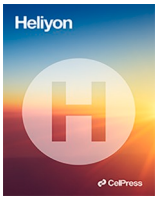
Heliyon 10 (2024) e29077

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

Novel approach for identifying VOC emission characteristics based on mobile monitoring platform data and deep learning: Application of source apportionment in a chemical industrial park

Deji Jing [a], Kexuan Yang [a], Zhanhong Shi [a], Xingnong Cai [a], Sujing Li [a], Wei Li [a,*], Qiaoli Wang [b,**]

[a] Key Laboratory of Biomass Chemical Engineering of the Ministry of Education, Institute of Industrial Ecology and Environment, College of Chemical and Biological Engineering, Zhejiang University (Zijingang Campus), Hangzhou, 310058, China
[b] College of Environment, Zhejiang University of Technology, Hangzhou, 310032, China

ARTICLE INFO

Keywords:
VOC emission characteristics
Mobile monitoring platform
Deep learning
Source apportionment
Chemical industrial park

ABSTRACT

Refined volatile organic compound (VOC) emission characteristics are crucial for accurate source apportionment in chemical industrial parks. The data from mobile monitoring platforms in chemical industrial parks contain pollution information that is not intuitively displayed, requiring further excavation. A novel approach was proposed to identify VOC emission characteristics using the class activation map (CAM) technology of convolutional neural network (CNN), which was applied on the mobile monitoring platform data (MD) derived from a typical fine chemical industrial park. It converts a large amount of monitoring data with high spatiotemporal complexity into simple and interpretable characteristic maps, effectively improving the identification effect of VOC emission characteristics, supporting more accurate source apportionment of VOC pollution around the park. Using this method, the VOC emission characteristics of eight key factories were identified. VOC source apportionment in the park was conducted for one day using a positive matrix factorization (PMF) model and seven combined factor profiles (CFPs) were calculated. Based on the identified VOC emission characteristics, the main pollution sources and their contributions to surrounding schools and residential areas were determined, revealing that one pesticide factory (named LKA) had the highest contribution ratio. The source apportionment results indicated that the impact of the chemical industrial park on the surrounding areas varied from morning to afternoon, which to some extent reflected the intermittent production methods employed for fine chemicals.

1. Introduction

In China, the $PM_{2.5}$ pollution remains severe and O_3 pollution is becoming increasingly prominent. Many studies have shown that VOCs are not only the primary precursors of $PM_{2.5}$ but also as the principal determinants of O_3 formation [1]. Therefore, it is crucial undertake substantial efforts to mitigate VOC emissions in order to enhance air quality.

* Corresponding author.
** Corresponding author.
E-mail addresses: w_li@zju.edu.cn (W. Li), wangqiaoli@zjut.edu.cn (Q. Wang).

https://doi.org/10.1016/j.heliyon.2024.e29077
Received 3 March 2024; Received in revised form 11 March 2024; Accepted 29 March 2024
Available online 4 April 2024

The chemical industry is a significant contributor to anthropogenic VOC emissions in China [2–6]. In 2010, VOC emissions from the chemical industry reached 2.63×10^6 t, accounting for 13.2% of those from industrial sources [7]. Chemical factories are generally clustered in industrial parks based on policies. According to the China Petroleum and Chemical Federation, over 600 national key chemical parks were dominated by the petroleum or chemical industry in 2020, with high-density factories and numerous pollution emission sources [8]. Chemical production is characterized by numerous product categories, intricate processes, a diverse range of raw materials and auxiliary substances. Complex VOC emissions from chemical industrial parks create favorable conditions for atmospheric chemical reactions, causing serious secondary pollution represented by $PM_{2.5}$ and O_3 in parks and surrounding areas. Therefore, reducing VOC emissions from the chemical industry is crucial for improving air quality in China. However, the emission characteristics and sources remain insufficiently identified, necessitating the adoption of scientific methods to accurately trace VOC emission sources in chemical parks.

Numerous studies have been conducted to investigate VOC source apportionment in large-scale regions such as provinces and cities [9–12]. However, the accuracy of source identification remains insufficient [13,14] due to the integrating disposal of industrial sources as one source, which is obviously unsuitable for chemical parks with a number of different emission sources. Limited studies applied receptor models and atmospheric dispersion models based on source profiles and emission inventories in chemical industrial parks. Yinzhi Huang [15] evaluated an emission profile-based source apportionment method, based on comparing similarities between source profiles and ambient air profiles, in a petrochemical industrial park. Dongfang Wang [16] conducted VOC source apportionment in a chemical park by coupling VOC source profiles, gas chromatography mass spectrometry (GC-MS) online measurement and Lagrange atmospheric dispersion model. Doudou Feng [17] conducted VOC source apportionment in a fine chemical park using PMF model combined with conditional bivariate probability functions, based on online VOC monitoring data and meteorological data. For receptor models, comparing the emission characteristics of VOCs is a key step in identifying pollution sources; however, the refined source profiles and emission inventories are extremely difficult to obtain, and there is a lack of emission characteristics that can distinguish various chemical production projects within the park. For atmospheric dispersion models, they generally do not consider chemical mechanisms due to the significant computational difficulties caused by complex and rapid chemical reactions at the park scale; however, in order to reflect the actual pollution process, simplified characteristics species would be beneficial for the models to call upon chemical mechanisms. In general, regardless of those methods, refined VOC emission characteristics are important prerequisites for performing accurate source apportionment.

VOC source profiles are important datasets describing emission characteristics, and relatively fine VOC source profiles have been established for petrochemical production, surface coating, solvent usage, and other industrial processes [18,19]. Combined with production process analysis, the refined emission characteristics of pollution sources in industrial parks can be identified. For example, Nana Cheng [20] employed manual sampling to construct the VOC source profiles of five production lines and two postprocessing units in a pharmaceutical factory. However, manual sampling is expensive and vulnerable to environmental fluctuations and human factors, resulting in small datasets and large uncertainty in source profiles [21], thereby complicating the extraction of accurate emission characteristics. Researchers have conducted static modeling on a large amount of monitoring data, establishing data-driven VOC source apportionment methods. They mainly use fixed or mobile gas sensors, utilizing pollutant concentration monitoring data, meteorological data and geographic location data, combined with atmospheric dispersion models, based on optimization theory, Bayesian theory, autonomous source search, etc., to directly estimate source parameters including emission source strength and source location [22–25]. For example, Bing Wang [26] developed an approach integrated of gas detectors, artificial neural network and gas dispersion models to achieve the real-time estimation of hazardous gas dispersion in chemical industries. Denglong Ma [27] proposed a method based on Tikhonov regularization coupled with particle swarm optimization algorithm. However, the key challenges of such methods are data authenticity and model interpretation. The emergence of VOC mobile monitoring platform technology in recent years has facilitated simultaneous monitoring of VOCs in the surrounding atmosphere using a mobile platform such as a vehicle [21]. This technology has considerably improved the temporal and spatial resolution of monitoring data and has been widely used in VOC monitoring of industrial parks. MD contain pollution information that is not intuitively displayed, requiring further excavation. In our previous study [28], we defined the process of VOC emission characteristic identification as selecting the features that gave the highest classification accuracy in machine learning. Specifically, we determined that VOC emission characteristics were specific species, which significantly enabled the classification of emission sources. CNN is suitable for visualizing VOC characteristics because it learns from the representation of visual concepts. For example, Yanlong Liu [29] utilized CNN and Fourier-transform infrared spectroscopy for spectral classification of large-scale blended microplastics. Weixiao Cheng [30] used CNN to classify the bioactivity of 3486 perfluoroalkyl and polyfluoroalkyl substances. In the field of atmospheric environmental science, CNNs have been applied in exposure assessment and concentration estimation of particulates [31,32]. For instance, Junshi Xu [33] used real-time street-level images paired with air quality data to predict short-term exposure to ultrafine particulate matter. CAM is a recently developed effective CNN visualization technology, which creates a two-dimensional fractional grid related to a specific output class [34]. The CAM calculates the position of any image and indicates the importance of each position to the class, which can be compared to the importance of VOC species for pollution source types.

