Supporting Information

Revealing Nighttime Construction-Related Activities from a Spatially Distributed Air Quality Monitoring Network

*Jintao Gu,[a] Bo Yuan,[a] Shefford P. Baker,[b] Shaojun Zhang,[cd] Xiaomeng Wu,[c] Ye Wu,[cd] and K. Max Zhang[a]**

[a]Sibley School of Mechanical and Aerospace Engineering, Cornell University, Ithaca, NY, 14853, United States

[b]Department of Materials Science and Engineering, Cornell University, Bard Hall, Ithaca, New York, 14853, United States

[c]School of Environment, State Key Joint Laboratory of Environment Simulation and Pollution Control, Tsinghua University, Beijing 100084, China

[d]State Environmental Protection Key Laboratory of Sources and Control of Air Pollution Complex, Beijing 100084, China

Section S1. The statistical analysis of PM data and data preprocessing

Table S1. The statistical analysis of original hourly PM data (unit: μg/m^3) in April-May 2020 for 56 urban monitoring stations and 109 other monitoring stations. Since PM_c is calculated from PM_{10}-$PM_{2.5}$, there are ~5% negative values for PM_c, which is normal due to measurement error. In Network Analysis, values falling below -30 (~0.50%) were eliminated, those values surpassing 99.95% (300 μg/m^3) were adjusted to 300 μg/m^3, and negative values over -30 were reset to 0.

	56 Urban Monitoring Stations			109 Other Monitoring Stations		
	$PM_{2.5}$	PM_{10}	PM_c	$PM_{2.5}$	PM_{10}	PM_c
count	75581	75581	75581	140159	140159	140159
mean	35.71	81.46	45.75	28.26	64.55	36.29
std	20.68	46.93	37.7	19.39	45.16	36.02
min	1	1	-545	1	1	-1272
0.05%	1	4	-133	1	2	-133
0.50%	5	10	-26	2	5	-24
5%	11	25	0	8	16	0
25%	22	49	22	16	34	14
50%	32	72	40	24	53	28
75%	45	104	62	36	83	50
95%	71	170	114	61	149	101
99.50%	113	260	187	100	255	187

99.95%	212	380	301	201	386	297
max	632	891	874	1281	960	944

Section S2. Time series clustering algorithm and implementation.

Each monitoring site records a PM_c measurement each hour and the 24 measurements from one site from one day are put into a single 24-dimensional time-series data point, $\boldsymbol{x_i}$. We want to identify the different types of behavior possible. To do this, we sort the data points into clusters in which each of the diurnal time series have similar behavior. We use the K-means algorithm[1], one of the most popular clustering algorithms, which has been widely applied for time series clustering[2,3]. Assume we have a set of *n* data points $\chi(\boldsymbol{x_1}, \boldsymbol{x_2}, \boldsymbol{x_3} \ldots \boldsymbol{x_n})$. The procedure is as follows:

Step 1. We first assume that there are *k* different types of behavior, and we generate *k* data points 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 data point which we initially populate with random values.

Step 2. We now divide the *n* data points into *k* clusters $\mathbb{C}(S_1, S_2, S_3 \ldots S_k)$ based on the centers: We calculate the distance between each data point and each of the centers using the Euclidean distance shown in (S2-1) and place each data point 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:

$$d(\boldsymbol{c}, \boldsymbol{x}) = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \cdots + (a_n + b_n)^2} \quad \text{(S2-1)}$$

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

$$\boldsymbol{c_i} = \frac{1}{|S_i|} \sum_{\boldsymbol{x} \in S_i} x \quad \text{(S2-2)}$$

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 [4] 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 data points *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. The "elbow" method results are shown below and number 25~35 seems like a good number to find the typical type of diurnal behaviors; thus, we used 30 as our final cluster count parameters.

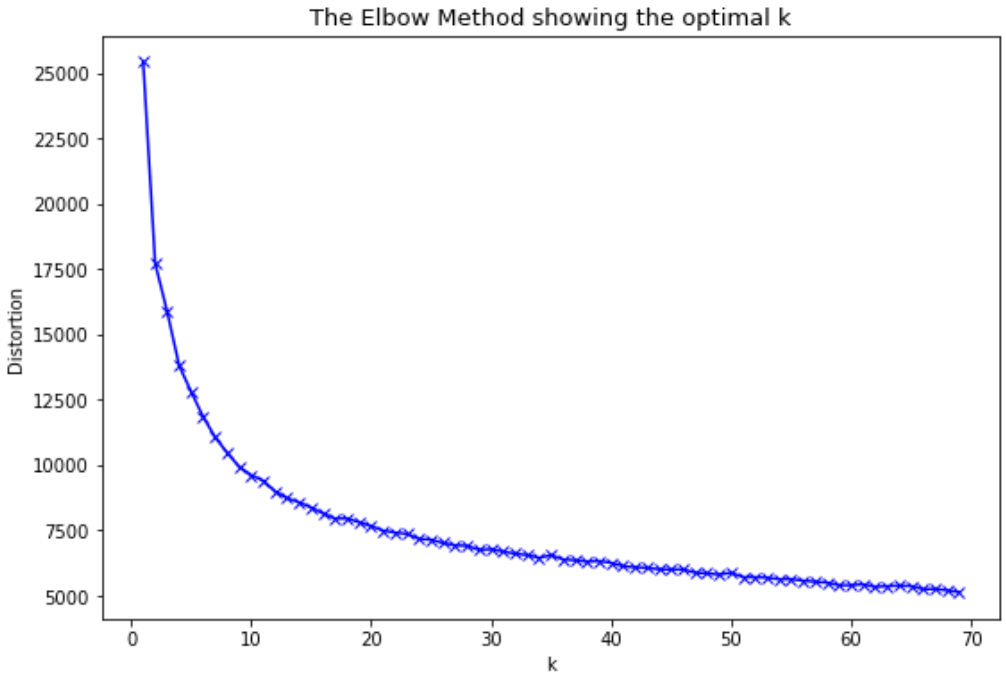

Figure S1. Elbow method results. Elbow Method plot showing the optimal number of clusters for PM_c diurnal variations, determined by the point where reduction in Within-Cluster-Sum of Squared Errors (WSS) begins to diminish. Number 25~35 seems like a good number here, and we selected 30.

Section S3. The emission sources and related features of local PM_c

Dust from the road and construction is supposed to be the main contributor to local urban PM_c emissions in China. Evidence is shown here. According to information released by the website of the Ministry of Ecology and Environment of the People's Republic of China[5], 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[6], 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, 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[7], 60% in the Sichuan Province[8], and 79% in Delhi[9], indicating a higher contribution rate to PM_c in those areas. In a recent study conducted in the Xiong'an new area, China, 98.49% of fugitive dust originated from road dust and construction dust and accounted for 29.38% of PM_{10} emissions[10].

Table S2. Traffic and construction PM_c emission source and related features.

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[a]), 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	

[a] Y/N means yes or no

Table S3. Traffic data for Xi'an City. The hourly traffic flow data were collected on a sample day for weekdays and weekends, respectively. The descriptive statistics data here is the road-based 24h average traffic counts for road sections. The visualization of roads with traffic counts is in **Figure S2**.

	Hourly Average Volume (in the sample day of weekdays)	Hourly Average Volume (in the sample day of weekends)
Number of road sections	77149	76840
mean	559	574
std	505	500
min	0	15
5%	121	154
25%	228	255
50%	375	384
75%	699	703
95%	1667	1720
max	4142	4103

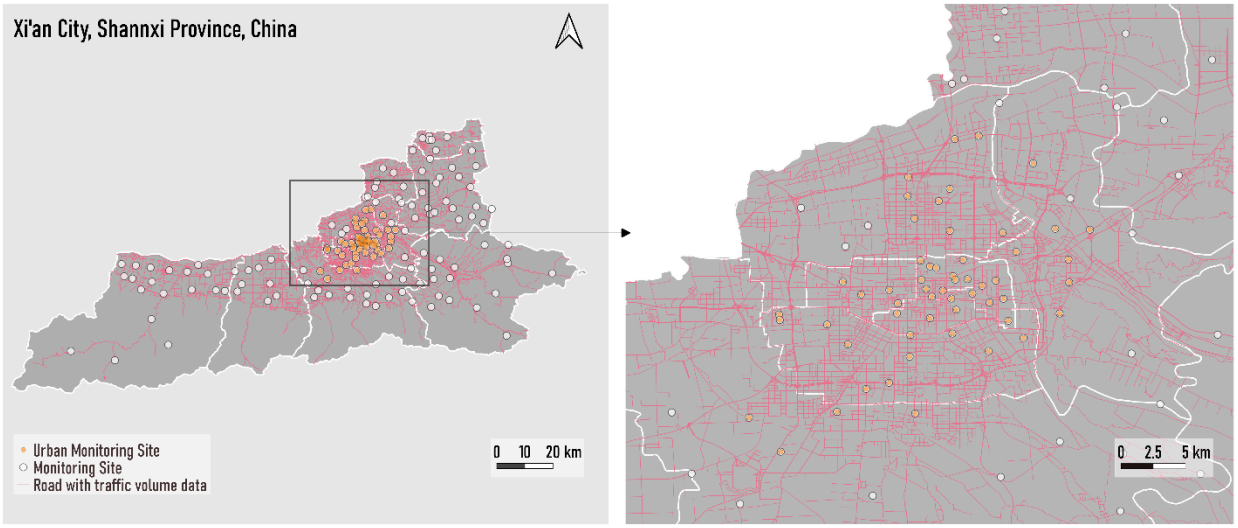

Figure S2. Roads with traffic flow data. Each road section has hourly-based traffic flow data on a sample day during workdays and a sample day during weekends. Both sample days are in April-May 2020.

