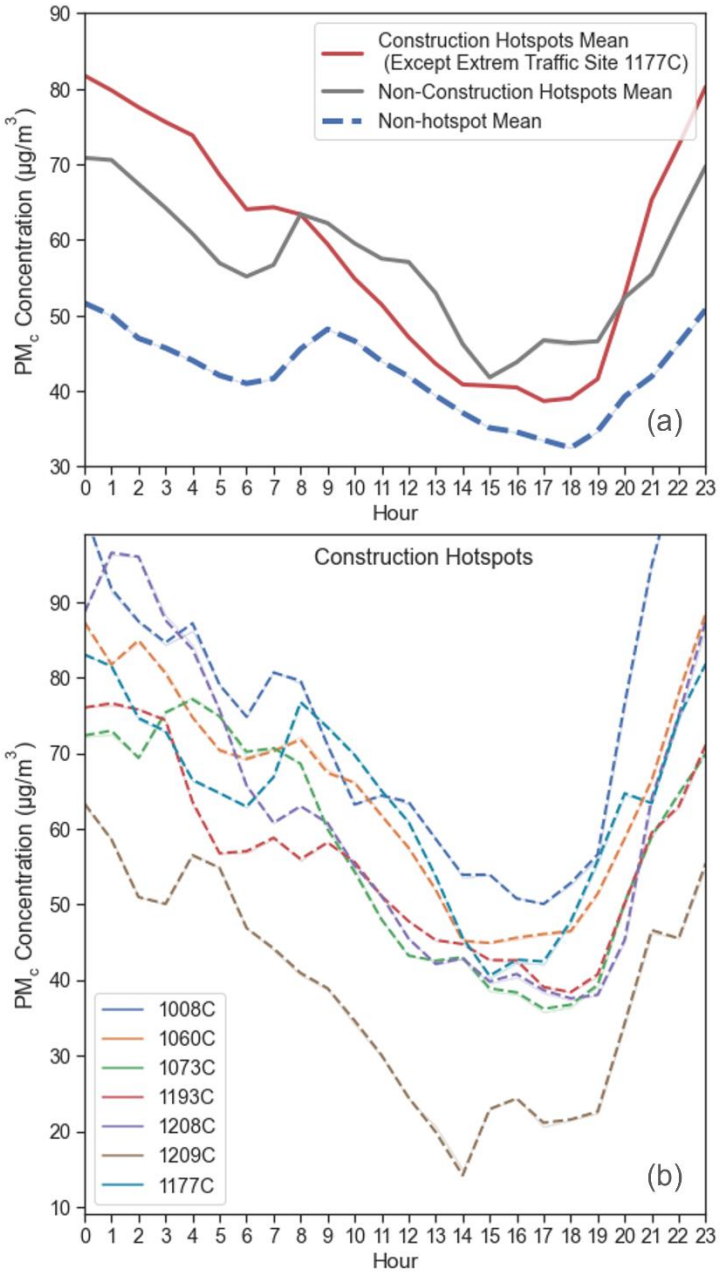


Figure 9 (a). The average PM_c daily diurnal patterns for non-construction hotspots and non-hotspot sites are different from construction hotspots, showing potential impact from construction site activities at night and early morning. (b). The average PM_c daily

diurnal patterns for all single construction hotspots are all of a high concentration during the night without an obvious morning peak except 1177C, an extremely high traffic site, showing the effect of heavy traffic in the morning.

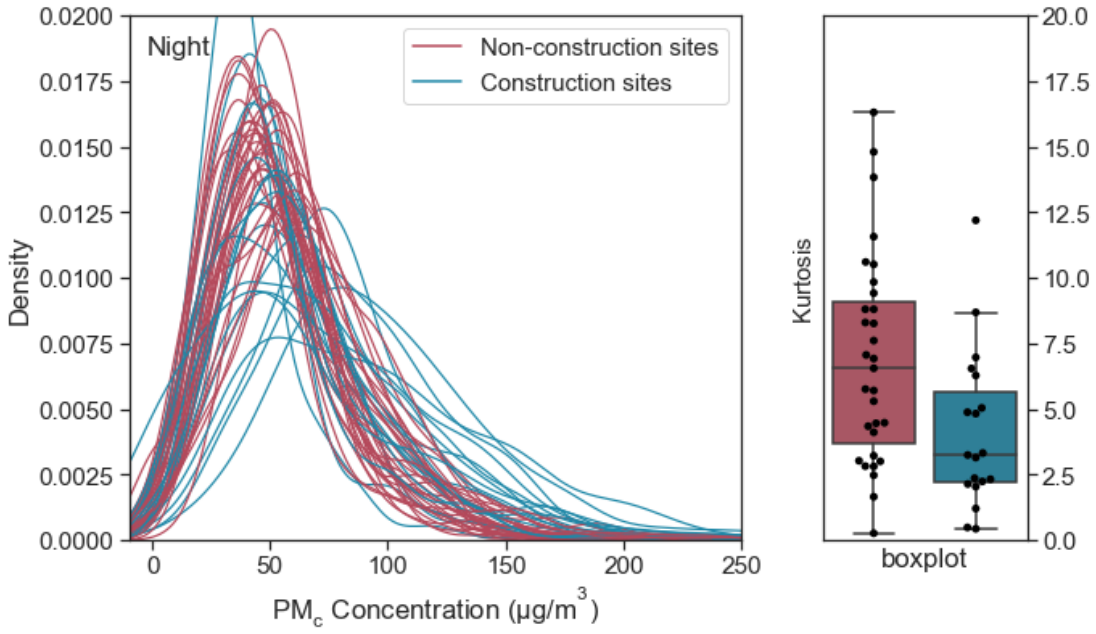
3.8. Construction leads to high concentrations during night and early morning: verification

An intuitive observation based on previous analysis targeting hotspots is that high PM_c concentration of Xi'an city is highly correlated with the construction site activities mainly during night and early morning. We verified this phenomenon statistically for all the urban sites.

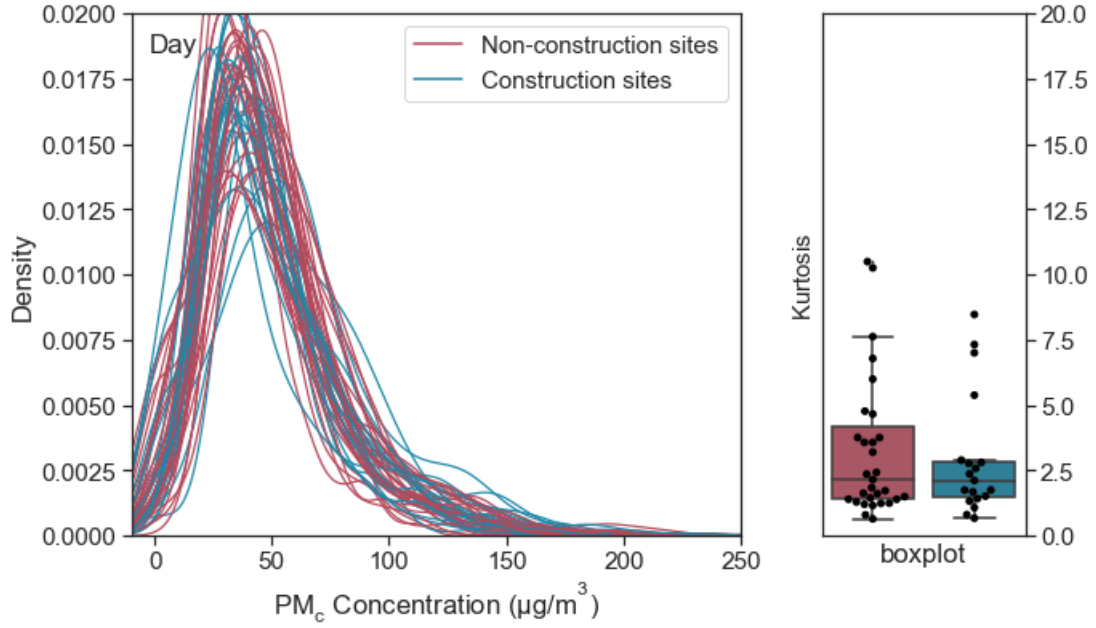
The probability density function of monitoring sites with a mean concentration higher than 10% sites is shown in Figure 10 by dividing hourly data into nighttime (20:00 to 6:00, early morning time included night peak time for general pattern included) and daytime (7:00 to 19:00, morning peak for general pattern included). As shown in Figure 9, during nighttime, construction stations have a heavier right-tail than non-construction stations. This phenomenon is not evident in the daytime, indicating a construction station is more likely to show high PM_c concentration at night.