Herein, a novel approach for VOC emission characteristic identification based on MD and a deep learning technique was proposed. Taking a typical fine chemical industrial park as the study area, the emission characteristics of eight key factories in the park were identified. VOC source apportionment in the park was conducted using PMF, which is a widely used source apportionment method with simple operation and a mature evaluation system. The results illustrated that this approach successfully exploited the emission characteristics and was applicable to exKisting source apportionment models. The study findings will support the development of VOC controls and air quality improvements in chemical industrial parks.

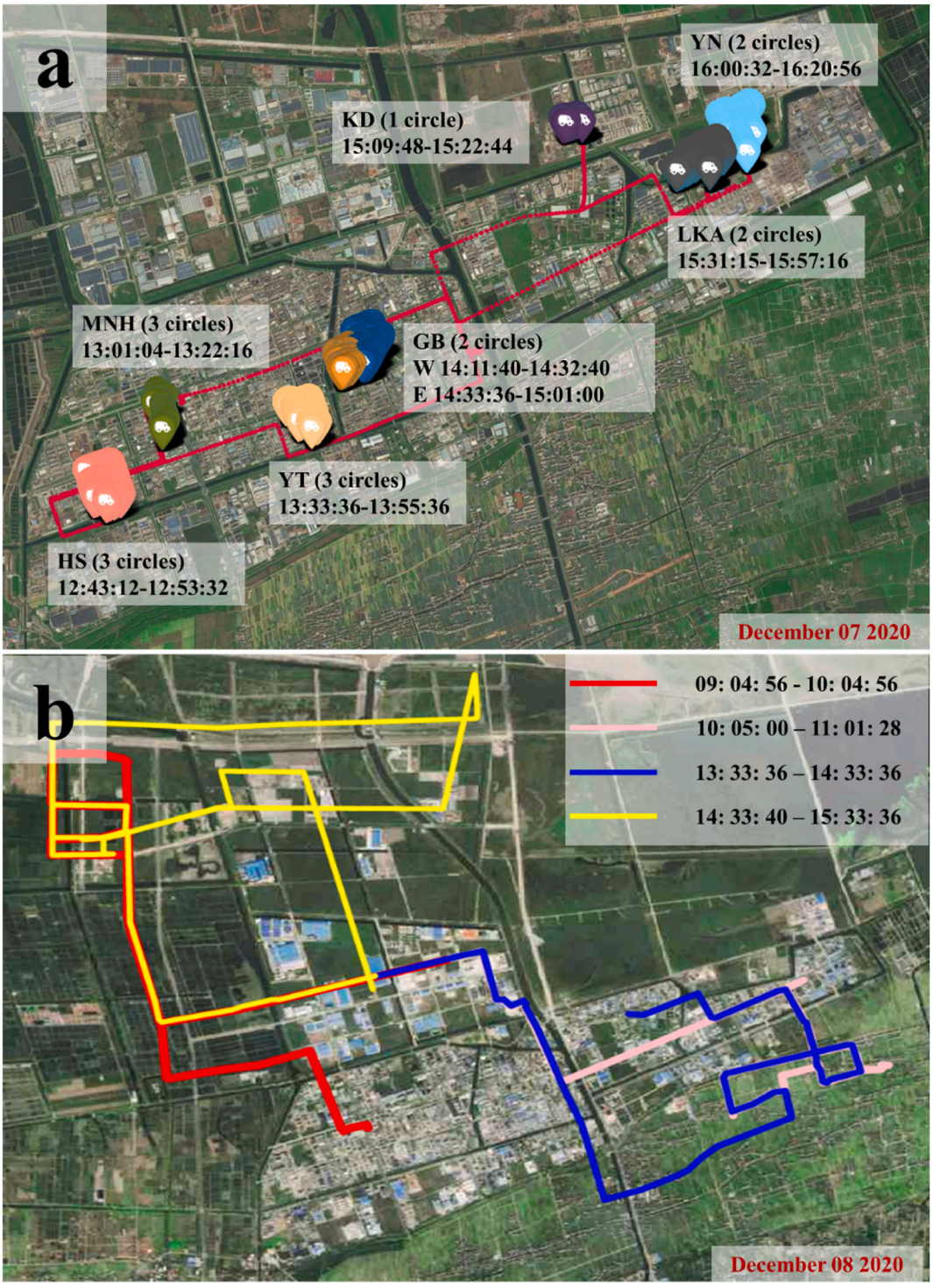


Fig. 1. Driving paths of the vehicle-mounted Proton Transfer Reaction Time-of-Flight Mass Spectrometer (PTR-ToF-MS). (a) Driving path in the park; (b) Driving path around the park.

2. Methodology

2.1. Study area and VOC sampling

A chemical industrial park located near Hangzhou Bay, China, with an annual output of over 100 billion yuan, was selected for this study. The park is dominated by the dye and pharmaceutical industries and has become the world's largest manufacturing base for disperse dye and quinolone (an antibiotic). The park selected for this study is one of the 15 key fine chemical industrial parks in China, with typical representativeness. The study employed a vehicle-mounted Proton Transfer Reaction Time-of-Flight Mass Spectrometer (PTR-ToF-MS) to obtain the VOC MD of the chemical park and surrounding areas. PTR-ToF-MS uses multiple ion sources to perform soft chemical ionization on VOC samples, including hydrocarbons, oxygen-containing hydrocarbons, halogenated hydrocarbons, nitrogen hydrocarbons, sulfur hydrocarbons, and low boiling polycyclic aromatic hydrocarbons. Its measurement principle relies on proton transfer reactions between precursor ions (H_3O^+, NO^+, O_2^+) and organic molecules. This process converts organic species into ions, which are subsequently quantified using mass spectrometry to determine the absolute concentration of organic compound. Based on the physical and chemical properties of the measured VOC, it automatically switches between H_3O^+, NO^+, and O_2^+ to achieve effective ionization of the measured object, achieving the purpose of measuring short chain alkanes and chlorinated hydrocarbons. The equipment has high sensitivity and quality resolution, and can obtain a set of VOCs concentration data in the atmosphere every 4 s. A