Section S4. Additional information for intra-ranking system analysis and time series clustering results

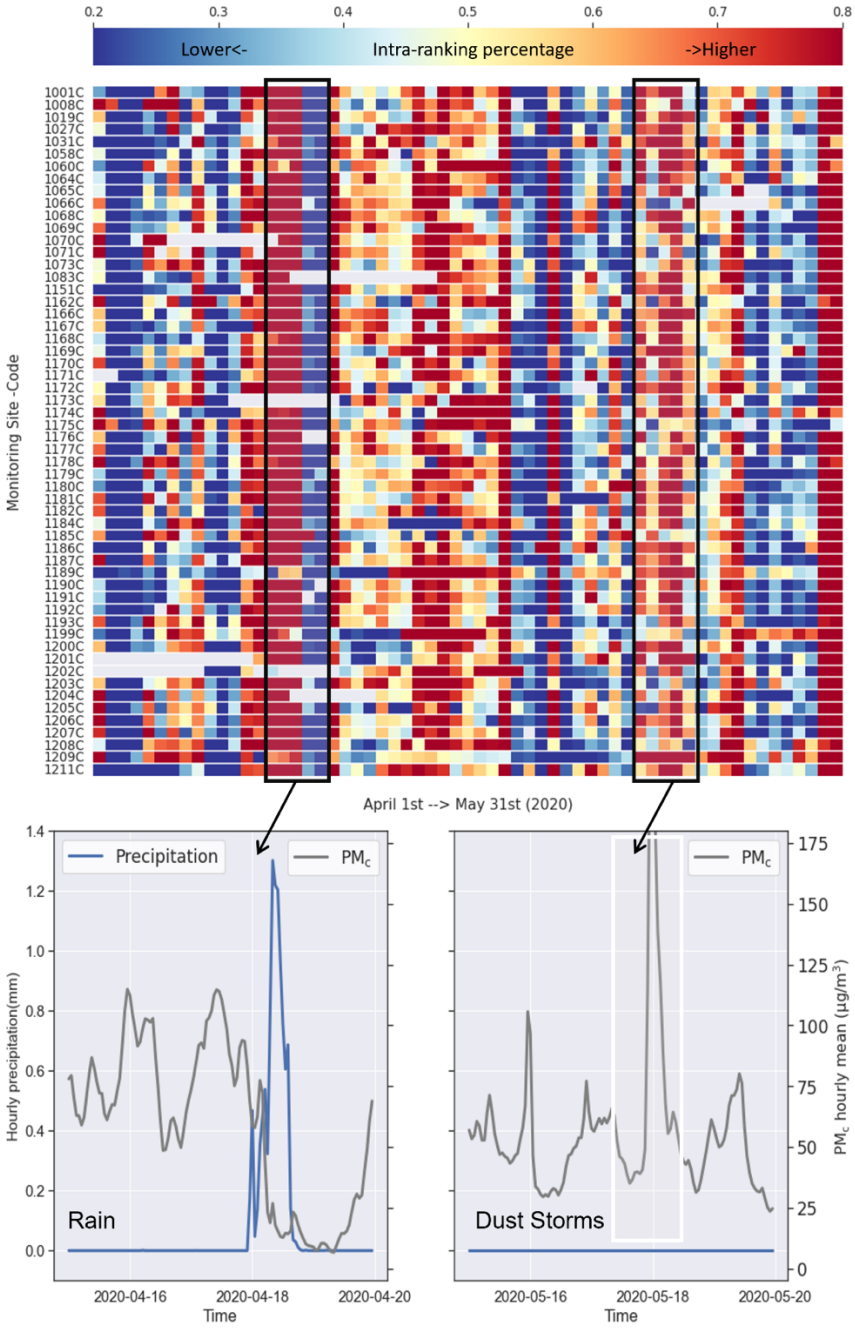

Figure S3. The intra-ranking results with more information. Days with extremely low (80% of sites are below 25% in a single day) or high PM_c concentration across the whole network (80% of sites are above 75% in a single day) were identified by the intra-ranking system and are excluded for further inter-ranking analysis and time series clustering if a record about regional events such as a dust storm (based on weather forecast history and news collection) or rain (based on ERA5 reanalysis data: https://www.ecmwf.int/en/forecasts/dataset/ecmwf-reanalysis-v5) could be found. Fifteen days were excluded before time series clustering.

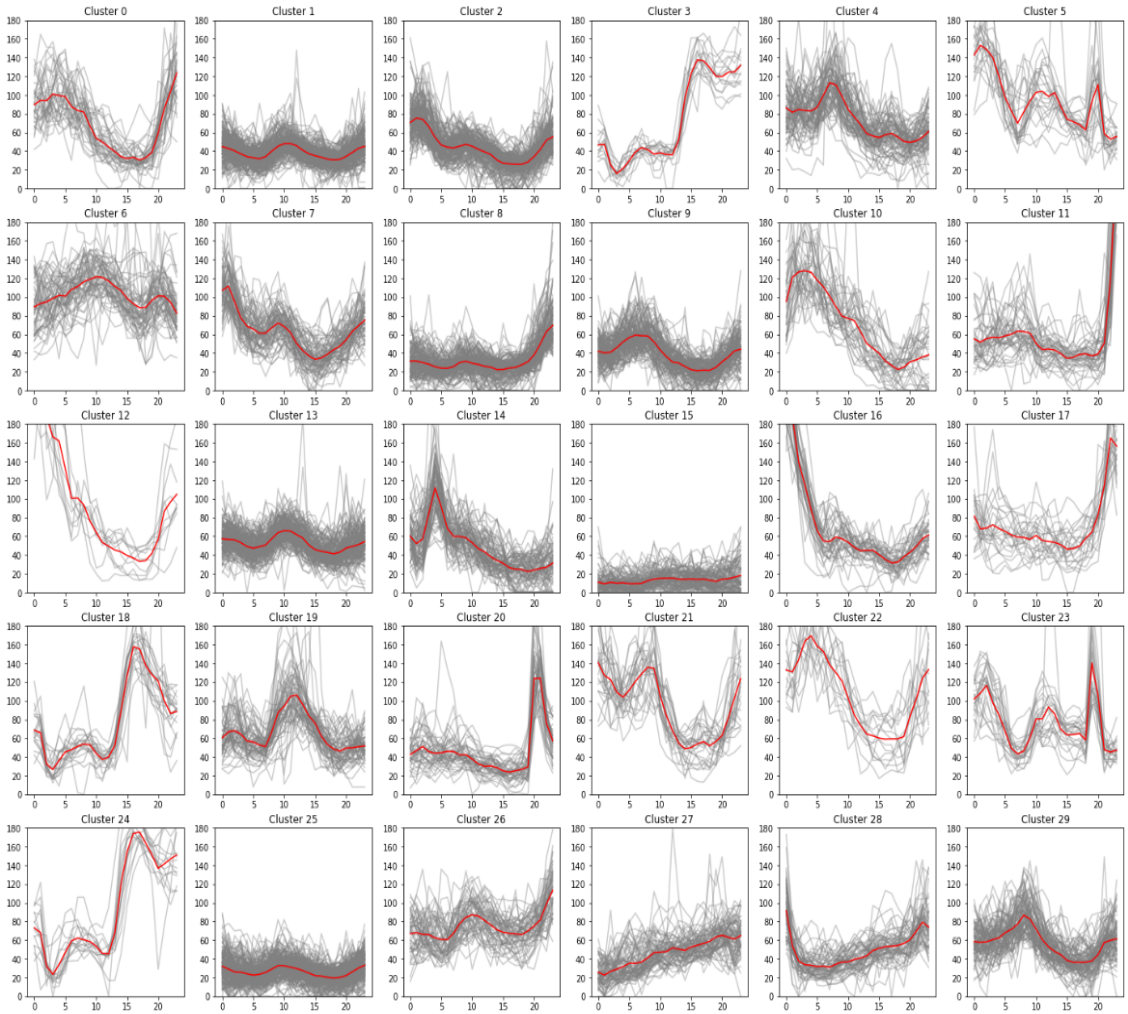

Figure S4. Clustering results. For each cluster, diurnal curves are shown in grey lines, and the red line is the mean. The patterns of those clusters are analyzed from red lines.