Kurtosis, a common statistical indicator to measure whether the data are heavy-tailed or light-tailed relative to a normal distribution, was applied to the distribution of each site in daytime and nighttime. We conducted Mann–Whitney U test[60] for kurtosis of each site to check whether construction station's PM_c concentration's distribution are significantly different from others. The results prove our observation: it is statistically significant (p value=0.002916, CI=99%) that the kurtosis of construction stations is

different from non-group during night. We cannot reject the hypothesis that the kurtosis of distributions from these two kinds of air quality monitoring sites is equal during the daytime ($p= 0.4061$). Thus, a construction station is more likely to differ from a non-construction station on PM_c concentration during the night than the daytime.


Night
Non-construction sites
Construction sites
Density
PMc Concentration (μg/m3)
Kurtosis
boxplot

(a)


Day
Non-construction sites
Construction sites
Density
PMc Concentration (μg/m3)
Kurtosis
boxplot

(b)

Figure 10: The probability density function for urban sites and their kurtosis during (a) nighttime and (b) daytime. The heavier tail (by visual inspection) and the lower kurtosis (generally means heavier tail) of construction sites (blue) during the night shows the night effect of construction sites. No apparent differences are found between the construction sites and other sites during the daytime, indicating construction activities impact mainly at night. (Sites with mean concentrations higher than 10% are shown in this image)

3.9. **What led to construction's night and early morning impact in Xi'an**

The factors behind construction site's night impact on PM_c are further analyzed here: both policies and construction practices could incentive construction contractors to implement earthwork overnight and transport construction materials mainly at night, leading to distinct peak PM_c concentrations from late night to early morning.

First, regarded as one of the three most common construction activities related to dust source, the vehicles used for transporting construction materials and rubbish is often operated during the night in many cities, including Xi'an. This is influenced by local policy and other factors, including law enforcement density and traffic flow. In Xi'an city, those vehicles are only permitted during nighttime (22:00~6:30) from 2011 to 2019 inside the third ring road. Since 2019, though additional time windows (10:00-16:00) have been implemented for those vehicles, based on many reported cases, restrictions still exist for those vehicles under many circumstances during the daytime due to . For

example, as shown in Section 3.6, in “Damaocheng” construction, people had to transport construction materials after 22:00 in 2021.

Secondly, as police are less likely to carry out law enforcement during late night and early morning, heavy vehicles used for construction materials and wastes (i.e., debris) transportation might choose to be more non-compliant with environmental regulations during this period: they choose to load more construction waste, do less cover, and less body washes, resulting in cost savings, increased efficiency and excessive PM_c pollution.

Thirdly, overnight operations are also common in construction sites in China. Research on China’s fatal accidents in building construction activities concluded relatively more machinery-related construction activities are still carried out from 22:00 to 8:00[61]. This could also be reflected in many media reports and cases (e.g., the example in Section 3.4), indicating that the construction site activities could happen frequently during the night and early morning.

Last, construction materials or rubbish accumulations lack of cover are common on construction sites. As a potential consequence, even if a specific construction site only operates during the daytime, because of lower mixing height during nighttime, it is still possible for those sites to generate a high PMc concentration from their accumulations of construction rubbish during the night due to the airflow. Furthermore, not allowing rubbish (such as debris) to be transported during the day could make the accumulation larger and thus increase the pollution at night, either

As the above factors show, unlike traffic PM_c emissions, construction site activities related PM_c emission sources are diverse and do not have a fixed time. For example, an active construction site may choose different time windows at night to transport construction rubbish in a period. This is consistent with the hotspots feature clusters: related peaks may occur at different times.

From this, we can know the nature of the heavy-tail distribution of construction stations shown in Section 3.8: the value of high PM_c pollution generated by construction sites is high. It does not have a dense interval like traffic because of the diversity and uncertainty of construction-related pollution sources. This also explains the hotspots feature clusters' difference in peak time reported in Section 3.5: construction sites can choose different time windows from 22:00 to 6:30 to implement overnight earthwork and materials and rubbish transportation.

3.10 Policy implications

Construction work, unpaved land, and heavy vehicles to transport construction materials and rubbish (i.e., debris) are three primary sources of dust (mainly PM_c) generated from construction site activities. Several technical measures have been proposed and partially required by the government, including building construction fencing and safety nets, moisturizing ground, washing vehicles, using chemical dust suppressants, and sprinkling water to suppress dust during construction and demolition. Though many engineering practices have proved the effectiveness of such measures, just

as Xi'an city urban hotspots' high PM_c concentration shown, more discussion is still needed to control PM_c emission further.

One of the primary emission sources worth further discussion is the construction materials and rubbish transportation. Based on previous research, they contribute a lot to total urban PM_c concentrations. For example, based on an analysis of dust sources in cities such as Shenzhen, construction vehicles carrying mud on the road can account for 40% of the local sources of dust. As our research has shown, this kind of pollution has an uneven level distribution during one day under the influence of policy, engineering practice, and metrological features. Thus, it's meaningful to discuss the policy options as well as their impact on PM_c emissions. For example, as we concluded that PM_c pollution spreads more easily during the daytime, a simple and intuitional policy option to reduce PM_c concentration level in Xi'an city is only allowing construction transportation vehicles operated during the daytime, which has already been applied in some cities. However, in policy practice, there is a complex trade-off between safety, construction efficiency, noise as well as dust pollution.

Here we have compiled information on the current or recent time limitation policies of heavy vehicles used for construction materials and rubbish transportation (Chinese name: 渣土车) in some large cities in China. As Table 3 shows, those policy options could be divided into five categories based on the differences in the time limit: 1), Night only. 2), Traffic peak time excluded (Night and some of the daytime). 3), Late night excluded. 4), Daytime only and a mixture of those types in different times or areas.

Table 3: The traffic restriction policies for construction-related transportation vehicles in China. Daytime or late night excluded type policy, which is less likely to cause heavy PM_c pollution, is less implemented.

Type	City	Urban population (million) (2020)	Updated time****	Operating time for urban area*
Night ONLY	Beijing	17.75	2020	23:00-6:00**
	Wuhan	9.95	2017	20:00 (21:00 for summer)-5:00 Only
	Nanjing	7.91	2014	22:00 -7:00
	Fuzhou	5.44	2014	Part: 21:00-7:00 Part: 24h ban
Traffic peak time excluded ***	Jinan	5.88	2019	9:00-17:00, 21:00-6:00
	Xi'an	9.28	2019	10:00-26:00, 22:00-6:00
	Chengdu	13.34	2020	Weekdays: 20:00-7:00, 9:00-17:00
				Weekends: 20:00-17:00
Late night excluded	Changsha	5.55	2020	Part A: 19:30-24:00
				Part B: 9:30-16:30, 19:30-22:00
				Part C: 9:30-22:00
Daytime ONLY	Hangzhou	8.74	2021	*****
Mixture	Zhengzhou	5.34	2019	Urban-center area: 22:00-5:00
				Other urban area: 9:00-20:00
	Luoyang	4.58	2020	October->March: 9:00-17:00
				April->September: 20:00-

				7:00
Other	Shenzhen	17.44	2021	24h ban, permit only

* Generally, it's the center-urban area which often refer to the area inside the third/forth/fifth ring road.

** Policy also works for other heavy vehicles.

*** Limitation during daytime may still exist under multiple circumstance.

**** If no documents, it means year of reporting of latest evidence

***** No direct documents founded but mentioned in multiple official media

Obviously, behind those different policy options are different considerations and local practices of local government. It could be found that the time limit policy for those heavy vehicles is not always the same in a single city. Once they change, some information about the consideration of local government will be reported. Some recorded policy option change and their reported reason are compiled and shown in Table 4.

Table 4: The official reason for local government to change construction transportation vehicles restriction policy. Trade-offs exists. More daytime windows mean less pollution and less noise disturbance, while more nighttime means more safety. safety.

City	Change time	Pre-change time	Post-change time	Reason from official report / Background
Jinan	2019-June	22:00-5:00	9:00-17:00,21:00-6:00	"Alleviate the problems of noise disturbance and air pollution"

Changsha	2020-January	22:00-4:00	Ban after 24:00, details seen in Table 4	"Dust control" "Reduce transportation noise at night" "Prevent traffic violations"
Xi'an	2011-December	No limitation	22:00-6:00	50 deaths in related traffic accidents
	2019-March	22:00-6:00	10:00-16:00, 22:00-6:00	"Enterprise work efficiency" "Dust control"
Nanjing (one of the subareas)	2020-June	22:00 -7:00	Daytime included	"Reduce noise complaints"

One of the typical change types that appeared in recent years is to extend the operation time to daytime even to ban those vehicles in a late-night under the consideration of dust pollution control, noise control, and construction efficiency, represented by Changsha city and Jinan city, which is consistent with our finding in Xi'an. Another typical change type that appeared years ago is to limit those vehicles during the daytime under the consideration of safety. For example, in Xi'an city, when we traced back to 2011, Xi'an city allowed those vehicles from construction sites to operate during the whole day. Then in 2011, those vehicles caused 50 deaths in various accidents in a single year and triggered a storm of public opinion. Consequently, the government published the new policy and banned all such vehicles in the urban-center area in the daytime. Engineering efficiency is also used as a reason for changing the

policy. In 2019, Xi'an city reopened the time window during the day (though the main time window remains at night) and used "improve the enterprise work efficiency" as one of the main reasons.