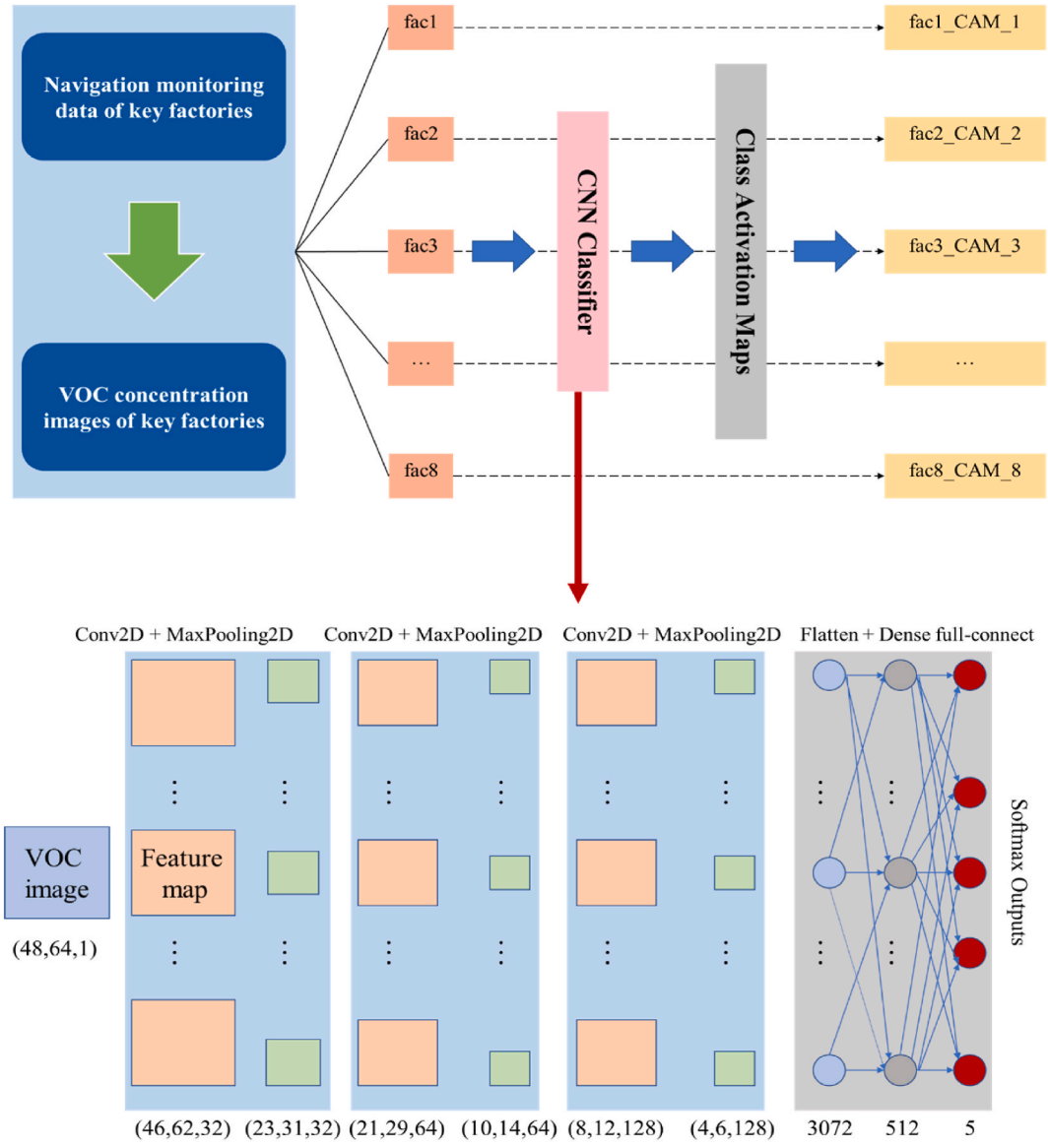


Fig. 2. Flowchart of the VOC emission characteristic identification method.

total of 48 VOC species were monitored, and detailed information was listed in Table S1. The driving path and time were shown in Fig. 1, which considered the key factories in the park as well as schools and residential areas around the park. In the park, in order to collect the VOC emission data closer to the real source, the data collected on the main road which is far from factories were not adopted. It is worth noting that the vehicle carrying a PTR-ToF-MS entered the factories and collected the VOC samples in close proximity to the production facilities. The driving speed remained below 5 $km \cdot h^{-1}$, minimizing the impact of vehicle driving on data accuracy.

The MD of key factories (MD_{KF}) in the park were obtained on December 7, 2020, for eight factories marked as HS, MNH, GB-W, GB-E, LKA, KD, YN, and YT, respectively. HS mainly produces small molecule drugs, including pharmaceutical intermediates and active pharmaceutical ingredients (APIs) for the treatment of AIDS, hepatitis C, and diabetes. MNH mainly produces pharmaceutical raw materials, including valsartan, candesartan ester, perindopril tert-butylamine, *etc*. GB develops and produces pharmaceutical and veterinary APIs, especially quinolone, macrolides, cephalosporin, and antineoplastic drugs, accounting for approximately 30% of the domestic and international market for enrofloxacin, azithromycin, and roxithromycin. GB had two factories, one in the west and one in the east. LKA mainly produces special surfactants and their derivatives, including agricultural chemicals, intermediates, polymers, catalysts, and additives. KD manufactures special surfactants and polyethers, including polypropylene and polyethylene glycol series, agricultural milk 600 series, dye intermediate series, *etc*. YN is a company that produces and processes pesticides, fine chemical intermediates, and herbicides. It also manufactures pesticide technical drugs, safety agents, and preparations. Some of the products it produces include glufosinate, aminochloropyridic acid, dichloropyridic acid, chloorofluoropyroxyacetate, alkynooxalic, herbicide water, herbicide emulsifiable concentrate, herbicide suspension agent, *etc*. YT mainly produces methoxy group acetone, 97% isopropyl alachlor, 97% dichloropropane, *etc*.

The MD of schools and residential areas (MD_{SR}) around the park were obtained on December 8, 2020. The sampling times were 09:04:56–11:01:28 and 13:33:36–15:33:36. Many atmospheric-sensitive locations exist around the park, including >10 villages in the southeast with >30,000 residents and several schools in the northwest.

2.2. VOC emission characteristic identification

The approach proposed in this study is illustrated in Fig. 2. The MD_{KF} were first converted into VOC images, followed by the construction of a CNN multiclassifier for the key factories. The CAMs were then calculated for each factory and were analyzed to obtain the VOC emission characteristics.

2.2.1. Construction of CNN

The CNN construction process was performed using Keras, presently one of the most popular artificial neural network libraries. As shown in Table 1, the MD_{KF} were divided into eight classes according to the actual factory categories, which were further labeled 0–7, with 70% and 30% of the samples divided for training and validation, respectively.

The training and evaluation process of the CNN classifier was as follows: (a) Constructed the training sample set in 55 batches, with each batch containing 30 randomly selected training samples. Input these samples into the CNN architecture, computed the trainable parameters using gradient descent, and completed one epoch of coverage over the training sample set. This process yielded the CNN classifier model. Calculated the model's classification accuracy and loss function value on the training set. (b) Constructed the validation sample set in 23 batches, with each batch containing 30 randomly selected validation samples. Input the validation sample set into the CNN classifier constructed in step (1) to compute the model's classification accuracy and loss function value. (c) Iterated steps (1) and (2) for 30 rounds to obtain the final CNN classifier, along with its classification accuracy and loss function values on both the training and validation sets.