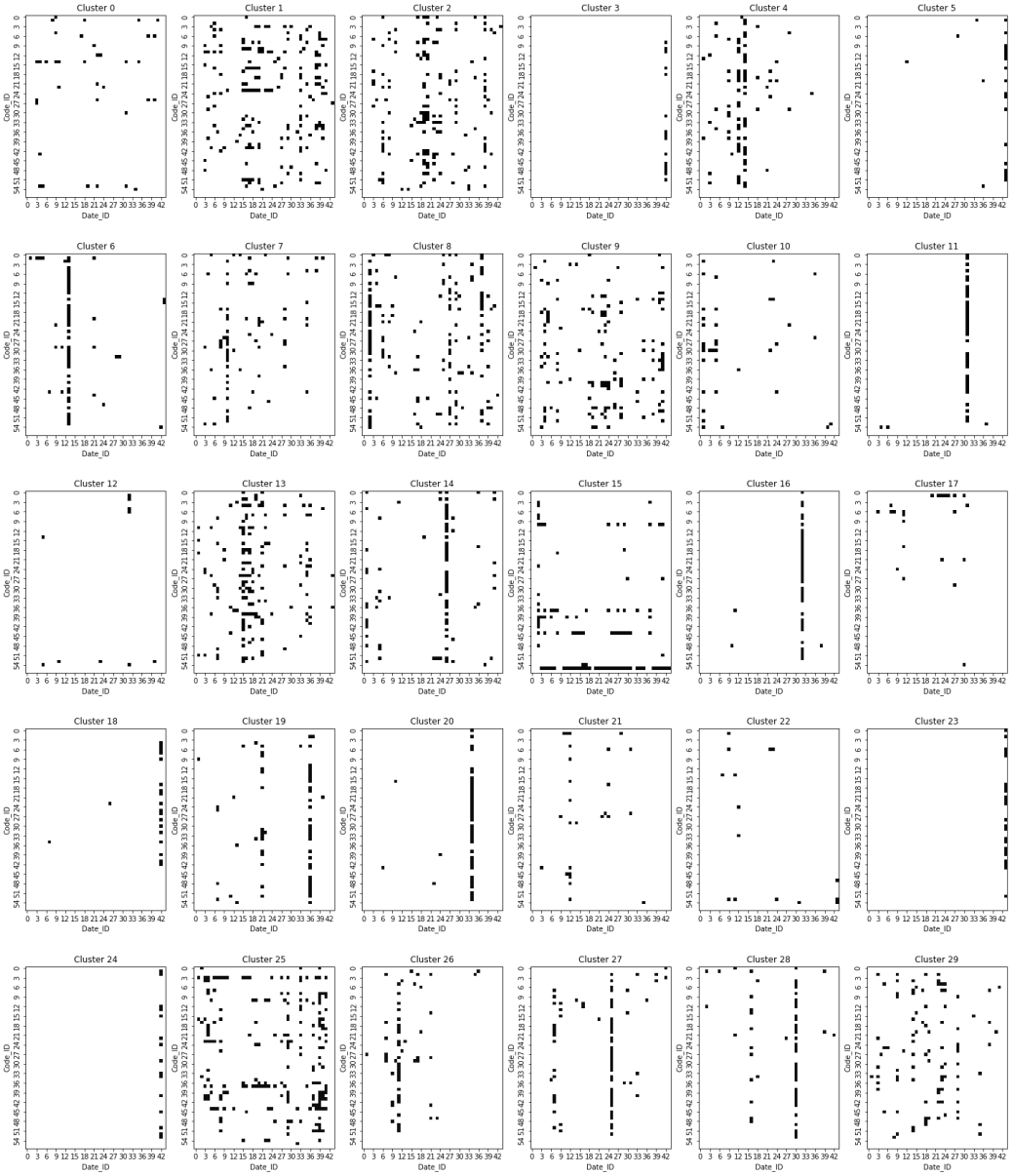

Figure S5. The spatial-temporal distribution of diurnal curves in each cluster. For each cluster, the x-axis is the date number (starting from the first day without regional events in April and ending on the last day without regional events in May), and the y-axis represents the Code for the monitoring station, ranked by their code number (small to large, e.g., "1008" in "1008C"), obeying a different order from **Figure 3.** Information could be extracted here. For example, for clusters that mainly have diurnal curves from a single day, such as Clusters 6, 11, 16, 20, 23, and 24, it may indicate regional events we didn't successfully identify.

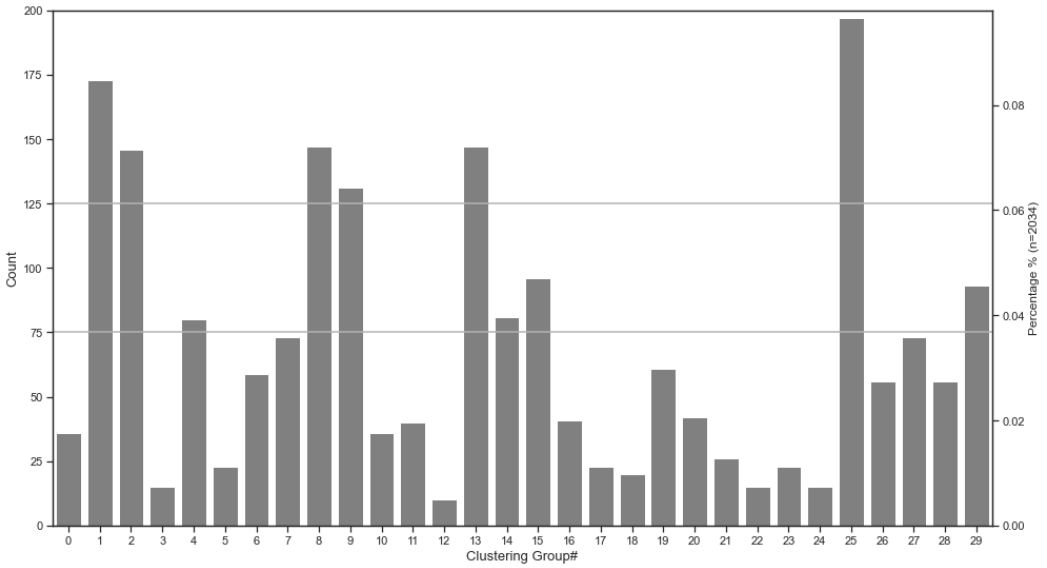

Figure S6. The sizes of clusters. Clusters 12,1 and 25 are the three largest clusters, and clusters 8, 9, and 2 are the 4th-6th, discussed in the main text and **Section S5** to identify the general diurnal pattern.

Section S5. Additional information for general pattern identification and explanation

Except 1st – 3rd largest clusters (13, 1, 25), here we also analyzed the 4th – 6th largest clusters (8, 9, 2), the relatively large clusters with a size range from 50 to 100 and their patterns (mean) to prove that a diurnal pattern with morning and evening peaks is the general pattern. The difference among these large clusters is consistent with our mixing height–traffic analysis in **Figure S7**. Furthermore, for other large clusters different from the six largest clusters, we got similar insight.

First, as shown in **Figure S8**, all clusters of large size (13, 1, 25, 8, 9, 2) exhibit some similarities in diurnal patterns of variation. As mentioned in the main body text, 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 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. The explanation from traffic and mixing height could partly interpret the formation of untypical general clusters 8 and 2. It could be found that clusters 1, 13, and 25 contain more weekdays, and clusters 2 and 8 have more weekends (see **Table S4**), and this is consistent with weekends' traffic flow shown in **Figure S9**: 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. Second, as shown in **Figure S6** and **Figure S10**, 35.8% of diurnal curves (count=728) were clustered into clusters of size larger than 50. Unlike the six largest clusters, it could be found that most of them, including clusters 4, 6, 14, 19, 26, 27, and 28, occurred 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, the shapes of the mean diurnal pattern are close to the general pattern we identified.

As a detailed version of **Figure 4(b)**, **Figure S11** shows the diurnal pattern and spatial-temporal distributions of key clusters. The largest six clusters are used as references in the figure.

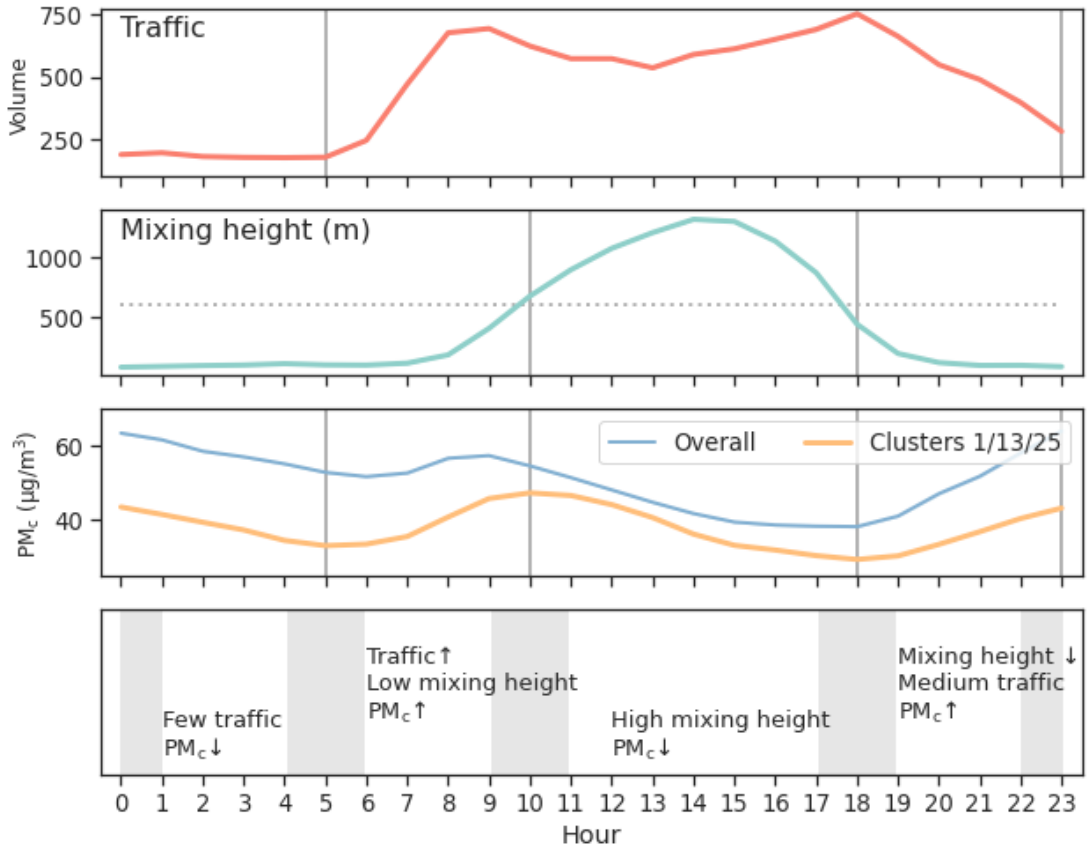

Figure S7. The traffic and mixing height variations lead to the general PM_c diurnal pattern (here represented by all sites mean and clusters of largest size 1/13/25 mean). Higher traffic volume means more emission sources, and the lower the mixing height, the longer the emissions could stay in the air. The shaded area represents the peak or" valley" time range caused by different traffic situations. The data source for mixing height here is "Boundary Layer Height" (an interchangeable term of "mixing height" in most cases) from ERA5 reanalysis data (https://www.ecmwf.int/en/forecasts/dataset/ecmwf-reanalysis-v5).