Undeniably, what is behind these policies are trade-offs of factors, including dust control, safety, engineering efficiency, and noise. To our extent, there is no doubt that the night-only or traffic peak-time-excluded/night-mainly type of policy will generate a tremendous amount of dust and thus increase the PM_c concentration level. This might be controlled by banning such vehicles late at night and allowing them on the road during the daytime. Considering the heavy night PM_c pollution in Xi'an city, such measures could be reassessed as a policy option.

3. **Conclusion**

In this study, a screening tool called network analysis was applied to process PM_c data derived from distributed air quality monitoring networks in the central urban area of Xi'an city and generate insights indicative of local emission sources, especially for construction sites. Ranking systems was created to identify the sites with high PM_c concentrations, and time series clustering was used to find the representative diurnal patterns for both hotspots and overall urban site. After comparing the representative diurnal patterns of hotspots with the general ones, we found that the site with construction nearby is more likely to show a higher PM_c concentration during the night and early morning, while the morning peak is caused by traffic is not apparent. This phenomenon was verified by a distribution plot, and multiple construction site-related

factors, including the policy and engineering practice, are considered as the explanation. Specifically, as one of the main emission sources, construction-related traffic produced by heavy vehicles transporting is often influenced by local policy. Thus, we discussed the available policy options and their trade-off and encouraged the policymakers to re-assess the construction-related traffic policy based on the trade-off.

Multiple limitations exist for this work. Three primary sources, including construction work, unpaved land, and heavy vehicles to transport construction materials and rubbish, were not taken apart in the local emission-related factors analysis due to difficulty in data collection, thus undermining the significance of policy discussions on traffic. We cannot totally exclude the impact of regional differences due to the relatively medium-scale (20 miles), though all selected monitoring sites are in the center of Xi'an and shows a similar trend in intra-ranking analysis.

Reference

[1] W. E. Wilson and H. H. Suh, “Fine particles and coarse particles: concentration relationships relevant to epidemiologic studies,” *J. Air Waste Manag. Assoc.*, vol. 47, no. 12, pp. 1238–1249, Dec. 1997, doi: 10.1080/10473289.1997.10464074.

[2] P. Kassomenos, S. Vardoulakis, A. Chaloulakou, G. Grivas, R. Borge, and J. Lumbreras, “Levels, sources and seasonality of coarse particles (PM 10-PM 2.5) in three European capitals - Implications for particulate pollution control,” *Atmos. Environ.*, vol. 54, pp. 337–347, Jul. 2012, doi: 10.1016/J.ATMOSENV.2012.02.051.

[3] N. Clements, M. P. Hannigan, S. L. Miller, J. L. Peel, and J. B. Milford, “Comparisons of urban and rural PM10-2.5 and PM2.5 mass concentrations and semi-volatile fractions in northeastern Colorado,” *Atmos. Chem. Phys.*, vol. 16, no. 11, pp. 7469–7484, Jun. 2016, doi: 10.5194/ACP-16-7469-2016.

[4] C. A. Keet, J. P. Keller, and R. D. Peng, “Long-term coarse particulate matter exposure is associated with asthma among children in medicaid,” *Am. J. Respir. Crit. Care Med.*, vol. 197, no. 6, pp. 737–746, Mar. 2018, doi: 10.1164/RCCM.201706-1267OC/SUPPL_FILE/DISCLOSURES.PDF.

[5] H. Powell, J. R. Krall, Y. Wang, M. L. Bell, and R. D. Peng, “Ambient Coarse Particulate Matter and Hospital Admissions in the Medicare Cohort Air Pollution Study, 1999-2010,” *Environ. Health Perspect.*, vol. 123, no. 11, pp.

1152–1158, Nov. 2015, doi: 10.1289/EHP.1408720.

[6] R. D. Peng *et al.*, "Coarse particulate matter air pollution and hospital admissions for cardiovascular and respiratory diseases among Medicare patients," *JAMA*, vol. 299, no. 18, pp. 2172–2179, May 2008, doi: 10.1001/JAMA.299.18.2172.

[7] Y. Zhao, S. Wang, L. Lang, C. Huang, W. Ma, and H. Lin, "Ambient fine and coarse particulate matter pollution and respiratory morbidity in Dongguan, China," *Environ. Pollut.*, vol. 222, pp. 126–131, 2017, doi: 10.1016/J.ENVPOL.2016.12.070.

[8] R. Chen *et al.*, "Associations between coarse particulate matter air pollution and cause-specific mortality: A nationwide analysis in 272 Chinese cities," *Environ. Health Perspect.*, vol. 127, no. 1, Jan. 2019, doi: 10.1289/EHP2711.

[9] P. Kassomenos, S. Vardoulakis, A. Chaloulakou, G. Grivas, R. Borge, and J. Lumbreras, "Levels, sources and seasonality of coarse particles (PM10–PM2.5) in three European capitals – Implications for particulate pollution control," *Atmos. Environ.*, vol. 54, pp. 337–347, Jul. 2012, doi: 10.1016/J.ATMOSENV.2012.02.051.

[10] P. Kumar, P. K. Hopke, S. Raja, G. Casuccio, T. L. Lersch, and R. R. West, "Characterization and heterogeneity of coarse particles across an urban area," *Atmos. Environ.*, vol. 46, pp. 449–459, Jan. 2012, doi: 10.1016/J.ATMOSENV.2011.09.018.

[11] K. Cheung *et al.*, "Spatial and temporal variation of chemical composition and

mass closure of ambient coarse particulate matter (PM10-2.5) in the Los Angeles area," *Atmos. Environ.*, vol. 45, no. 16, pp. 2651–2662, May 2011, doi: 10.1016/J.ATMOSENV.2011.02.066.

[12] T. A. Pakkanen *et al.*, "Sources and chemical composition of atmospheric fine and coarse particles in the Helsinki area," *Atmos. Environ.*, vol. 35, no. 32, pp. 5381–5391, 2001, doi: 10.1016/S1352-2310(01)00307-7.

[13] W. C. Malm, M. L. Pitchford, C. McDade, and L. L. Ashbaugh, "Coarse particle speciation at selected locations in the rural continental United States," *Atmos. Environ.*, vol. 41, no. 10, pp. 2225–2239, Mar. 2007, doi: 10.1016/J.ATMOSENV.2006.10.077.

[14] S. Choung *et al.*, "Comparison of physicochemical properties between fine (PM2.5) and coarse airborne particles at cold season in Korea," *Sci. Total Environ.*, vol. 541, pp. 1132–1138, Jan. 2016, doi: 10.1016/J.SCITOTENV.2015.10.021.

[15] V. Jalali Farahani, A. Altuwayjiri, S. Taghvaee, and C. Sioutas, "Tailpipe and Nontailpipe Emission Factors and Source Contributions of PM10on Major Freeways in the Los Angeles Basin," *Environ. Sci. Technol.*, Jun. 2021, doi: 10.1021/ACS.EST.1C06954/SUPPL_FILE/ES1C06954_SI_001.PDF.

[16] Y. Fujitani *et al.*, "Contribution of industrial and traffic emissions to ultrafine, fine, coarse particles in the vicinity of industrial areas in Japan," *Environ. Adv.*, vol. 5, p. 100101, Oct. 2021, doi: 10.1016/J.ENVADV.2021.100101.

[17] A. Piscitello, C. Bianco, A. Casasso, and R. Sethi, "Non-exhaust traffic

emissions: Sources, characterization, and mitigation measures," *Sci. Total Environ.*, vol. 766, p. 144440, Apr. 2021, doi: 10.1016/J.SCITOTENV.2020.144440.

[18] V. R. J. H. Timmers and P. A. J. Achten, "Non-exhaust PM emissions from electric vehicles," *Atmos. Environ.*, vol. 134, pp. 10–17, Jun. 2016, doi: 10.1016/J.ATMOSENV.2016.03.017.

[19] H. Hagino, M. Oyama, and S. Sasaki, "Laboratory testing of airborne brake wear particle emissions using a dynamometer system under urban city driving cycles," *Atmos. Environ.*, vol. 131, pp. 269–278, Apr. 2016, doi: 10.1016/J.ATMOSENV.2016.02.014.

[20] R. Li, C. Wiedinmyer, K. R. Baker, and M. P. Hannigan, "Characterization of coarse particulate matter in the western United States: A comparison between observation and modeling," *Atmos. Chem. Phys.*, vol. 13, no. 3, pp. 1311–1327, 2013, doi: 10.5194/ACP-13-1311-2013.

[21] M. Eeftens *et al.*, "Development of land use regression models for PM2.5, PM 2.5 absorbance, PM10 and PMcoarse in 20 European study areas; Results of the ESCAPE project," *Environ. Sci. Technol.*, vol. 46, no. 20, pp. 11195–11205, Oct. 2012, doi: 10.1021/ES301948K/SUPPL_FILE/ES301948K_SI_001.PDF.