As a CNN is especially effective in image input processing, the MD_{KF} was converted into VOC weight percentage images (WPIs) using opencv, a classic specialized library in computer vision. Each image was divided into 6×8 grids, with each grid representing a VOC species weight percentage. The position of each VOC in the image was arranged according to the average weight percentage of all samples from high to low and from the center to the surrounding areas. An example of a VOC WPI was shown in Fig. 3. The lighter colored grid indicated a higher weight percentage of the VOC species, while a darker colored grid indicated a lower weight percentage. Preliminary experiments showed that this approach distributed the characteristic species as much as possible in the center of the image, which was beneficial for VOC emission characteristic identification without affecting the credibility and accuracy of the results.

Table 1
Pretreatment of MD_{KF} for CNN classifier.

Factory	Number of samples	Number of training samples	Number of validation samples	Label
HS	151	106	45	0
MNH	312	218	94	1
GB-E	403	282	121	2
GB-W	305	214	91	3
KD	189	132	57	4
LKA	382	267	115	5
YN	304	213	91	6
YT	317	222	95	7
Sum	2363	1654	709	\

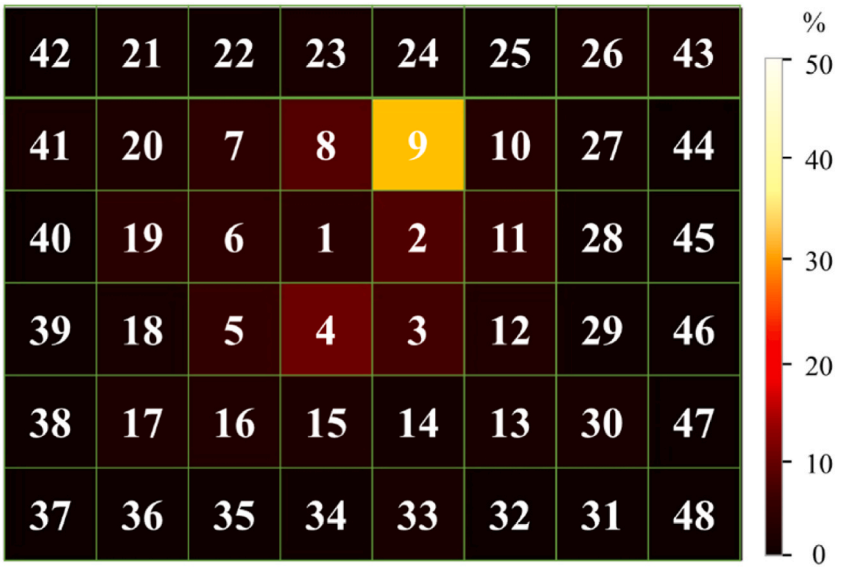


Fig. 3. Example of a VOC weight percentage image converted from $MD_{KF.}$ (1–5: chloroform, ethyl acetate, acetone, ethyl formate, and pentene; 6–10: hexane, vinyl acetate, xylene, butane, and methylbenzene; 11–15: methyl cyclopentane, acetaldehyde, diethyl sulfide, chloromethane, and butene; 16–20: cyclohexanone, isoprene, formaldehyde, 1,3-butadiene, and benzene; 21–25: methyl mercaptan, pentane, dichloroethane, dichlorodifluoromethane, and acetonitrile; 26–30: dichloropropene, trimethylbenzene, chloroethylene, methyl chloroform, and acrolein; 31–35: diethylbenzene, butenal, chlorobenzene, dichlorobenzene, and dimethyl sulfide; 36–40: tetrachloroethylene, methyl methacrylate, trichlorobenzene, styrene, and trichlorofluoromethane; 41–45: carbon tetrachloride, trichloro ethylene, carbon disulfide, hexanal, and methyl bromide; 46–48: dibromochloromethane, tetrachloroethane, and dibromoethane. The VOC species numbers are consistent with the following text.)

The CNN's specific parameters were shown in Table S2, comprising three combinations of convolutional and max-pooling layers, one flattened layer, and two dense fully connected layers, with a total of 1,669,189 parameters for training. As shown in Fig. 3, the WPI is processed into a 48 × 64 two-dimensional tensor to input the network. The 1st layer is a convolutional layer that using 2 × 2 convolutional kernel, followed by a max pooling layer of 2 × 2 grid. The 3rd to 6th layers consist of the same combination of convolutional layers and pooling layers. The 7th layer is a smooth connection layer that flattens the output two-dimensional tensor into a one-dimensional column vector. The 8th and 9th layers are fully connected layers, forming a dense connected classifier network. In terms of activation function selection, the 1st to 8th layers choose the Relu activation function, and the 9th layer chooses the Softmax function. The final output of the network is an 8 × 1 column vector, which uses one-hot encoding to output the sample classes.

2.2.2. *Characteristic identification by CAM*

The CAM method employed in this study was Grad-CAM, which was proposed by Selvaraju et al., in 2017. Grad-CAM uses the gradients of any target concept, i.e., the class of factory, flowing into the final convolutional layer to produce a coarse localization map highlighting important regions in the image, i.e., the VOC species characteristics, in order to predict the concept.

The ultimate goal is to obtain the class-discriminative localization map Grad-CAM $L^c_{Grad-CAM} \in \mathbb{R}^{u \times v}$ of height u and width v for any class c, where u and v represent the grid dimensions of the input VOC image, which were 6 and 8, respectively, in this study.

First, the gradient of the score for class c, y^c, before the Softmax, with respect to feature map A^k of a convolutional layer, i.e., $\frac{\partial y^c}{\partial A^k}$, was computed. Then, these gradients flowing back were global-average-pooled to obtain the neuron importance weight, α^c_k, as follows:

$$\alpha^c_k = \frac{1}{z}\sum_i \sum_j \frac{\partial y^c}{\partial A^k_{ij}} \tag{1}$$

where z refers to the number of feature maps, i refers to the ith row, and j refers to the jth column. This weight, α^c_k, represents a partial linearization of the deep network downstream from A and captures the "importance" of feature map k for target c. Finally, we performed a weighted combination of forward activation maps, followed by application of a rectified linear unit (ReLU) to obtain:

$$L^c_{Grad-CAM} = \mathrm{ReLU}\left(\sum_k \alpha^c_k A^k\right) \tag{2}$$

2.3. *VOC source apportionment*

The EPA PMF5.0 receptor model was employed in this study. In previous PMF source apportionment studies, large time ranges were often selected for calculation due to the difficulty of obtaining monitoring data. The data were usually derived from one or several fixed