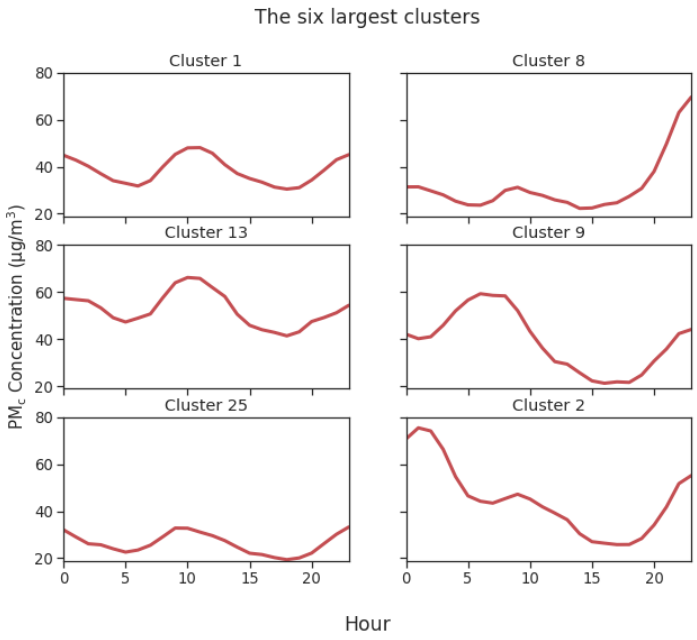

Figure S8. 1st-6th largest clusters (13, 1, 25, 8, 9, and 2). Custer 1, 13, and 25 share a very similar mean diurnal pattern with both evening and morning peaks, while clusters 8, 0, and 2 are more different. The difference among these large clusters can be partly explained by **Figure S7**, **Figure S10**, and **Table S4**, where discussion is shown in **Section S5**.

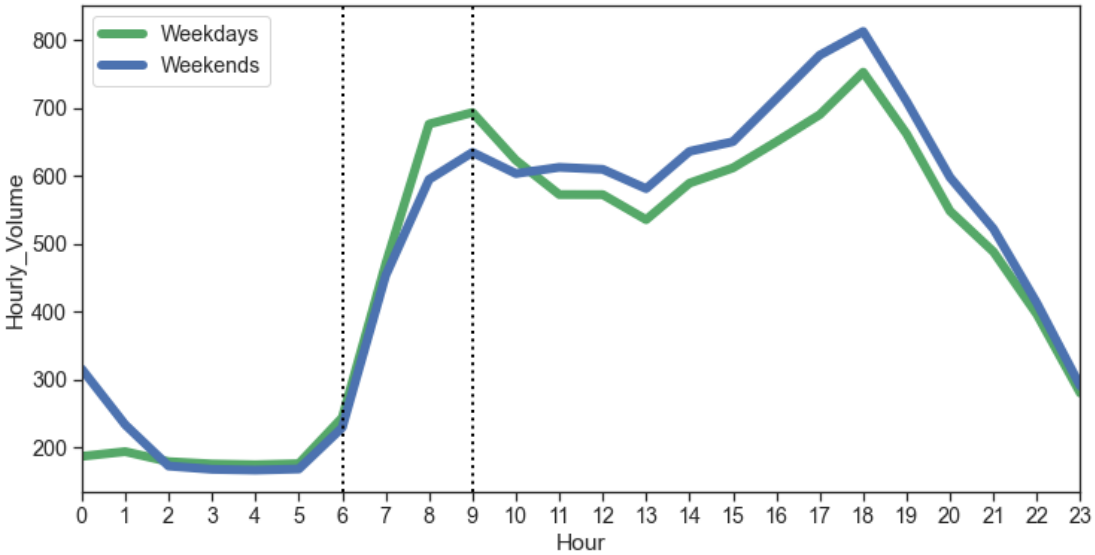

Figure S9. The average diurnal traffic volume for sample days during weekdays and weekends, respectively. Following our analysis for **Figure S7** and **S9**, this Figure, combined with **Table S4** shows why cluster 2 and cluster 8 are slightly different from clusters 1, 13 and 25. During weekends, 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.

Table S4: The week-of-day distribution for the diurnal curves in the six largest clusters. Clusters 1, 13, and 25 contain more weekdays, and clusters 2 and 8 have more weekends.

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

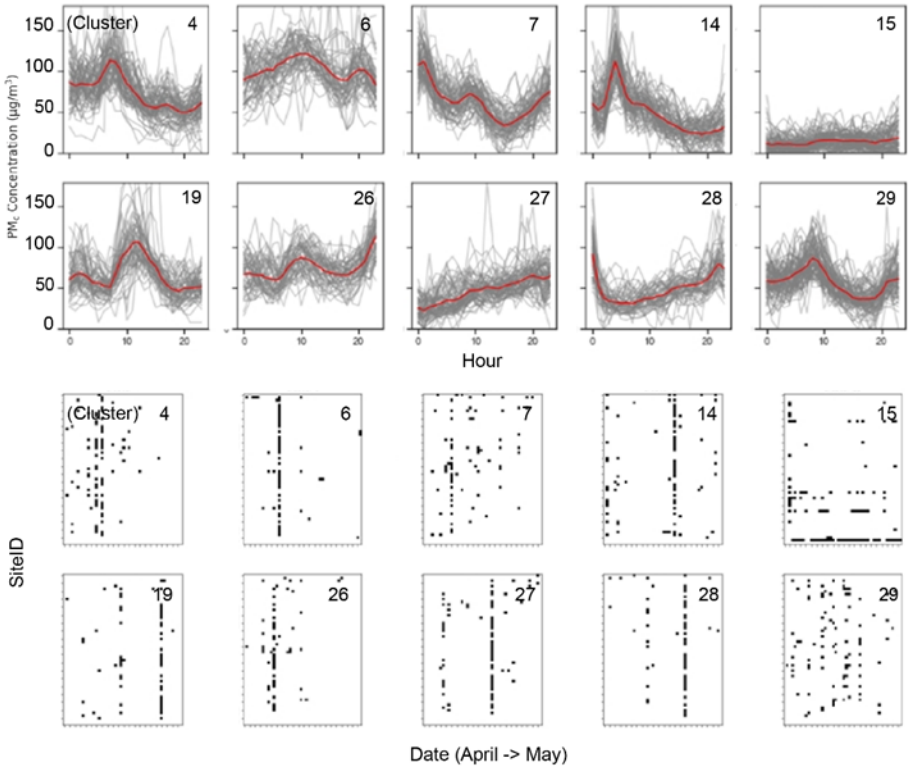

Figure S10. The diurnal curves and their spatial-temporal distributions of relatively large clusters. Most of them, including clusters 4, 6, 14, 19, 26, 27, and 28, occurred mainly in a single day or two days, indicating the potential unrecorded regional events' impact. For the clusters with a relatively scattered distribution, including cluster 7 and cluster 29, the shapes of the mean diurnal pattern are close to the general pattern we identified.

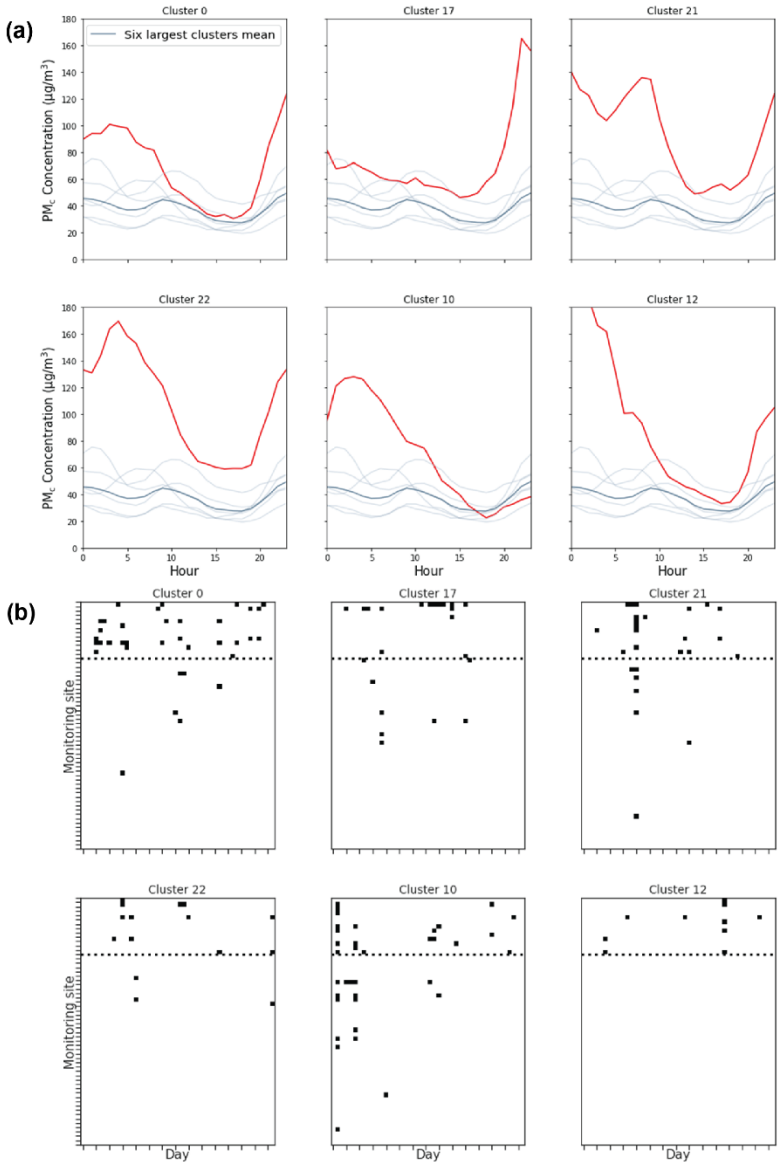

Figure S11: (a) The key clusters and (b) corresponding temporal-spatial distribution (sites above the dotted line area are hotspots). Regional events cannot explain the high PM_c concentration level for late night and early morning: these clusters mainly contain hotspots and don't contain samples mainly from a single day. The monitoring site ranking (y-axis in (b) is consistent with **Figure 3**.