[22] M. Stafoggia *et al.*, "Estimation of daily PM10 and PM2.5 concentrations in Italy, 2013–2015, using a spatiotemporal land-use random-forest model," *Environ. Int.*, vol. 124, pp. 170–179, Mar. 2019, doi: 10.1016/J.ENVINT.2019.01.016.

[23] G. Hoek *et al.*, “Land use regression model for ultrafine particles in Amsterdam,” *Environ. Sci. Technol.*, vol. 45, no. 2, pp. 622–628, Jan. 2011, doi: 10.1021/ES1023042/ASSET/IMAGES/LARGE/ES-2010-023042_0002.JPEG.

[24] H. Teichgraeber and A. R. Brandt, “Clustering methods to find representative periods for the optimization of energy systems: An initial framework and comparison,” *Appl. Energy*, vol. 239, pp. 1283–1293, Apr. 2019, doi: 10.1016/J.APENERGY.2019.02.012.

[25] J. Duan, Y. Chen, W. Fang, and Z. Su, “Characteristics and Relationship of PM, PM10, PM2.5 Concentration in a Polluted City in Northern China,” *Procedia Eng.*, vol. 102, pp. 1150–1155, Jan. 2015, doi: 10.1016/J.PROENG.2015.01.239.

[26] R. Dobson *et al.*, “Diurnal variability of fine-particulate pollution concentrations: Data from 14 low- And middle-income countries,” *Int. J. Tuberc. Lung Dis.*, vol. 25, no. 3, pp. 206–214, Mar. 2021, doi: 10.5588/IJTLD.20.0704.

[27] M. I. Manning, R. V. Martin, C. Hasenkopf, J. Flasher, and C. Li, “Diurnal Patterns in Global Fine Particulate Matter Concentration,” *Environ. Sci. Technol. Lett.*, vol. 5, no. 11, pp. 687–691, Nov. 2018, doi: 10.1021/ACS.ESTLETT.8B00573/SUPPL_FILE/EZ8B00573_SI_001.PDF.

[28] G. H. Wang *et al.*, “Evolution of aerosol chemistry in Xi’an, inland China, during the dust storm period of 2013 – Part 1: Sources, chemical forms

and formation mechanisms of nitrate and sulfate," *Atmos. Chem. Phys.*, vol. 14, no. 21, pp. 11571–11585, Nov. 2014, doi: 10.5194/ACP-14-11571-2014.

[29] M. Wang *et al.*, "Slower than expected reduction in annual PM2.5 in Xi'an revealed by machine learning-based meteorological normalization," *Sci. Total Environ.*, vol. 841, Oct. 2022, doi: 10.1016/j.scitotenv.2022.156740.

[30] Z. Wang *et al.*, "The seasonal variation, characteristics and secondary generation of PM2.5 in Xi'an, China, especially during pollution events," *Environ. Res.*, vol. 212, Sep. 2022, doi: 10.1016/j.envres.2022.113388.

[31] P. Wang *et al.*, "Spatial and seasonal variations of PM2.5 mass and species during 2010 in Xi'an, China," *Sci. Total Environ.*, vol. 508, pp. 477–487, Mar. 2015, doi: 10.1016/J.SCITOTENV.2014.11.007.

[32] H. Xu *et al.*, "Inter-annual variability of wintertime PM2.5 chemical composition in Xi'an, China: Evidences of changing source emissions," *Sci. Total Environ.*, vol. 545–546, pp. 546–555, Mar. 2016, doi: 10.1016/J.SCITOTENV.2015.12.070.

[33] X. Niu *et al.*, "PM2.5 from the Guanzhong Plain: Chemical composition and implications for emission reductions," *Atmos. Environ.*, vol. 147, pp. 458–469, Dec. 2016, doi: 10.1016/j.atmosenv.2016.10.029.

[34] D. Wang *et al.*, "Source contributions to primary and secondary inorganic particulate matter during a severe wintertime PM2.5 pollution episode in Xi'an, China," *Atmos. Environ.*, vol. 97, pp. 182–194, 2014, doi: 10.1016/J.ATMOSENV.2014.08.020.

[35] Z. Shen *et al.*, “Chemical characteristics of fine particles (PM1) from Xi’an, China,” *Aerosol Sci. Technol.*, vol. 44, no. 6, pp. 461–472, Jun. 2010, doi: 10.1080/02786821003738908.

[36] W. Huang *et al.*, “Seasonal variation of chemical species associated with short-term mortality effects of PM 2.5 in Xi’an, a central city in China,” *Am. J. Epidemiol.*, vol. 175, no. 6, pp. 556–566, Mar. 2012, doi: 10.1093/AJE/KWR342.

[37] X. Liu, Y. Hui, Z. Y. Yin, Z. Wang, X. Xie, and J. Fang, “Deteriorating haze situation and the severe haze episode during December 18–25 of 2013 in Xi’an, China, the worst event on record,” *Theor. Appl. Climatol.*, vol. 125, no. 1–2, pp. 321–335, Jul. 2016, doi: 10.1007/S00704-015-1509-8.

[38] Y. Chen, J. Cao, R. Huang, F. Yang, Q. Wang, and Y. Wang, “Characterization, mixing state, and evolution of urban single particles in Xi’an (China) during wintertime haze days,” *Sci. Total Environ.*, vol. 573, pp. 937–945, Dec. 2016, doi: 10.1016/j.scitotenv.2016.08.151.

[39] J. J. Cao *et al.*, “Impacts of aerosol compositions on visibility impairment in Xi’an, China,” *Atmos. Environ.*, vol. 59, pp. 559–566, Nov. 2012, doi: 10.1016/j.atmosenv.2012.05.036.

[40] Q. Dai *et al.*, “Chemical nature of PM2.5 and PM10 in Xi’an, China: Insights into primary emissions and secondary particle formation,” *Environ. Pollut.*, vol. 240, pp. 155–166, Sep. 2018, doi: 10.1016/J.ENVPOL.2018.04.111.

[41] M. Elser *et al.*, “New insights into PM2.5 chemical composition and sources in

two major cities in China during extreme haze events using aerosol mass spectrometry," *Atmos. Chem. Phys.*, vol. 16, no. 5, pp. 3207–3225, Mar. 2016, doi: 10.5194/ACP-16-3207-2016.

[42] "Xi'an." http://en.shaanxi.gov.cn/as/cities/201705/t20170505_1595150.html (accessed Jul. 28, 2022).

[43] "经济社会发展统计图表：第七次全国人口普查超大、特大城市人口基本情况 - 求是网." http://www.qstheory.cn/dukan/qs/2021-09/16/c_1127863567.htm (accessed Jul. 28, 2022).

[44] E. S. Macias and R. B. Husar, "Atmospheric Particulate Mass Measurement with Beta Attenuation Mass Monitor," *Environ. Sci. Technol.*, vol. 10, no. 9, pp. 904–907, Sep. 1976, doi: 10.1021/ES60120A015/ASSET/ES60120A015.FP.PNG_V03.

[45] "eCFR :: 40 CFR Part 53 -- Ambient Air Monitoring Reference and Equivalent Methods." https://www.ecfr.gov/current/title-40/chapter-I/subchapter-C/part-53 (accessed Jul. 11, 2022).

[46] D. Arthur and S. Vassilvitskii, "K-means++: The advantages of careful seeding," *Proc. Annu. ACM-SIAM Symp. Discret. Algorithms*, vol. 07-09-Janu, pp. 1027–1035, 2007.

[47] R. Xu and D. C. Wunsch, "Partitional Clustering," *Clustering*, pp. 63–110, Jan. 2009, doi: 10.1002/9780470382776.CH4.

[48] R. L. Thorndike, "Who belongs in the family?," *Psychom. 1953 184*, vol. 18, no. 4, pp. 267–276, Dec. 1953, doi: 10.1007/BF02289263.

[49] “蓝天保卫战专家谈（5）丨扬尘是颗粒物的重要来源，应持续强化管控_中华人民共和国生态环境部.” https://www.mee.gov.cn/ywgz/dqhjbh/dqhjzlgl/202003/t20200312_768760.shtml (accessed Mar. 09, 2022).

[50] “广东省环境保护厅关于对广东省十二届人大三次会议第1218号代表建议的答复摘要- 广东省生态环境厅公众网.” http://gdee.gd.gov.cn/jgzy/content/post_2335098.html (accessed Jul. 26, 2022).

[51] D. Chen, S. Cheng, Y. Zhou, X. Guo, S. Fan, and H. Wang, “Impact of Road Fugitive Dust on Air Quality in Beijing, China,” *https://home.liebertpub.com/ees*, vol. 27, no. 10, pp. 825–834, Oct. 2010, doi: 10.1089/EES.2009.0122.