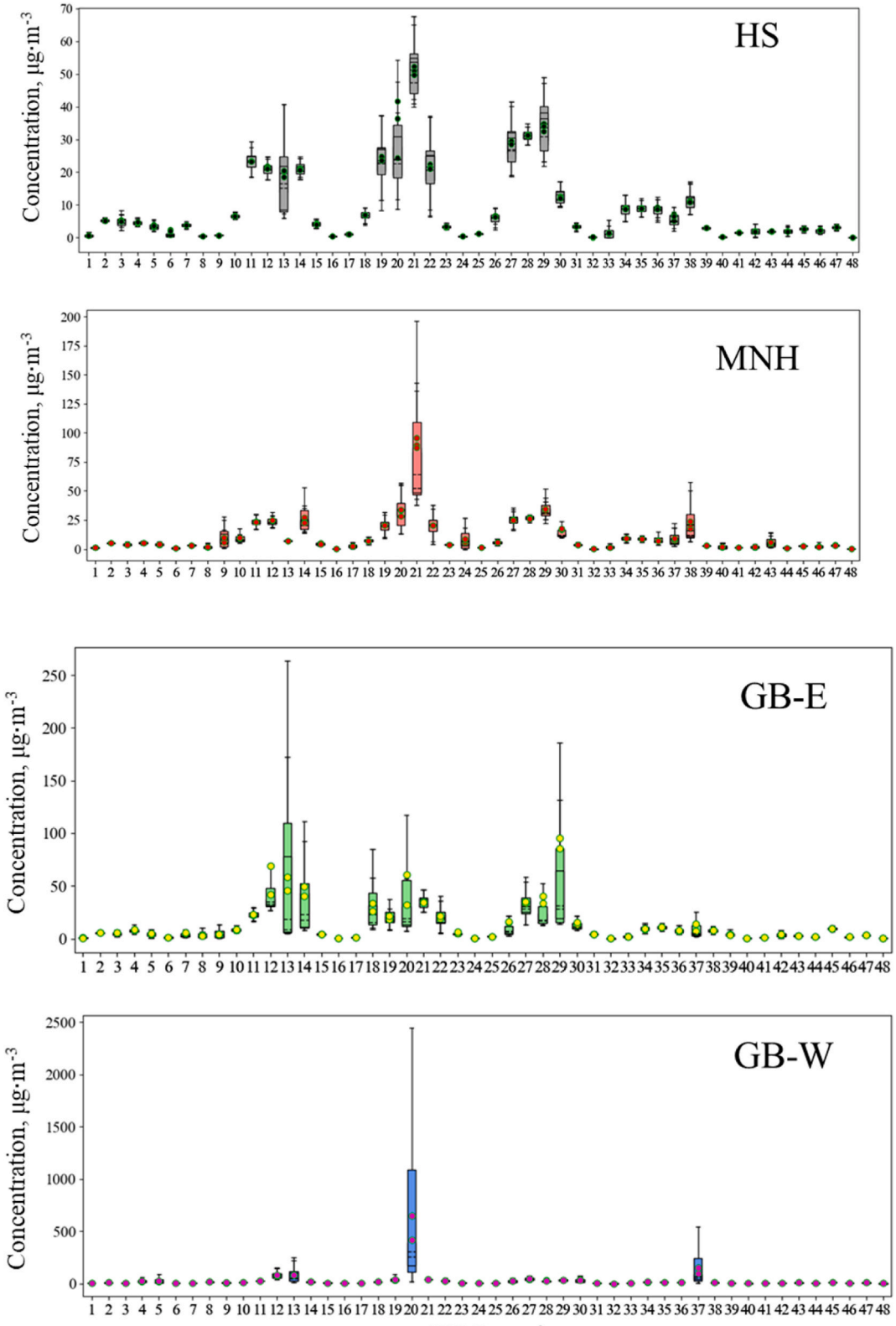


Fig. 4. VOC concentration distribution in key factories.

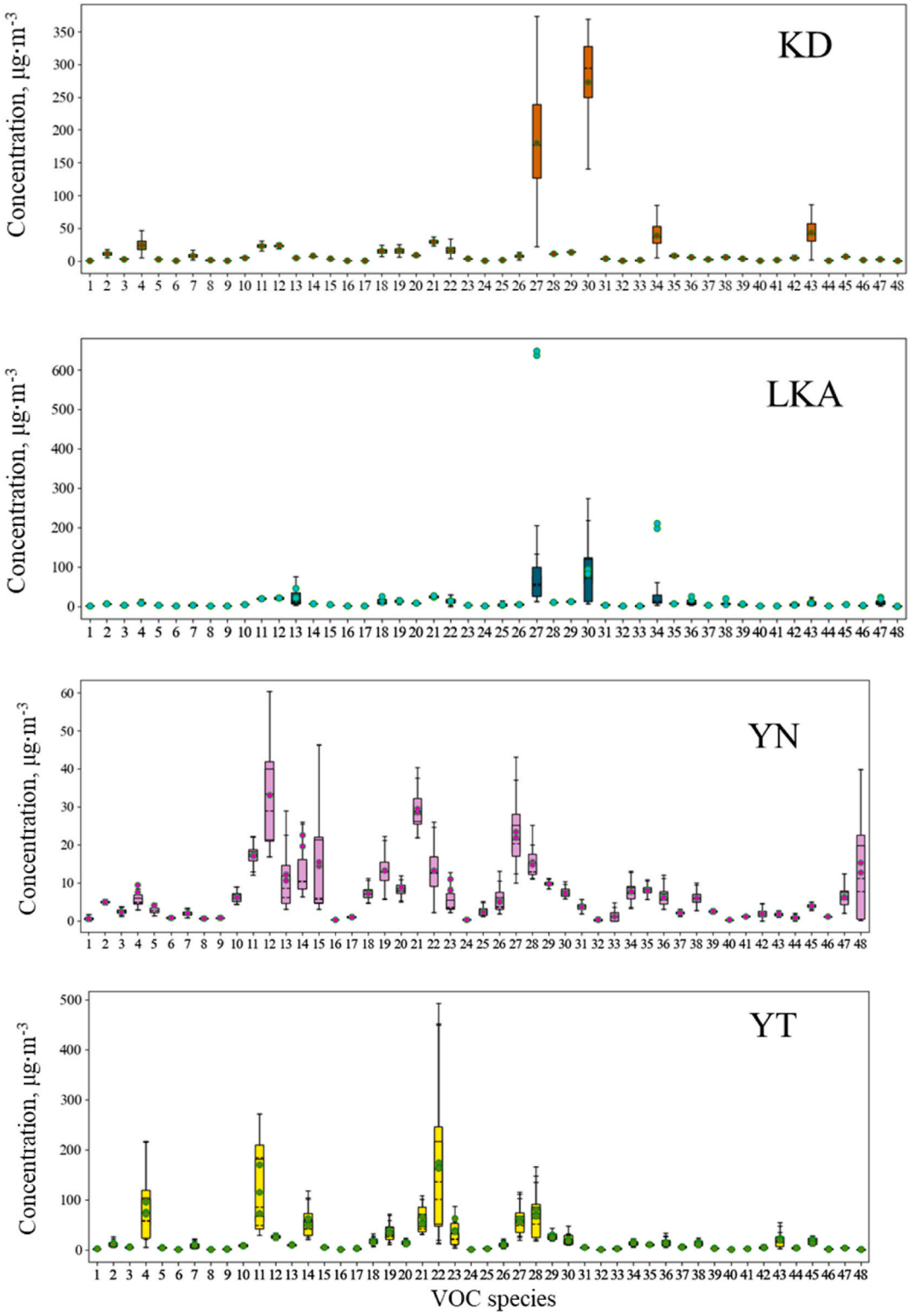


Fig. 4. (*continued*).