Section S6. Additional information for hotspots' features and images

Table S5. 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 B, Olympics Axis Park	Foundation construction/Main body construction	Y
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 B, Public building	Main body construction 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 Residential building	Main body construction Earth excavation	 Y
1209C				Y	Metro construction Residential building	Main body construction Earth excavation	 Y

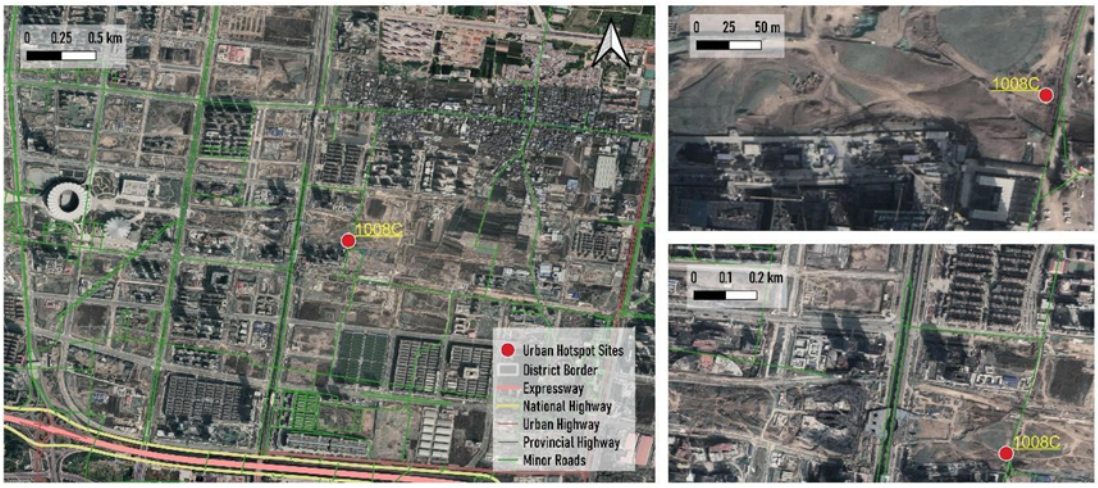

Figure S12. Satellite-based figures of construction hotspot 1008C

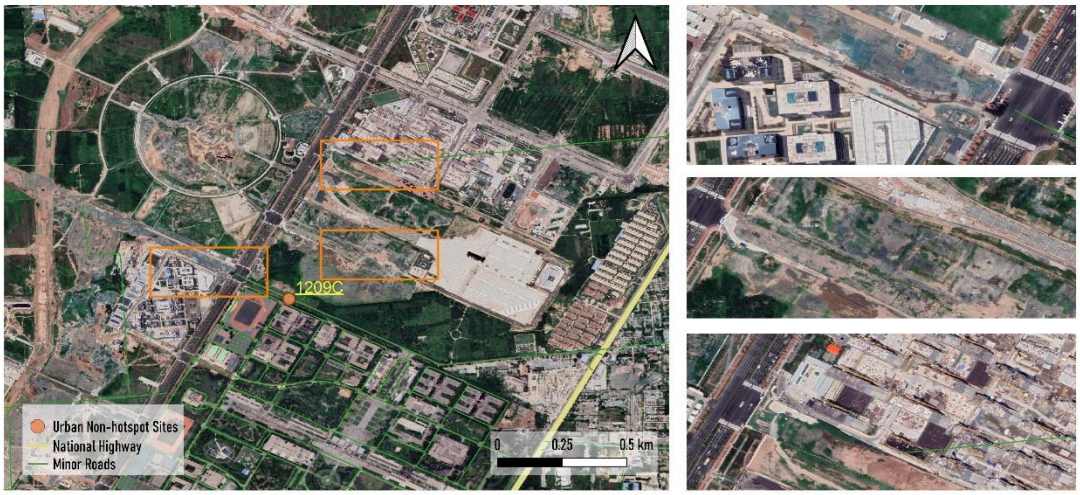

Figure S13. Satellite-based figures of construction hotspot 1209C

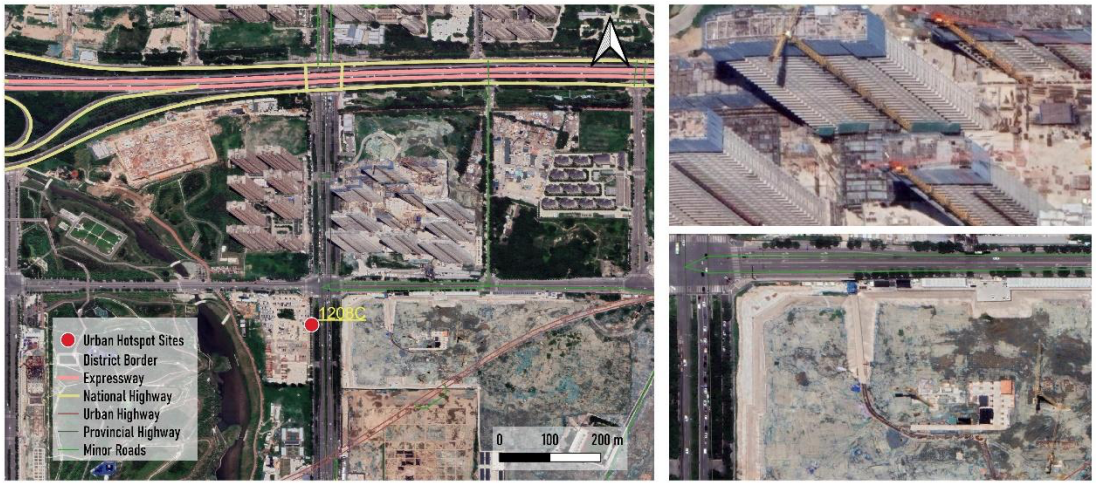

Figure S14. Satellite-based figures of construction hotspot 1208C

Figure S15. Satellite-based figures of construction hotspot 1060C

Section S7. Additional information on individual statistics data and test

Ranked by overall mean value, the mean, median, 25th percentile (Q1), 75th percentile (Q3), and the percentage of readings higher than 100 μg/m^3 and 150 μg/m^3 in daytime/nighttime for 56 urban stations are shown in **Table S7** and **S8**. Except mean shown in **Figure 6,** other station-individual statistical indicators, including median, Q1 and Q3 were all examined by Mann-Whitney U test[11]. As shown in **Figure S16,** during nighttime, construction stations had a higher mean, median, and Q3, all with significant differences (CI=99%, 95%, and 99%, respectively), while during the daytime, none of the individual statistical indicators showed a significant difference between construction stations and non-construction ones (p-value=0.13, 0.27 and 0.16,

respectively). The individual Q1 values between the two groups were less statistically significant for both time periods (p=0.06 for nighttime and 0.30 for daytime) than the median and 75th percentile, indicating that construction emission events tend to generate high-level PMc concentrations for event-corresponding times rather than largely increasing the pollution value for each time interval, which is consistent with previous studies for construction site's PM_{10} (mainly PM_c). [12,13] For each individual dataset, with the exception of the mean, at least one group (construction or non-construction stations during daytime or nighttime) did not meet the normality assumption test (Shapiro-Wilk test[14]). This test is a prerequisite for parametric tests. Therefore, in **Figure 6's inset figures** and **Figure S16**, we used the Mann-Whitney U test instead of the t-test.

There are two interesting observations that are not discussed in the main text but can support our conclusion: First, compared to the non-construction stations with close values on the mean of PM_c concentration at nighttime, many construction stations such as 1170C, 1193C, and 1209C can have a much higher percentage of readings larger than 150 $\mu g/m^3$, all of the 20%+ readings are from construction stations even if some of their mean values are not that high. Secondly, the construction stations generally have a higher ranking in nighttime compared to daytime, such as 1209C and 1193C. Both suggested the night impact of construction site activities.

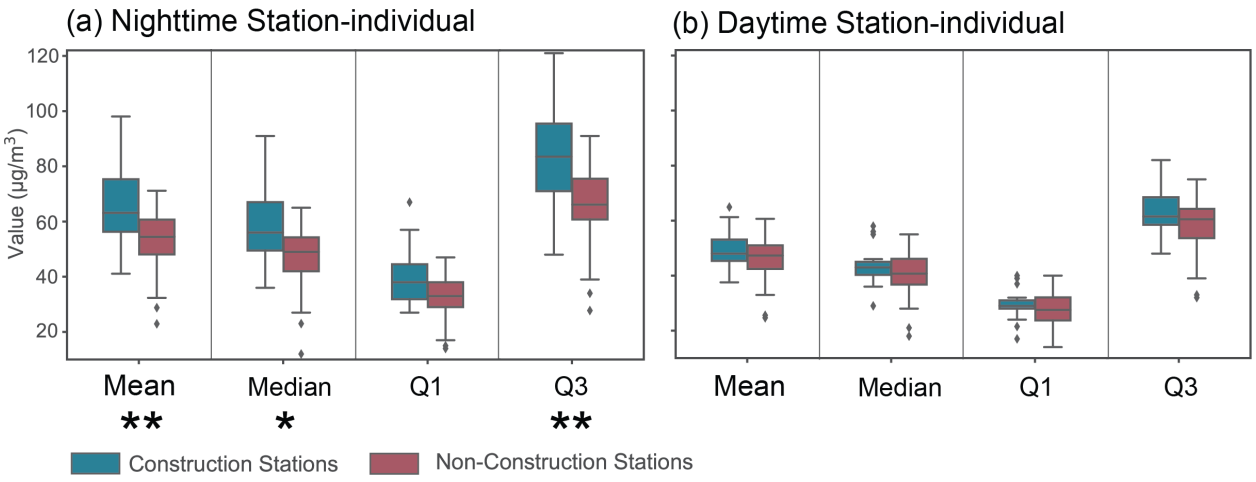

Figure S16. The station-individual mean, median, Q1, and Q3 percentile values for construction stations and non-construction stations during (a) nighttime and (b) daytime (**: p-value ≤ 0.01, *: p-value ≤ 0.05).