[52] Z. Zhou, Q. Tan, Y. Deng, K. Wu, X. Yang, and X. Zhou, “Emission inventory of anthropogenic air pollutant sources and characteristics of VOCs species in Sichuan Province, China,” *J. Atmos. Chem. 2019 761*, vol. 76, no. 1, pp. 21–58, Jan. 2019, doi: 10.1007/S10874-019-9386-7.

[53] V. Singh, A. Biswal, A. P. Kesarkar, S. Mor, and K. Ravindra, “High resolution vehicular PM10 emissions over megacity Delhi: Relative contributions of exhaust and non-exhaust sources,” *Sci. Total Environ.*, vol. 699, p. 134273, Jan. 2020, doi: 10.1016/J.SCITOTENV.2019.134273.

[54] T. Li *et al.*, “Application and validation of the fugitive dust source emission inventory compilation method in Xiong’an New Area, China,” *Sci. Total*

Environ., vol. 798, p. 149114, Dec. 2021, doi: 10.1016/J.SCITOTENV.2021.149114.

[55] G. P. Compo, "Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory.," *International Surface Pressure Databank (ISPDv2).*, 2010. https://doi.org/10.5065/D6SQ8XDW (accessed May 02, 2022).

[56] "Non-exhaust Particulate Emissions from Road Transport," *Non-exhaust Part. Emiss. from Road Transp.*, Dec. 2020, doi: 10.1787/4A4DC6CA-EN.

[57] K. Schäfer, S. Emeis, H. Hoffmann, and C. Jahn, "Influence of mixing layer height upon air pollution in urban and sub-urban areas," *Meteorol. Zeitschrift*, vol. 15, no. 6, pp. 647–658, Dec. 2006, doi: 10.1127/0941-2948/2006/0164.

[58] P. Wagner and K. Schäfer, "Influence of mixing layer height on air pollutant concentrations in an urban street canyon," *Urban Clim.*, vol. 22, pp. 64–79, Dec. 2017, doi: 10.1016/J.UCLIM.2015.11.001.

[59] J. Rost *et al.*, "Variability of PM10 concentrations dependent on meteorological conditions," *Int. J. Environ. Pollut.*, vol. 36, no. 1–3, pp. 3–18, 2009, doi: 10.1504/IJEP.2009.021813.

[60] H. B. Mann and D. R. Whitney, "On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other," *https://doi.org/10.1214/aoms/1177730491*, vol. 18, no. 1, pp. 50–60, Mar. 1947, doi: 10.1214/AOMS/1177730491.

[61] B. Shao, Z. Hu, Q. Liu, S. Chen, and W. He, "Fatal accident patterns of

building construction activities in China," *Saf. Sci.*, vol. 111, pp. 253–263, Jan. 2019, doi: 10.1016/J.SSCI.2018.07.019.

Supporting Information

S1. Data statistics and processing

Table S1-S4 shows that the PM_c/PM_{10} ratio is higher in April-May 2020 than in December 2019 and July 2019. The statistical data for April-May 2020 are shown in Table S5. Since PM_c is generated by PM_{10}-$PM_{2.5}$, there are less than 5% negative values for PM_c, which is normal due to measurement error. In time series clustering, extremely low value (lower than 0.05%) was removed, and extreme high value (higher than 99.95%, 300 $\mu g/m^3$) was converted to 300 $\mu g/m^{3.}$ A negative value higher than -30 was converted to 0.

Table S1

Apr-20	$PM_{2.5}$	PM_{10}	PM_c	PM_c/PM_{10}
count	37711.00	37711.00	37711.00	37711.00
mean	39.55	85.28	45.73	0.46
std	21.37	45.10	38.14	0.37
min	1.00	1.00	-545.00	-14.57
5%	15.00	28.00	-5.00	-0.14
25%	25.00	52.00	21.00	0.39
50%	36.00	78.00	42.00	0.55
75%	50.00	111.00	65.00	0.65
95%	74.00	167.00	111.00	0.77
max	632.00	891.00	874.00	0.98

Table S2

May-20	$PM_{2.5}$	PM_{10}	PM_c	PM_c/PM_{10}
count	40347.00	40347.00	40347.00	40347.00
mean	32.53	78.60	46.08	0.53
std	19.10	47.58	36.63	0.37
min	1.00	1.00	-250.00	-28.00
5%	10.00	23.00	4.00	0.14
25%	19.00	47.00	24.00	0.47
50%	29.00	69.00	39.00	0.59
75%	41.00	97.00	60.00	0.67
95%	68.00	173.00	117.00	0.79
max	304.00	737.00	705.00	0.99

Table S3

Jul-19	$PM_{2.5}$	PM_{10}	PM_c	PM_c/PM_{10}
count	36949.00	36949.00	36949.00	36949.00
mean	24.89	42.12	17.24	0.33
std	14.71	23.33	20.36	0.55
min	1.00	1.00	-331.00	-46.00
5%	8.00	14.00	-6.00	-0.27
25%	16.00	27.00	7.00	0.27
50%	23.00	38.00	15.00	0.43
75%	31.00	53.00	26.00	0.55
95%	47.00	83.00	47.00	0.70
max	374.00	874.00	853.00	0.99

Table S4

Dec-19	$PM_{2.5}$	PM_{10}	PM_c	PM_c/PM_{10}
count	39502.00	39502.00	39502.00	39502.00

mean	72.14	126.44	54.30	0.45
std	46.97	64.46	30.83	0.18
min	1.00	4.00	-102.00	-12.75
5%	17.00	42.00	15.00	0.18
25%	36.00	76.00	33.00	0.35
50%	60.00	116.00	49.00	0.46
75%	97.00	165.00	71.00	0.56
95%	168.00	250.00	110.00	0.69
max	301.00	772.00	734.00	0.98

Table S5

Original (April and May 2020)	PM_c	After preprocessing	PM_c
count	75581	count	75267
mean	45.753165	mean	46.600462
std	37.701932	std	35.344297
min	-545	min	0
0.10%	-80.84	0.10%	0
0.50%	-26	0.50%	0
1%	-18	1%	0
2%	-11	2%	0
3%	-6	3%	0
4%	-2	4%	0
4.50%	-1	4.50%	0
5%	0	5%	1
20%	18	20%	19
40%	33	40%	33
50%	40	50%	40
60%	48	60%	48
80%	69	80%	69

99.50%	187	99.50%	187
99.95%	301.21	99.95%	300
max	874	max	300

S2. Traffic flow count dataset

77149 Xi'an's Road sections' hourly traffic flow data were collected in weekdays and weekends, respectively. Table S6 shows the length and average volume for traffic data. Other information include longitude, latitude and road level. The road sections with traffic flow records are shown in Figure S1.

Table S6

	Length (m)	AveVolume
count	77149	77149
mean	155.953117	559.132652
std	207.902063	504.900141
min	0.011714	0
5%	8.455662	121
25%	24.48773	228
50%	83.146839	375
75%	197.497815	699
95%	576.290309	1667
max	3248.538202	4142