monitoring stations, with little difference between the data. Currently, few PMF calculations use MD. The $MD_{KF/SR}$ dataset was large, with high time accuracy and a wide spatial range, which could be further subdivided. Intermittent production was the main production method employed in the studied park, wherein feeding, reaction, separation, and other operations were performed at various intervals

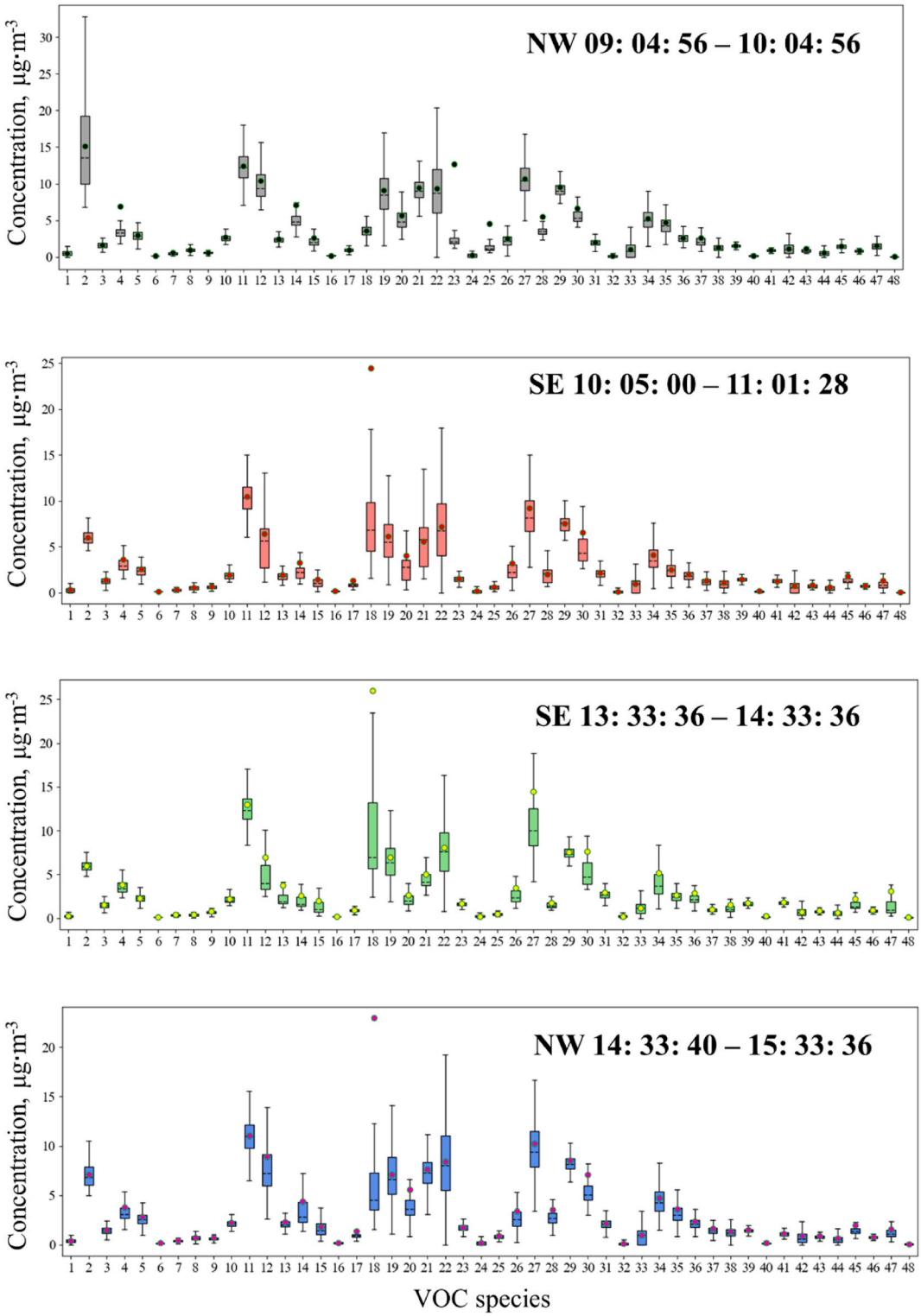


Fig. 5. VOC concentration distribution around the chemical park.

throughout the day. Moreover, the scale of the park was relatively small, requiring improved refinement of time and space for the source apportionment results. Therefore, the MD_{SR} were divided into four time periods for VOC source apportionment, maintaining sufficient data in each time period to ensure the accuracy of PMF calculations.

The input data files containing VOC species weight percentages, uncertainty, and optimal number of source factors were determined according to the EPA PMF5.0 User Guide. The factor profiles (FPs) calculated from the four time periods were combined through hierarchical clustering to determine the final combined factor profiles (CFPs). See **S1.1** and Fig. S1-S9 for details of the above. Referring to the VOC emission characteristics of the factories displayed by their CAMs, combined with analysis of the production processes and tracer species, the FPs were mapped to the actual emission sources and the VOC tracing results were obtained.

3. Results and discussion

3.1. VOC concentration distribution

During the mobile monitoring of key factories in the chemical park, the total VOC concentrations ranged from 223.64 to 1704.63 $\mu g \cdot m^{-3}$, with an average concentration of 847.75 $\mu g \cdot m^{-3}$.