Table S6. The mean, median, 25th percentile (Q1), 75th percentile (Q3), the percentage of readings higher than 100 μg/m^3 and 150 μg/m^3 in daytime for 56 urban stations.

Code	Overall Mean	Construction Station	Daytime					
			Mean	Median	Q1	Q3	Proportion>100	Proportion>150
1008C	80.42	*	65.97	59.50	41.00	82.00	15.36%	2.45%
1060C	72.32	*	62.10	57.00	40.00	75.00	11.22%	1.79%
1208C	70.56	*	55.18	46.00	31.00	73.50	9.75%	1.68%
1177C	69.19	*	61.91	58.00	37.00	82.00	12.73%	2.21%
1073C	64.21	*	55.07	44.00	29.00	72.00	11.48%	1.66%
1168C	64.04		60.12	55.00	40.00	74.00	8.26%	0.66%
1058C	64.03		58.42	53.00	36.00	76.00	9.52%	1.21%
1019C	63.70		60.74	56.00	37.00	73.00	11.49%	2.36%
1170C	62.45	*	56.57	47.00	32.00	69.00	11.49%	2.75%
1193C	61.91	*	53.21	44.00	29.00	69.00	10.21%	0.81%
1174C	59.81		53.74	48.00	36.00	65.00	6.06%	1.01%
1070C	57.09		52.13	42.00	30.00	64.75	9.76%	1.79%
1187C	56.93		55.38	51.00	36.00	69.00	6.53%	0.16%
1178C	56.59		49.22	44.00	30.00	64.25	5.36%	0.32%
1201C	56.08	*	48.28	44.00	28.00	61.00	5.99%	0.44%
1176C	55.92	*	51.66	45.50	31.00	63.25	7.55%	1.17%
1181C	55.36		49.80	45.00	28.00	65.00	7.98%	0.98%
1172C	54.65	*	48.62	44.00	29.00	64.00	6.59%	1.04%
1083C	54.59		52.96	44.50	29.00	68.25	7.08%	2.36%
1175C	54.53	*	45.94	41.00	22.00	62.00	7.39%	0.86%
1192C	53.37		51.44	47.00	33.00	65.00	5.33%	0.36%
1182C	53.08		50.74	46.00	33.00	60.00	6.04%	0.82%
1200C	52.56		46.20	40.00	25.00	63.00	6.48%	0.68%
1202C	52.40		49.26	41.00	29.00	65.00	6.35%	0.71%
1180C	52.35		53.03	47.00	31.00	65.00	8.67%	0.51%
1066C	52.16		45.90	41.00	24.00	64.00	6.33%	1.30%
1207C	52.13	*	49.90	45.00	31.25	62.00	5.70%	0.49%
1031C	51.98		48.69	47.00	34.00	60.00	2.31%	0.66%
1203C	51.82	*	47.28	37.00	24.00	61.00	8.84%	0.65%
1166C	51.37		48.33	42.00	29.00	62.00	6.19%	
1204C	51.06	*	46.50	42.00	28.00	60.25	5.22%	
1064C	50.95		49.82	46.00	31.00	63.00	4.61%	0.49%
1179C	50.83		48.99	44.00	31.00	59.00	4.93%	0.33%
1071C	49.96		47.28	41.00	24.00	64.00	7.73%	1.26%
1162C	49.40		44.62	38.00	23.00	60.00	5.39%	0.65%
1209C	49.31	*	39.05	29.00	17.00	49.00	6.15%	2.11%
1173C	49.12	*	45.06	41.00	28.00	59.00	2.77%	0.40%
1151C	48.28		44.89	40.00	28.00	57.00	4.24%	

1167C	48.22		45.77	41.00	27.00	58.00	5.01%	0.65%
1069C	48.07		48.60	45.00	32.00	61.00	4.43%	0.33%
1001C	47.26		36.72	33.00	23.00	46.00	1.00%	
1184C	47.15		43.21	39.00	26.00	56.00	5.29%	0.50%
1065C	46.45		44.68	40.00	28.00	55.00	3.77%	0.17%
1190C	46.36	*	45.68	40.00	29.00	56.00	4.41%	0.65%
1169C	45.40		40.20	34.00	26.00	50.00	2.60%	0.81%
1206C	44.99		43.59	39.00	26.00	56.00	3.12%	
1191C	43.79		43.00	39.00	26.00	54.00	3.46%	0.16%
1205C	43.53		40.63	36.00	24.00	51.00	4.07%	0.51%
1185C	42.81		38.54	34.00	24.00	51.00	1.32%	
1171C	42.68	*	43.37	39.00	28.00	52.00	2.72%	0.68%
1068C	35.80		36.40	31.00	21.00	45.00	2.29%	
1189C	35.10		33.80	28.50	15.25	48.75	1.71%	0.17%
1027C	32.68		32.53	29.00	22.00	38.25	0.68%	0.51%
1186C	27.04		25.58	21.00	14.00	33.00	0.83%	
1199C	23.84		24.29	18.00	7.00	31.00	3.02%	0.84%
1211C	9.13		9.62	8.00	4.00	13.00		

Table S7. The mean, median, 25th percentile (Q1), 75th percentile (Q3), the percentage of readings higher than 100 μg/m^3 and 150 μg/m^3 in the nighttime for 56 urban stations.

Code	Overall Mean	Construction Station	Nighttime				
			Median	Q1	Q3	Proportion>100	Proportion>150
1008C	80.42	*	92.0	69.0	123.0	41.7%	12.8%
1060C	72.32	*	78.0	56.5	110.5	31.9%	6.9%
1208C	70.56	*	82.0	48.0	123.0	38.4%	15.9%
1177C	69.19	*	74.0	55.3	98.0	23.5%	5.9%
1073C	64.21	*	67.0	43.0	101.8	26.7%	6.3%
1168C	64.04		65.5	47.8	91.0	16.3%	1.8%
1058C	64.03		66.0	46.8	89.3	17.7%	3.5%
1019C	63.70		63.0	46.0	85.0	13.1%	1.9%
1170C	62.45	*	59.0	40.0	90.0	20.2%	4.3%
1193C	61.91	*	66.0	46.0	95.0	21.2%	4.5%
1174C	59.81		59.0	47.0	82.0	13.6%	1.4%
1070C	57.09		57.0	37.0	77.8	13.1%	3.8%
1187C	56.93		55.0	39.0	71.0	10.1%	0.9%
1178C	56.59		61.0	41.0	83.0	15.1%	4.0%
1201C	56.08	*	55.0	38.5	85.5	17.4%	5.8%
1176C	55.92	*	57.0	39.0	77.0	12.8%	1.9%
1181C	55.36		56.0	38.3	79.5	12.4%	2.0%
1172C	54.65	*	56.0	40.0	75.0	11.6%	4.4%
1083C	54.59		47.0	34.0	66.3	9.9%	2.7%

1175C	54.53	*	61.0	36.0	88.0	19.4%	2.5%
1192C	53.37		52.0	37.0	67.0	8.9%	1.5%
1182C	53.08		51.0	36.0	68.0	8.2%	0.9%
1200C	52.56		53.0	37.0	76.0	11.6%	3.3%
1202C	52.40		49.0	35.0	69.0	6.5%	3.6%
1180C	52.35		44.0	30.0	66.0	7.2%	0.7%
1066C	52.16		57.0	39.0	80.0	11.7%	0.6%
1207C	52.13	*	48.0	34.0	69.0	9.3%	0.7%
1031C	51.98		53.0	40.0	67.0	4.6%	1.8%
1203C	51.82	*	50.0	28.8	75.0	13.1%	2.7%
1166C	51.37		49.0	34.0	68.0	9.7%	0.9%
1204C	51.06	*	49.0	33.3	72.8	11.5%	1.5%
1064C	50.95		48.0	30.0	69.0	7.5%	0.9%
1179C	50.83		50.0	35.0	66.0	6.4%	1.1%
1071C	49.96		49.0	32.0	65.0	8.6%	2.8%
1162C	49.40		53.0	30.0	71.0	9.4%	1.4%
1209C	49.31	*	52.5	28.0	85.8	20.4%	7.1%
1173C	49.12	*	51.0	32.0	70.0	6.3%	1.4%
1151C	48.28		47.0	34.0	62.0	8.9%	1.4%
1167C	48.22		46.0	32.0	62.8	8.6%	0.9%
1069C	48.07		43.0	29.0	61.0	4.7%	0.5%
1001C	47.26		54.0	38.0	79.5	13.2%	1.8%
1184C	47.15		49.0	32.0	71.0	6.6%	1.2%
1065C	46.45		45.0	30.0	61.0	3.8%	0.7%
1190C	46.36	*	42.0	29.0	55.3	6.2%	1.4%
1169C	45.40		43.0	30.0	68.0	6.1%	2.0%
1206C	44.99			29.5	59.0	4.4%	0.7%
1191C	43.79		40.0	26.0	54.0	4.6%	0.7%
1205C	43.53		42.0	28.0	58.0	6.8%	1.0%
1185C	42.81		42.0	27.0	63.0	5.5%	0.7%
1171C	42.68	*	36.0	27.0	48.5	2.7%	1.0%
1068C	35.80		28.0	17.0	40.0	4.0%	0.5%
1189C	35.10		30.5	14.0	53.0	4.1%	1.0%
1027C	32.68		30.0	21.0	39.0	1.9%	0.7%
1186C	27.04		23.5	14.0	35.0	2.3%	0.9%
1199C	23.84		12.0	2.0	28.0	3.6%	1.1%
1211C	9.13		7.0	3.0	11.0		

Section S8. LUR Models Development

The predictor variable pre-selection and selection procedure is illustrated as follows.