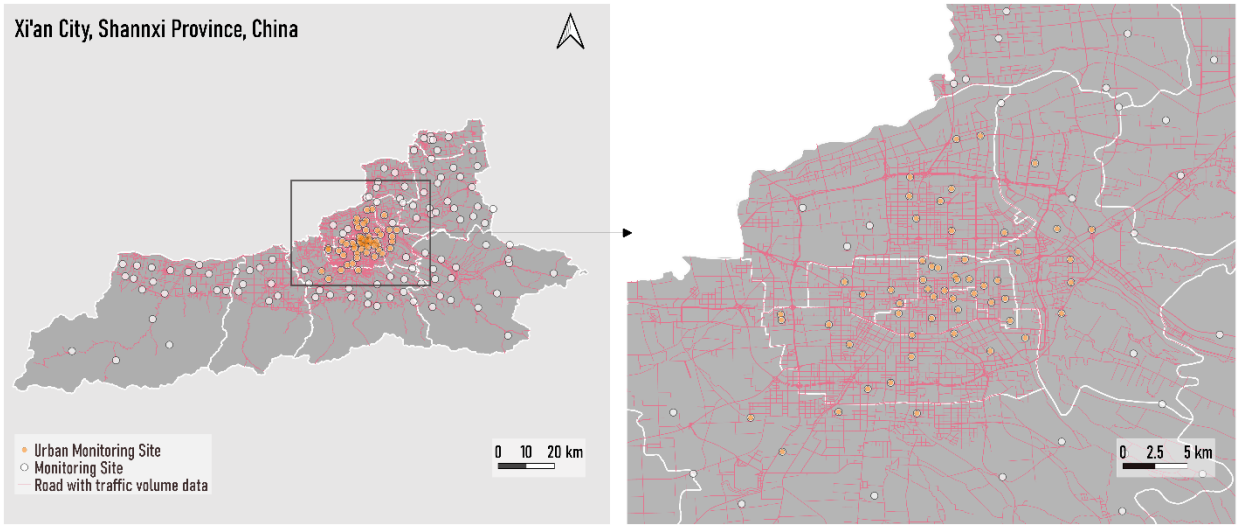


Figure S1. Roads with traffic flow data

S3. Time series clustering results

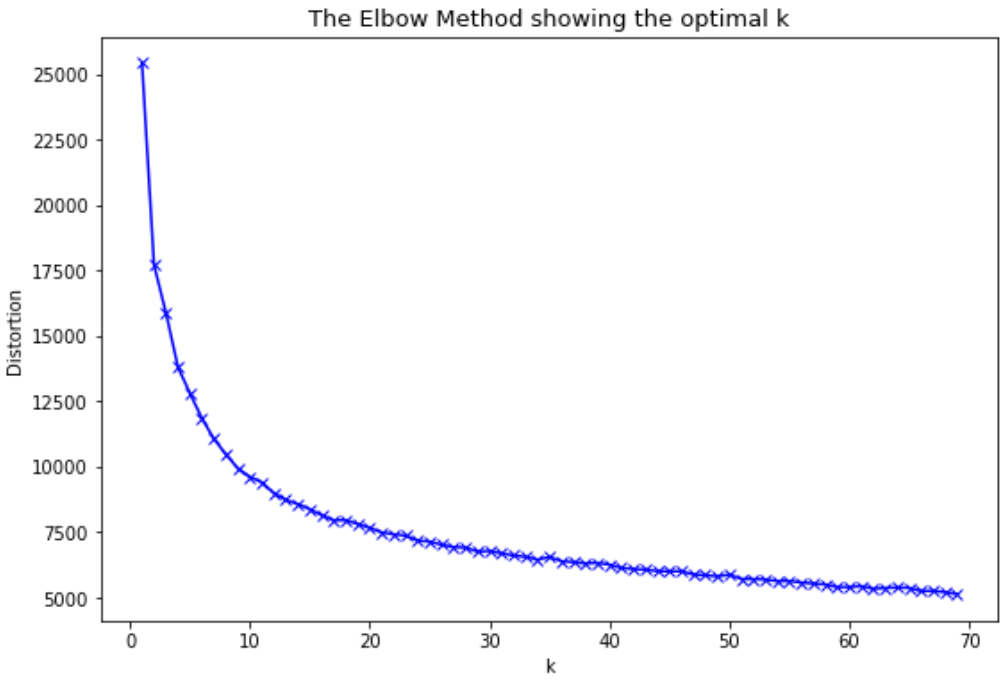


Figure S2. Elbow method results

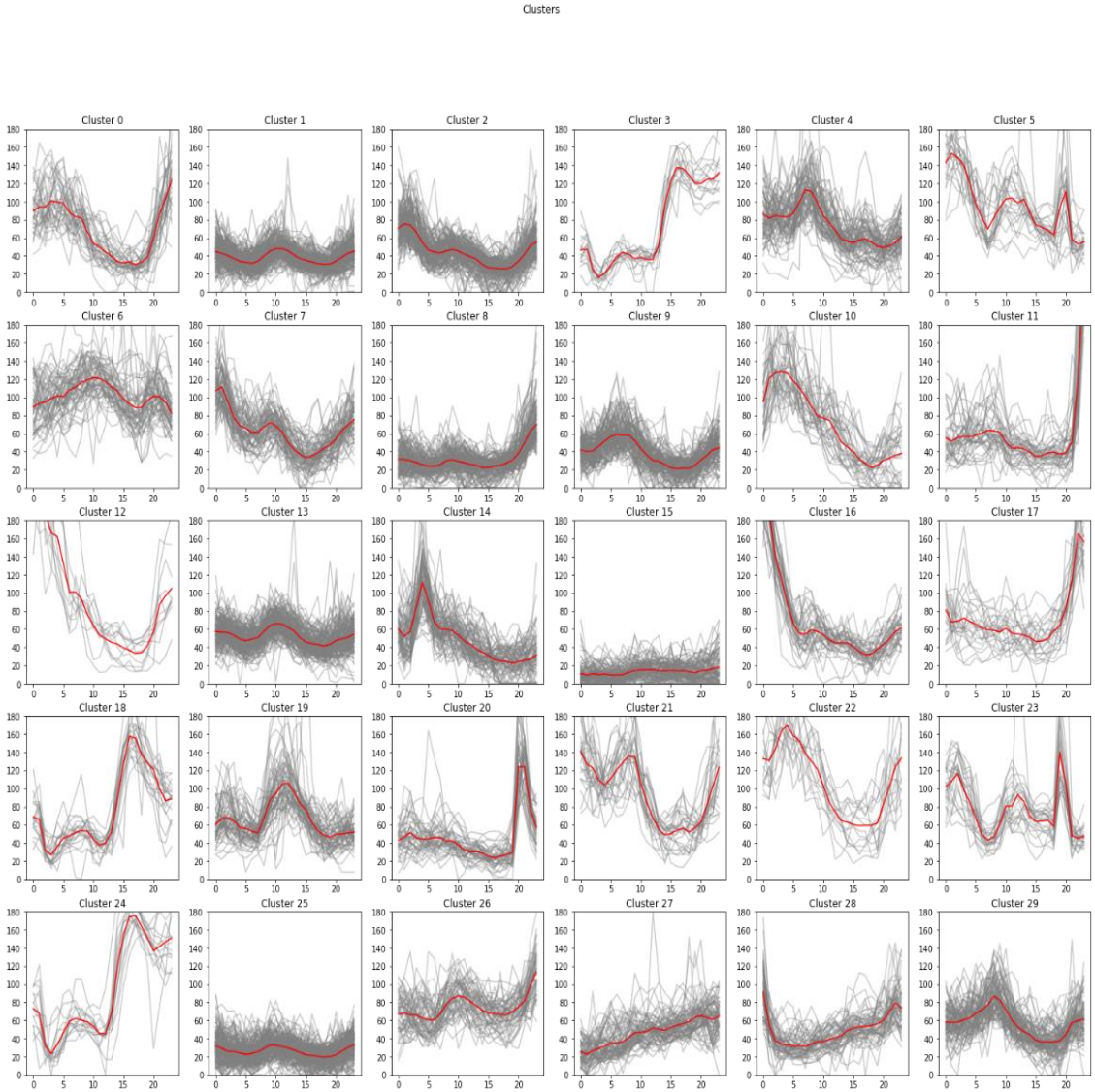


Figure S3. Clustering results: diurnal patterns

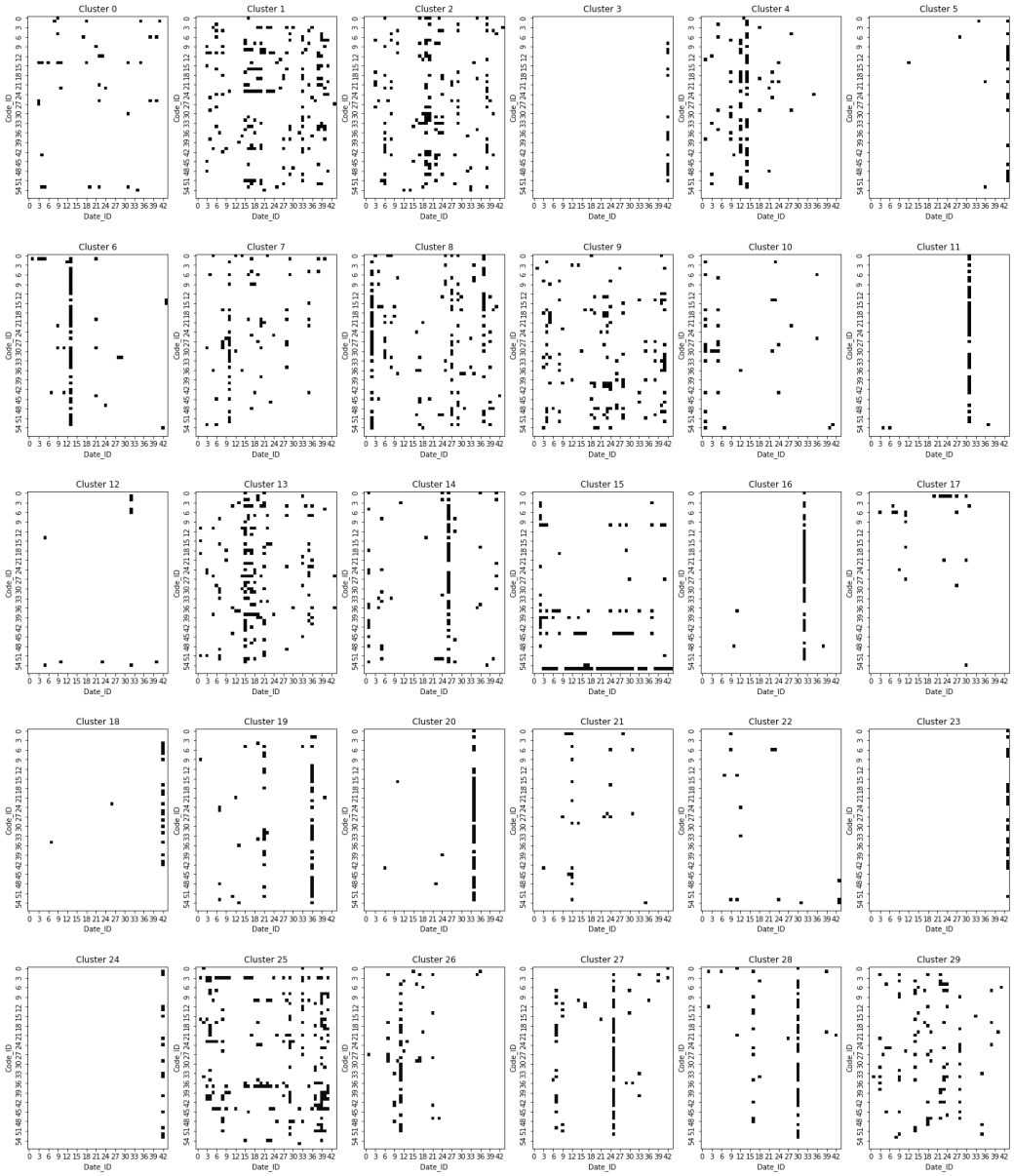


Figure S4. Clustering results: Spatial-temporal distribution

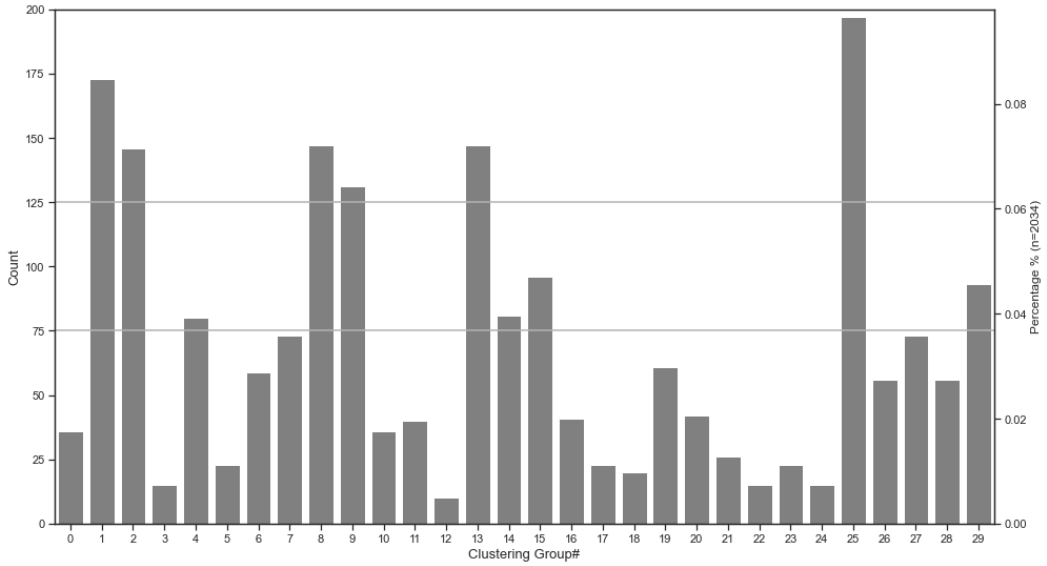


Figure S5. The sizes of clusters

S4. General diurnal pattern identification

Here we analyzed the relatively large clusters with a size range from 50 to 100 and mean diurnal patterns to prove that diurnal patterns with morning and evening peaks are the general pattern.

As shown in Figures S5, S6, and S7, 35.8% of diurnal patterns (count=728) were clustered into relatively large clusters. Unlike the six largest clusters, it