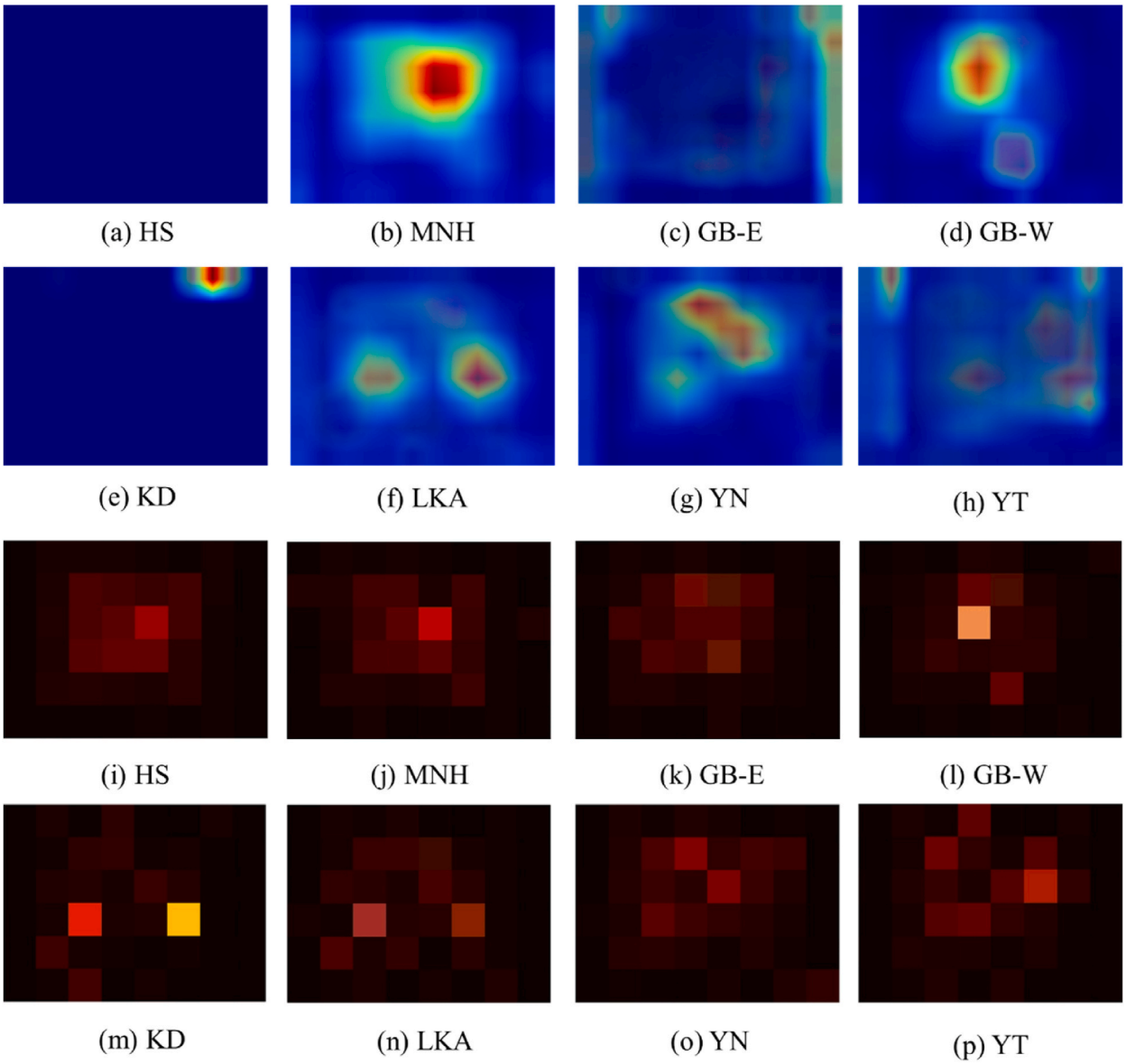


Fig. 6. Average CAM and WPI of each factory.

The distribution of VOC concentrations among the 8 factories in the park was depicted in Fig. 4. Most species exhibited significant concentration fluctuations. Some outliers suggested intermittent high VOC emissions due to sporadic production activities in the chemical park. In HS, the VOC concentrations were generally low. Notable species with relatively higher concentrations included ethyl acetate, acetone, and chloroform. Ethyl acetate exhibited the highest concentration, ranging from 39.98 to 68.12 $\mu g \cdot m^{-3}$, with an average of 49.70 $\mu g \cdot m^{-3}$. In MNH, ethyl acetate concentrations significantly surpassed other species, ranging from 37.98 to 196.65 $\mu g \cdot m^{-3}$, with an average of 95.50 $\mu g \cdot m^{-3}$. In GB-E, species with relatively higher concentrations included butane, acetone, toluene, and chloroform. Their concentrations ranged from 4.62 to 264.40 $\mu g \cdot m^{-3}$, with average values of 45.57 $\mu g \cdot m^{-3}$, 85.58 $\mu g \cdot m^{-3}$, 49.08 $\mu g \cdot m^{-3}$, and 61.13 $\mu g \cdot m^{-3}$, respectively. In GB-W, chloroform concentrations significantly exceeded other species. Their concentrations ranged from 16.25 to 2452.65 $\mu g \cdot m^{-3}$, with an average of 648.85 $\mu g \cdot m^{-3}$. Chloromethane followed, with concentrations ranging from 2.78 to 541.24 $\mu g \cdot m^{-3}$, averaging 150.73 $\mu g \cdot m^{-3}$. In KD, concentrations of pentene and acetaldehyde significantly surpassed other species. Their concentrations ranged from 22.49 to 373.18 $\mu g \cdot m^{-3}$ for pentene and 10.93 to 369.74 $\mu g \cdot m^{-3}$ for acetaldehyde. Averages were 180.59 $\mu g \cdot m^{-3}$ and 272.94 $\mu g \cdot m^{-3}$, respectively. In LKA, similarly, pentene and acetaldehyde dominated. Their concentrations ranged from 12.88 to 209.15 $\mu g \cdot m^{-3}$ for pentene and 6.08 to 272.74 $\mu g \cdot m^{-3}$ for acetaldehyde. Averages are 66.95 $\mu g \cdot m^{-3}$ and 82.11 $\mu g \cdot m^{-3}$, respectively. In YN, the VOC concentrations were generally low across all species. Xylene exhibited relatively higher concentrations, ranging from 16.82 to 60.75 $\mu g \cdot m^{-3}$, with an average of 33.11 $\mu g \cdot m^{-3}$. In YT, methyl cyclopentane, vinyl acetate, and dichloroethane dominated. Their concentrations ranged from 12.12 to 493.75 $\mu g \cdot m^{-3}$, 29.44 to 271.85 $\mu g \cdot m^{-3}$, and 5.28 to 220.02 $\mu g \cdot m^{-3}$, respectively. Averages are 163.13 $\mu g \cdot m^{-3}$, 72.14 $\mu g \cdot m^{-3}$, and 71.41 $\mu g \cdot m^{-3}$, respectively.

Overall, HS and YN consistently exhibited lower VOC concentrations, all below 100 $\mu g \cdot m^{-3}$. Other factories showed moderate average concentrations but occasional spikes. Emission characteristics were pronounced in MNH, GB-W, KD, LKA, and YT. KD and LKA shared similar characteristic species, making them indistinguishable based solely on concentration data. HS, GB-E, and YN lack distinct characteristic species, hindering source differentiation based on concentration alone.

The monitoring of the surrounding areas near the industrial park covered four time periods: northwest in the morning, southeast in the morning, southeast in the afternoon, and northwest in the afternoon, as shown in Fig. 5. The total concentration of VOCs ranged from 72.67 to 730.79 $\mu g \cdot m^{-3}$, with an average value of 153.13 $\mu g \cdot m^{-3}$. Notably, these concentrations were significantly lower than those observed in the factories. Across the four time periods, the distribution of VOC species exhibited overall similarities. Ethyl acetate, xylene, butane and methyl cyclopentane showed relatively higher concentrations. However, it was important to note that the concentration fluctuations for each species were complex and varied. Despite the general consistency in species distribution, the dynamic nature of VOC levels warranted further investigation.

3.2. VOC emission characteristics

As shown in Fig. S10, the constructed CNN showed good performance with a classification accuracy of 99.99%, which was the key premise to identifying VOC emission characteristics. The average CAM and WPI of each factory were shown in Fig. 6. In Fig. 6 (a)–(h), a highlighted grid indicated that the corresponding VOC species had a high impact weight on the CNN classifier's correct classification of the VOC WPI, i.e., the VOC emission characteristic species of the factory. In Fig. 6 (i)–(p), a grid with a high thermal value indicated that the corresponding VOC species had a high weight percentage. The characteristic species were obtained, as shown in Table 2.