1. **Pre-selection Process**: For each predictor variable represented by multiple buffer radii, we assessed their relationship with pollutant concentrations by computing the Pearson correlation coefficient (r). The buffer radius that exhibited the highest correlation for each variable was selected. Subsequent buffer radii of the same variable, if highly correlated ($r \geq 0.6$) with the initially selected radius, were eliminated. This step ensured that only the primary buffer radius and those with low correlation remained for the next stage of analysis. Additionally, variables needed at least three non-zero values to proceed before pre-selection; those failing to meet this criterion were excluded.

2. **Selection Methodology**: We implemented an exhaustive search subset selection based on the highest adjusted R^2 derived from Ordinary Least Squares (OLS) regression. The "statsmodels" package in Python was utilized to calculate the adjusted R^2 here. Our selection criteria limited the number of predictors to a maximum of one per ten samples, resulting in a cap of five predictors for our dataset. This approach diverges from others by not strictly follow the hypotheses about the predictors, recognizing that specific PM_c sources, like unpaved fields, may coexist more with areas of relatively lower population density. Only predictor combinations that yielded statistically significant p-values below 0.1 were considered, thereby excluding sets with insufficient p-values despite potentially higher adjusted R^2 values.

The four final OLS models are detailed in **Table S9**.

Table S8. Summary of predictor variables other than 'Construction Site'. Data sources for Table S8 include: Land use and partial road information from OpenStreetMap (openstreetmap.org). Road and Points of Interest (POIs) data provided by the Institute of Earth Environment, Chinese Academy of Sciences (english.ieecas.cn). The 2019 annual population density data, with a resolution of 1 km, sourced from the WorldPop project (worldpop.org).

Category	Description	Sub-category (abbreviation)	Buffer radii
Geographic	Geographic location (degrees)	Latitude, Longitude	N/A
Land Use	Square (km^2) of land cover types within a buffer radius	Grass, Park, Forest, Farmland	100m, 200m, 500m, 1000m, 1500m, 2000m, 3000m
		Residential, Industrial, Commercial, Cemetery, Quarry, Retail	100m, 200m, 500m, 1000m,

			1500m, 2000m, 3000m
Road Types	Length (km) of different types of roads within a buffer radius	City Mainway, National Highway, Provincial Highway, Other Way, Railway Cycleway, Footway, Living Street, Pedestrian, Primary, Primary Link, Residential Road, Secondary, Service, Steps, Tertiary, Track, Trunk, Trunk Link, Motorway,	100m, 200m, 500m, 1000m, 1500m, 2000m, 3000m
Points of Interest (POIs)	Number of POIs within a buffer radius	Railway, Gas Station, Traffic Light, Mall, Bus Stop, Restaurant, Building, Store, River, Lake, Reservoir, Central Park	100m, 200m, 500m, 1000m, 1500m, 2000m, 3000m
Traffic Volume	Hourly Average for Nearby Traffic Volume (if distance to any road<100m)		N/A
Distance to Road	Distance to closest road (km)	Closest National Highway Distance, Closest Provincial Highway Distance, Closest City Mainway Distance, Closest Motorway Distance, Closest Other Way Distance	N/A
Population Density	The number of people per km^2 in 2019	Population Density	N/A

Table S9. The four final OLS models.

Model	Variables	Coefficient	Standard Error	P-value	R-squared	Adj. R-squared
Nighttime With	Intercept	67.6260	3.434	0.0000	0.647	0.611
	Construction Site	8.8826	2.785	0.0020		
	Store_3000	-0.7991	0.132	0.0000		
	Forest_1000	-102.8501	20.259	0.0000		
	Service_3000	0.4712	0.141	0.0020		
	Grass_2000	-32.7693	8.544	0.0000		
Nighttime Without	Intercept	73.0033	3.005	0.0000	0.599	0.558
	Store_3000	-0.7457	0.144	0.0000		
	Forest_1000	-116.2961	20.982	0.0000		
	Primary_100	17.7311	9.998	0.0820		
	Service_3000	0.2980	0.148	0.0490		
	Grass_2000	-33.5428	9.122	0.0010		
Daytime With	Intercept	45.2206	1.651	0.0000	0.471	0.417
	Grass_2000	-19.6601	5.933	0.0020		
	Closest Provincial Highway Distance	0.0022	0.001	0.0290		

	Restaurant_100	1.963	0.537	0.0010		
	Store_200	-6.1148	2.126	0.0060		
	Construction Site	3.213	1.837	0.0870		
Daytime	Intercept	49.8038	1.675	0.0000		
Without	Forest_1000	-56.1561	11.882	0.0000		
	Grass2000	-29.7274	5.222	0.0000	0.611	0.571
	Building_3000	-0.09	0.016	0.0000		
	Retail_2000	30.6975	14.571	0.0400		
	Residentialroad_1000	0.6074	0.204	0.0040		

Section S9. Collection of Chinese news articles on dump trucks

The rapid urbanization in China is accompanied by substantial construction activities. Consequently, dump trucks, ubiquitous on construction sites, have become a focal point of media discussions and governmental regulations over the past two decades due to their noise pollution, dust pollution, countless fatal accidents, and recurring traffic and environmental violations. We selected a few representative news articles and web posts here.

S9.1. The considerations of dump trucks night-mainly policy

1. http://www.jzlj.org.cn/Item/Show.asp?m=1&d=2930
   2014-03-25

   "With the aggravation of smog hazards, urban muck trucks are getting more and more attention. In addition to pollution, rampant muck trucks also pose a threat to traffic safety. to deal with these problems. Various places have introduced measures against dump trucks."
   "Establish and expand restricted areas" - Fengbu City, Anhui Province: Prohibited vehicles are prohibited from entering prohibited areas and road sections from 4 to 21 o'clock every day"

   The first news clip from 2014 shows areas restricting dump truck operation to nighttime due to pollution and safety concerns.

2. http://jiangsu.china.com.cn/html/2016/scnews_1116/8091655.html
   2016-11-16

“Chengdu responds to heavy pollution days: starting from the 16th, dump trucks will be prohibited from passing around the city during the day”

The 2016 news excerpt reveals Chengdu's daytime ban on dump trucks in response to high pollution days, underlining environmental concerns.

3. https://news.ifeng.com/c/7fbxXOmJUpY
   2012-04-12

   “A front-line traffic police said that dump trucks are not allowed on the road during the day, mainly because of conflicts with the ban on trucks during the day, and because there are too many residents in the urban area, allowing dump trucks to drive on the road during the day is a great safety hazard.”

   The 2012 news piece shows the safety motivation behind the nighttime operation of dump trucks due to potential hazards in residential areas.

4. https://news.sina.com.cn/c/2005-11-16/06587451295s.shtml
   2005-11-6

   “At 8 o'clock every night, the dump trucks were already on the road. At this time, there were still many pedestrians and vehicles on the road, and the extremely fast speed and the flying dust they brought up made them very distressed. On the one hand, it is easy to cause collision accidents, on the other hand, it also makes people afraid to breathe, and is forced to cover their noses for a long time.”

   “The city appearance department and the transportation department are planning a "notice" to limit the time and limit the road section. For example, dump trucks are not allowed to go on the road during the day, and they can only work after 10:00 at night, and they must "stop work" before 5:00 am the next day. For example, muck trucks are not allowed on the main road with the prefix "中".”

   The 2005 news item highlights both safety and pollution issues of dump trucks and suggests that it’s important to implement daytime (including evening)restrictions.

5. http://news.cjn.cn/sywh/201210/t2070473.htm
   2012-10-11

"What if the construction period cannot keep up? He said that what needs to be done is actually very simple, as long as the dump truck is cleaned up so as not to pollute the environment, you can also transport muck during the day, which is equivalent to extending the construction period."

The 2012 news report suggests the key issue of extending dump truck operation into the daytime is dust– as the logic from the government is if you are clean enough, then you can apply for an operation in the daytime.

6. https://people.rednet.cn/front/messages/detail?id=995930
   2011-11-07

"The Changsha Municipal Government's "Notice on Prohibiting the Passage of Waste and Sand and Gravel Transport Vehicles in Some Areas of the Urban Area during the Daytime" (hereinafter referred to as the "Notice"). This document issued on April 4, 2003 stipulates that vehicles transporting muck, sand and gravel are prohibited from operating in the urban area from 4:00 to 20:00 every day, and the prohibited areas are delineated. In this way, the dump trucks in Changsha City can only work at night, and the corresponding construction sites can only be constructed at night. This document was formulated "in order to alleviate the traffic pressure in Changsha City and keep the city clean and tidy", but it was unexpected that the problem of noise pollution at night would appear immediately.

The 2011 piece talks about a daytime ban by Changsha to reduce traffic and maintain cleanliness while acknowledging the resulting nighttime noise pollution.

S9.2. The traffic accidents related to dump trucks in Xi'an City after 2011

7. https://www.sohu.com/a/425647866_99939235

At approximately 3:48 AM on October 18th 2020, a dump truck running a red light in Xi'an collided with a taxi, resulting in the death of a passenger and minor injuries to the taxi driver.

8. https://new.qq.com/rain/a/20201027A075X900

On the early morning of October 25th 2020, a netizen's dashcam captured two incidents of heavy-duty trucks, not standard dump trucks, running red lights and driving in the wrong direction on Xi'an's Xiangyun Road to Baozi Village route, posing a significant risk to road safety.

9. https://m.jiazhao.com/news42881/

On October 11, 2014, around 4 AM, a dump truck heading north on the eastern second ring road spilled a load of

mud near the Interchange Bridge, leading to multiple vehicles slipping and colliding due to the muddy conditions, causing a multi-car pile-up without any reported injuries.