could be found that most of them, including clusters 4, 6, 14, 19, 26, 27, and 28, happened mainly in a single day or two days, indicating the potential unrecorded regional events' impact. For the clusters with a relatively scattered distribution, including cluster 7 and cluster 29, their shapes of the mean diurnal pattern are close to the general diurnal pattern we identified.

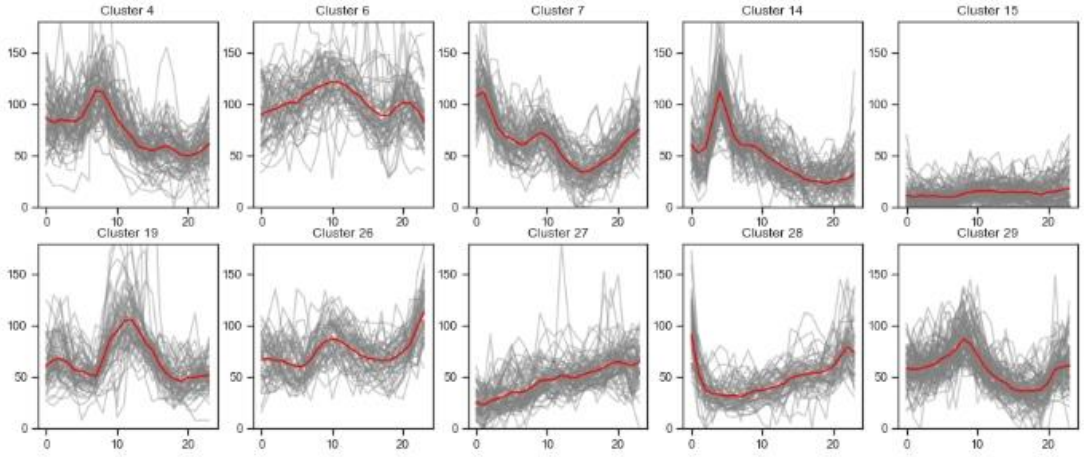


Figure S6

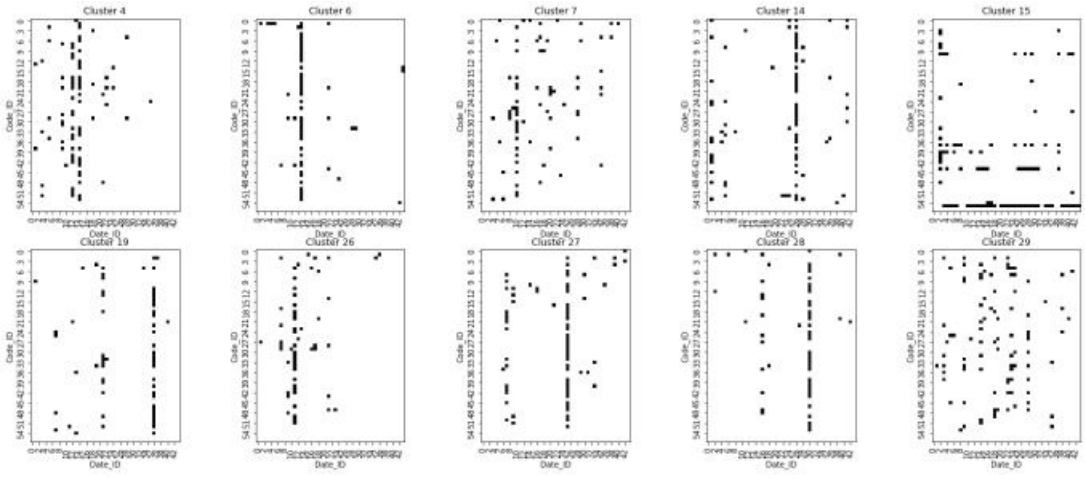


Figure S7

The mean diurnal pattern for all urban sites in selected period (2020/04-2020/05) is shown in Figure S8. Each grey line represents a single site's mean diurnal pattern while the red line represents the overall. Most of the sites and the overall showed a diurnal pattern with morning peaks and evening peaks, which is consistent with our analysis of general patterns.

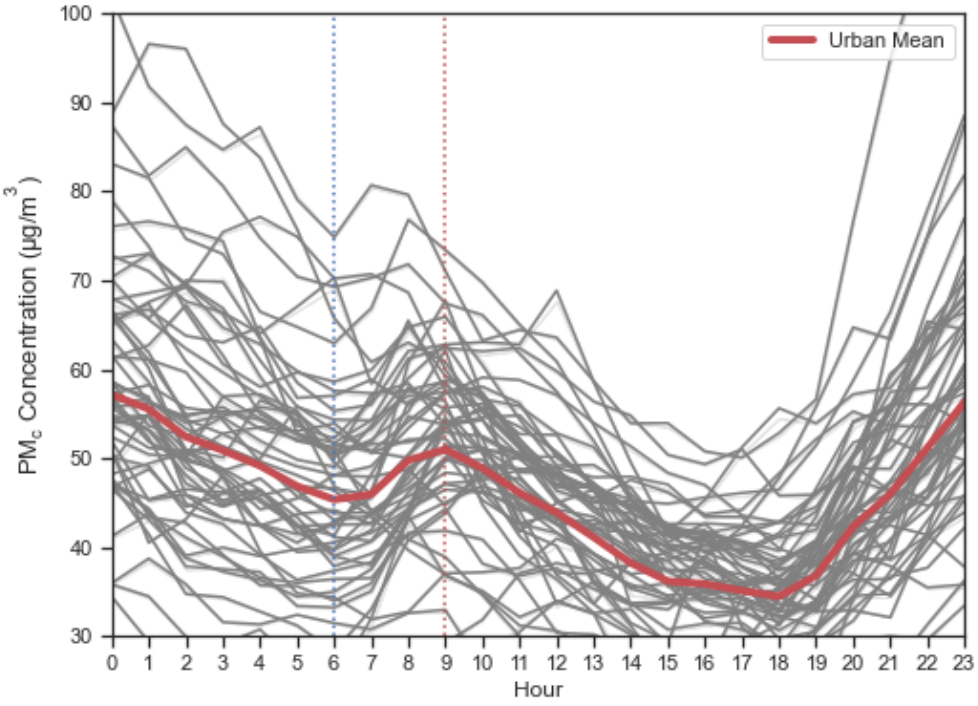

Urban Mean
100
90
80
70
60
50
40
30
PM$_c$ Concentration (µg/m^3)
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
Hour

Figure S8

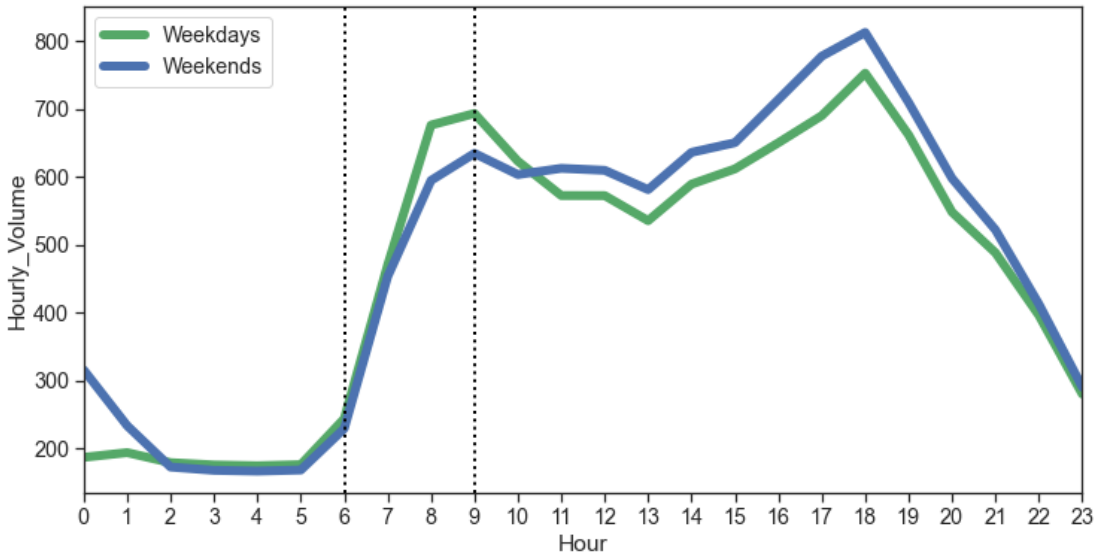


Figure S9

Table S7

Cluster	Monday	Tuesday	Wedenesday	Thursday	Friday	Saturday	Sunday	Sum	Notes
1	8%	16%	28%	14%	12%	16%	7%	100 %	
13	11%	20%	20%	16%	17%	9%	7%	100%	General patterns
25	15%	11%	26%	17%	11%	8%	13%	100%	
8	18%	16%	13%	16%	2%	25%	11%	100%	General patterns (higher night peak)
9	15%	11%	11%	8%	24%	14%	17%	100%	Unresolved
2	3%	17%	14%	13%	11%	17%	25%	100%	General patterns (delayed night peak at 1:00)

S5. Emission-related features

Table S8. The emission-related features of hotspots

Code	Traffic	Construction

	Roadside	Road type	Traffic Flow Counts Hourly Average	Construction site (~300m)	Type	Stage	Unpaved area
1008C	Y	Minor roads	N/A	Y	A, Metro station complex	Foundation construction/Main body construction	Y
					B, Olympics Axis Park		
1060C	Y	Urban highway	1569	Y	Residential building	Main body construction	Y
1168C		Minor roads	206				
1177C	Y	Urban highway	2579	Y	Unknown	Main body construction	Y
1019C							
1208C	Y	Urban highway	786	Y	A, Residential building	Main body construction	
					B, Public building	Earth excavation	Y
1058C	Y	Minor roads	49				
1174C							
1193C				Y	Residential building	Foundation construction/Main body construction	
1170C	Y	Provincial highway	973	Y (~600m)	Xi'an Train Station	Earth excavation/Main body construction	Y

1073C	Y	Minor roads	544	Y	Metro station	Earth excavation/Foundation construction/Earthwork backfill	Y
1175C				Y (Not active in May)	Residential building	Main body construction	
					Residential building	Earth excavation	Y
1209C				Y	Metro construction	Main body construction	
					Residential building	Earth excavation	Y

S6. Examples of hotspots with construction

Figure S10

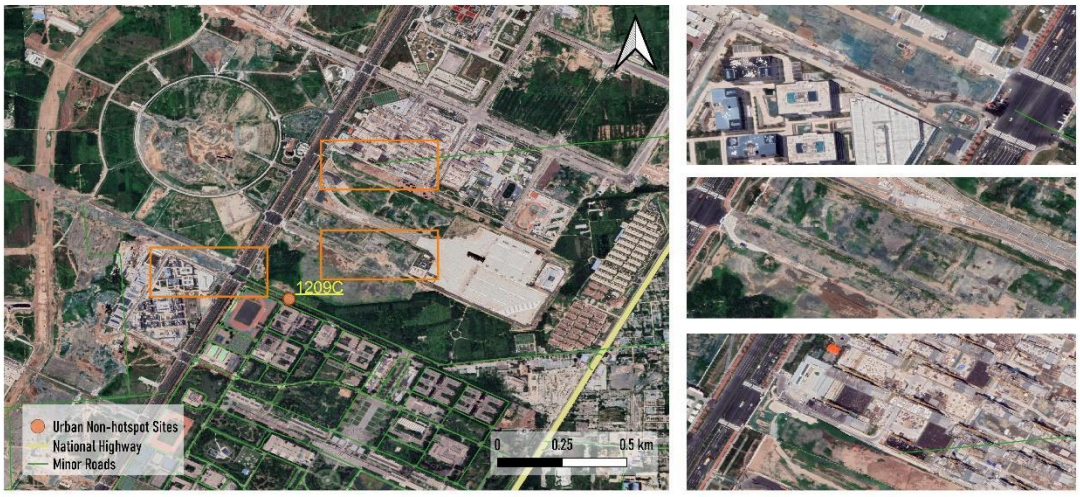


Figure S11

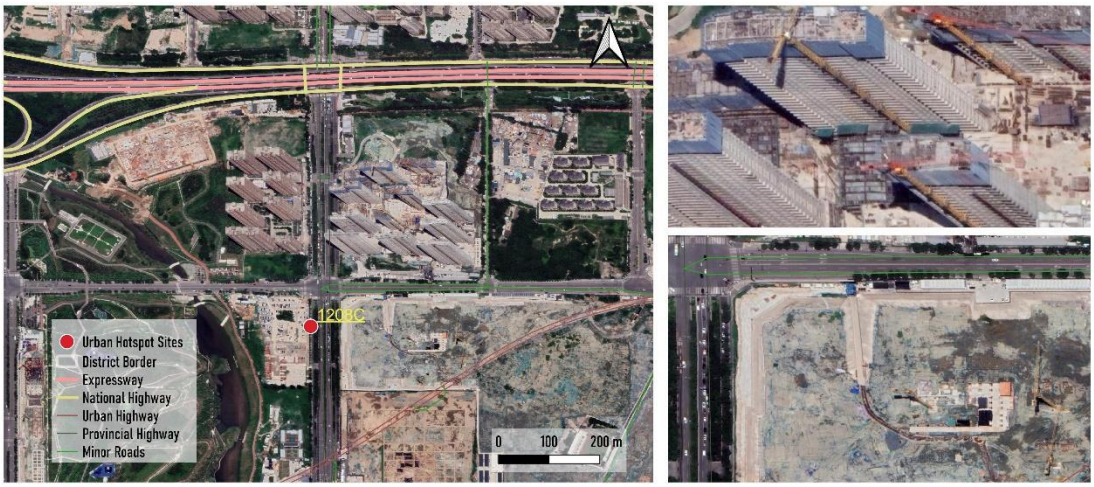


Figure S12


1060C
1174C
Urban Hotspot Sites
District Border
Expressway
National Highway
Urban Highway
Provincial Highway
Minor Roads
0 250 500 m
1174C
0 50 m
1060C
0 100 200 m

Figure S13