10. https://m.thepaper.cn/newsDetail_forward_1753502

On the night of August 2, 2017, a 22-year-old female teacher was fatally hit by a dump truck while riding a shared bicycle in Xi'an, causing significant public concern and sparking discussions about road safety, especially during night time.

11. http://energy.people.com.cn/n/2014/0328/c71890-24766064.html

On March 13th, 2014, at approximately 1:15 am, law enforcement officers in Xi'an, including traffic police and urban management officials, carried out a surprise inspection of key areas where illegal dumping of construction waste was frequently reported. While conducting the inspection, a sedan, driven by a man named Pan, deliberately collided with the enforcement vehicles in an attempt to disrupt the operation and evade detection of his own illegal dumping activities.

12. http://xian.baogaosu.com/xinwen/%E5%87%8C%E6%99%A8%E8%A5%BF%E5%AE%89%E4%B8%9C%E4%B8%89%E7%8E%AF%E7%A6%8F%E5%85%8B%E6%96%AF%E6%B8%A3%E5%9C%9F%E8%BD%A6/31892407/

On April 17, 2015, around 4 a.m., a severe traffic accident occurred near the Banpo Overpass on the Eastern Third Ring Road in Xi'an, where a white unregistered Ford sedan collided with a dump truck, resulting in one fatality and multiple injuries in the sedan, three of them severe. The accident is currently under investigation by the police.

13. https://www.sohu.com/a/348435694_120385152

On the night of October 19th 2019, around 11 p.m., a major traffic accident involving a dump truck occurred in Xi'an when the vehicle ran a red light at the intersection of Qinhan Avenue and Xincheng Road, resulting in the destruction of a car and the tragic loss of four lives.

14. http://sx.sina.cn/news/xian/2018-04-27/detail-ifztkpin9978299.d.html?wm=3049_0016&from=qudao

On the early morning of April 20, 2018, around 4:30 a.m., a dump truck running a red light in Xi'an collided with a vehicle at the intersection of Taibei North Road and Zhaoyuanmen, resulting in six casualties and highlighting the ongoing problem of dump trucks as "road killers" in the city.

S9.3. Some policy change examples

15. https://k.sina.cn/article_2810373291_a782e4ab02001j9bj.html?from=news&subch=onews

In 2020, Changsha city in China revised its longstanding policy related to dump truck operations, which had been

in effect for about two decades. Previously, the city only allowed dump truck operations to take place at night, between 10 PM and 4 AM, to minimize dust pollution during the day. However, the nighttime operations led to increased noise pollution and were prone to traffic violations, causing public complaints.

In response, Changsha implemented a new set of regulations, referred to as the "Twelve Articles for Dust Control", which included a prohibition of dump truck operations after midnight. From January 6th, 2020, dump trucks were permitted to operate during the day, a significant shift from the previous regulation. This move was aimed at improving traffic safety, reducing nighttime noise disturbances for residents, and managing dust pollution under the daylight, thereby addressing the issues caused by the previous schedule of nighttime operations.

16. https://xian.xcar.com.cn/201112/news_354513_1.html

In 2011, Xi'an city in China enforced restrictions on dump trucks due to serious safety concerns caused by multiple fatal accidents. Measures included a temporary halt on construction waste transport, extensive driver training, and the suspension of new entrants to the industry. A daily operating ban was also enforced from 6 AM to 10:30 PM within the city's third ring road to ensure public safety and smooth traffic flow.

17. https://www.sohu.com/a/301775233_336604

In March 2019, citizens in Xi'an noticed dump trucks operating during daylight hours, which prompted inquiries as it was unusual. The local government explained that post the "winter defense period", the operational hours for construction waste transportation within the third ring road had been adjusted to 10 AM to 4 PM and 10:30 PM to 6 AM the next day. Outside of the third ring road, waste transportation was permitted 24/7. This policy change resulted from a substantial increase in construction sites and a corresponding rise in waste output, combined with reduced effective waste transportation times.

Notes: according to information from the local forum and examples such as “Damaocheng” in section 2.2.3, it’s highly likely that dump trucks still face many limitations during the daytime after 2019.

18. https://sd.ifeng.com/a/20190629/7568405_0.shtml

In June 2019, the Jinan Traffic Police Brigade introduced a new policy to allow construction dump trucks to operate during the day. The policy, known as the "Six Public Systems" for major construction waste transportation projects, aimed to address issues of nighttime noise and disturbance caused by these vehicles. As a pilot, the policy was first applied to the construction project of the Comprehensive Building of the Western Hospital of Shandong University of Traditional Chinese Medicine. The policy aimed to alleviate issues of noise disturbance and air pollution by extending dump truck operating hours to avoid concentrated nighttime transit. The local community is encouraged to monitor and report any violations. This change was necessitated by a rapid increase in construction activities in Jinan, with dump truck transportation being a crucial part of urban development.

19. http://www.xdkb.net/p1/nj/20200629/97695.html

In June 2020, the Qinhuai District in Nanjing started to explore a policy change to reduce noise complaints from nighttime construction waste transportation. They plan to allow selected construction sites to operate during the day, provided they meet certain management criteria. Alongside this initiative, a reporting hotline has been established for residents to report violations, incentivized by rewards for confirmed cases.

Table S10 The examples of current dump truck policy for some large cities. A daytime ban (Night ONLY) or designate nighttime as the primary operating window for dump trucks is common.

Type	City	Urban population 2020 (million)	Updated time[d]	Operating time for urban area[a]
Night ONLY	Beijing	17.75	2020	23:00-6:00[b]
	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 A: 21:00-7:00 Part B: 24h ban
Night as Main Window[c]	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
Daytime and Evening	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
Mixture	Zhengzhou	5.34	2019	Urban-center area: 22:00-5:00
				Other urban areas: 9:00-20:00
	Luoyang	4.58	2020	October->March: 9:00-17:00
				April->September: 20:00-7:00

[a] Generally, it's the center-urban area which often refers to the area inside the third/fourth/fifth ring road.

[b] Policy also works for other heavy vehicles.

[c] Limitations during the daytime may still exist under multiple circumstances. For example, the example of "Damaocheng" in Section 2.2.3

[d] If there are no documents found, it means the year of reporting the latest evidence

(1) Arthur, D.; Vassilvitskii, S. K-Means++: The Advantages of Careful Seeding. *Proc. Annu. ACM-SIAM Symp. Discret. Algorithms* **2007**, *07-09-Janu*, 1027–1035.

(2) Xu, R.; Wunsch, D. C. Partitional Clustering. *Clustering* **2009**, 63–110. https://doi.org/10.1002/9780470382776.CH4.

(3) Teichgraeber, H.; Brandt, A. R. Clustering Methods to Find Representative Periods for the Optimization of Energy Systems: An Initial Framework and Comparison. *Appl. Energy* **2019**, *239*, 1283–1293. https://doi.org/10.1016/J.APENERGY.2019.02.012.

(4) Thorndike, R. L. Who Belongs in the Family? *Psychom. 1953 184* **1953**, *18* (4), 267–276. https://doi.org/10.1007/BF02289263.

(5) *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 2022-03-09).

(6) *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 2022-07-26).

(7) Chen, D.; Cheng, S.; Zhou, Y.; Guo, X.; Fan, S.; Wang, H. Impact of Road Fugitive Dust on Air Quality in Beijing, China. *https://home.liebertpub.com/ees* **2010**, *27* (10), 825–834. https://doi.org/10.1089/EES.2009.0122.

(8) Zhou, Z.; Tan, Q.; Deng, Y.; Wu, K.; Yang, X.; Zhou, X. Emission Inventory of Anthropogenic Air Pollutant Sources and Characteristics of VOCs Species in Sichuan Province, China. *J. Atmos. Chem. 2019 761* **2019**, *76* (1), 21–58. https://doi.org/10.1007/S10874-019-9386-7.

(9) Singh, V.; Biswal, A.; Kesarkar, A. P.; Mor, S.; Ravindra, K. High Resolution Vehicular PM10 Emissions over Megacity Delhi: Relative Contributions of Exhaust and Non-Exhaust Sources. *Sci. Total Environ.* **2020**, *699*, 134273. https://doi.org/10.1016/J.SCITOTENV.2019.134273.

(10) Li, T.; Dong, W.; Dai, Q.; Feng, Y.; Bi, X.; Zhang, Y.; Wu, J. Application and Validation of the Fugitive Dust Source Emission Inventory Compilation Method in Xiong'an New Area, China. *Sci. Total Environ.* **2021**, *798*, 149114. https://doi.org/10.1016/J.SCITOTENV.2021.149114.

(11) Mann, H. B.; Whitney, D. R. On a Test of Whether One of Two Random Variables Is Stochastically Larger than the Other. *https://doi.org/10.1214/aoms/1177730491* **1947**, *18* (1), 50–60. https://doi.org/10.1214/AOMS/1177730491.

(12) Font, A.; Baker, T.; Mudway, I. S.; Purdie, E.; Dunster, C.; Fuller, G. W. Degradation in Urban Air Quality from Construction Activity and Increased Traffic Arising from a Road Widening Scheme. *Sci. Total Environ.* **2014**, *497–498*, 123–132. https://doi.org/10.1016/j.scitotenv.2014.07.060.

(13) Fuller, G. W.; Green, D. The Impact of Local Fugitive PM 10 from Building Works and Road Works on the Assessment of the European Union Limit Value. *Atmos. Environ.* **2004**, *38* (30), 4993–5002. https://doi.org/10.1016/j.atmosenv.2004.06.024.

(14) Shapiro, S. S.; Wilk, M. B. An Analysis of Variance Test for Normality (Complete Samples). *Biometrika* **1965**, *52* (3/4), 591. https://doi.org/10.2307/2333709.