The WPI of HS had ethyl acetate with the highest average weight percentage. The CAM of MNH had the most obvious activation of ethyl acetate, which was consistent with its WPI, and the grids of methyl cyclopentane and butane were partially activated. In the CAM of GB-E, there were few obvious areas of activation, with the highlighted regions mainly displayed on the right side and relatively high activation levels for hexanal and carbon disulfide. Similarly, few species had significantly high weight percentages in the WPI of GB-E, in which xylene and acetone were ranked in the top two. In the CAM of GB-W, the highlighted regions were displayed in the areas of chloroform and xylene, and the lower active area in the map was represented by chloromethane. Meanwhile, the top three species in weight percentages in the WPI of GB-W were also chloroform, xylene, and chloromethane. In the CAM of KD, the area represented by dichloropropene was obviously activated; however, acetaldehyde and pentene had the highest weight percentages in its WPI. In the CAM of LKA, the highlighted regions were clearly divided into two parts on the left and right, and the species included acetaldehyde, methyl chloroform, methyl cyclopentane, and pentene. Meanwhile, in the WPI of LKA, acetaldehyde and pentene had the highest

Table 2
VOC emission characteristic species in CAMs and WPIs.

Factory	Characteristic VOC species
HS	Ethyl acetate[b]
MNH	Ethyl acetate[a,b], Methyl cyclopentane[a], Butane[a]
GB-E	Hexanal[a], Carbon disulfide[a], Xylene[b], Acetone[b]
GB-W	Chloroform[a,b], Xylene[a,b], Chloromethane[a,b]
KD	Dichloropropene[a], Acetaldehyde[b], Pentene[b]
LKA	Acetaldehyde[a,b], Methyl chloroform[a], Methyl cyclopentane[a], Pentene[a,b]
YN	Methyl cyclopentane[a], Butane[a], Ethyl acetate[a,b], Xylene[a,b], Pentene[b]
YT	Trichloro ethylene[a], Methyl mercaptan[a], Dichloropropene[a], Carbon disulfide[a], Acrolein-tetrachloroethane[a], Methyl cyclopentane[b]

[a] Identified from CAMs.
[b] Identified from WPIs.

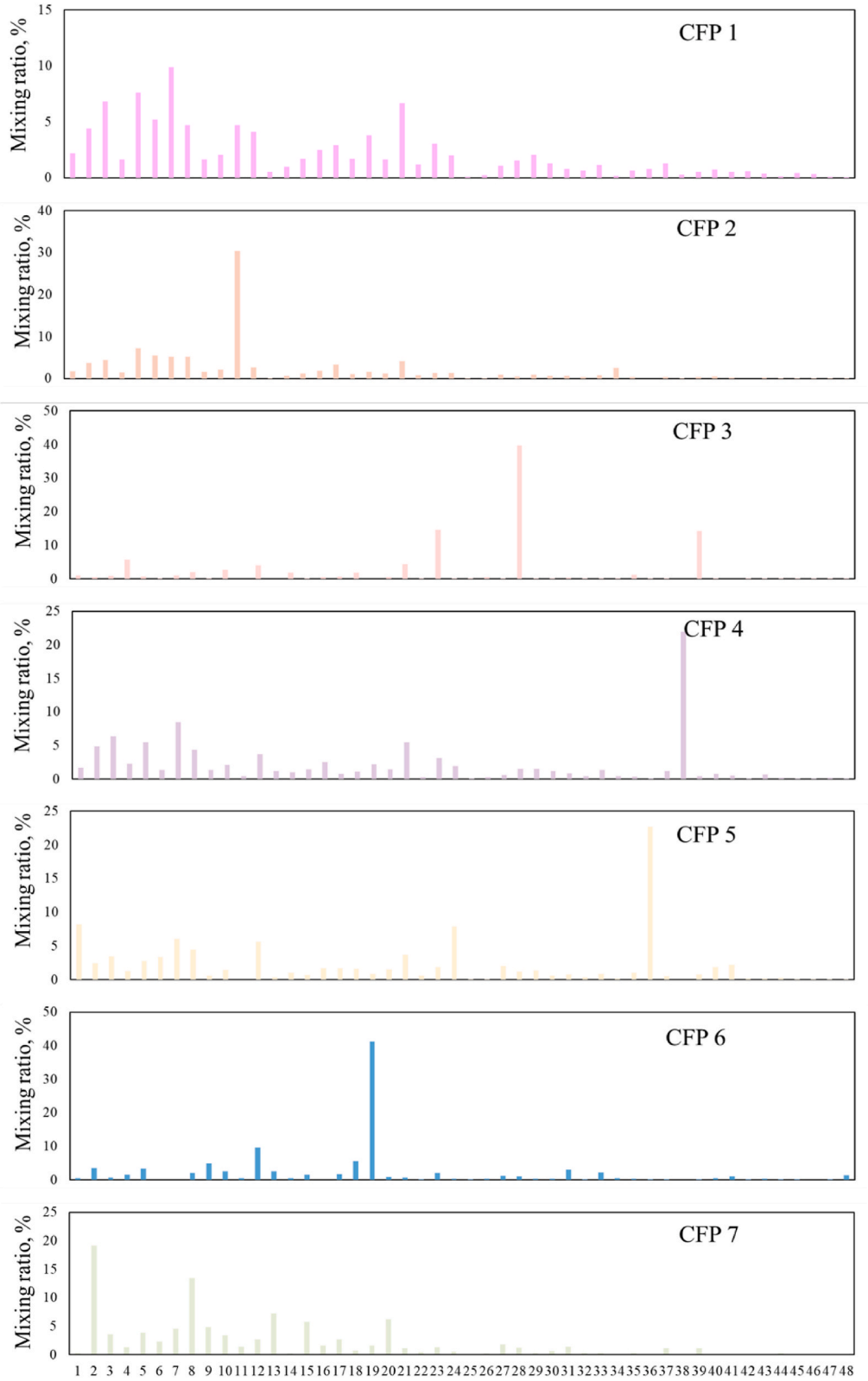


Fig. 7. Combined factor profiles calculated by PMF.

weight percentages, which was the same as in the WPI of KD. In the CAM of YN, the highlighted regions were also located in the middle of the map, representing species such as methyl cyclopentane, butane, ethyl acetate, and xylene. Meanwhile, in the WPI of YN, the top three species in weight percentages were xylene, ethyl acetate, and pentene. In the CAM of YT, three activation areas were located at the junctions of trichloro ethylene and methyl mercaptan, dichloropropene and carbon disulfide, and acrolein and tetrachloroethane, respectively. In the WPI of YT, methyl cyclopentane had the highest weight percentage.

In summary, the characteristic species generally corresponded to species with high weight percentages, as shown in Fig. 6 (b), (d), (f), and **(g).** However, species with high weight percentages from different emission sources were similar in some cases, complicating classification of the pollution source, such as in Fig. 6 (e)–(f). Moreover, no obvious species had high weight percentages from emission sources in some cases, such as in Fig. 6 (c), and **(h)**, and the characteristic species were dispersed. The above phenomena indicated that, for a single source, species with high weight percentages might not be well differentiated from other sources. When multiple sources had similar emissions of this species, differences in other species with lower weight percentages were more distinguishable.

3.3. Source identification of CFPs

The numbers of factors for each time period were set to 7, 7, 7, and 8, respectively. The Q(true)/Q(exp) values were 4.38, 5.07, 4.25, and 2.72, respectively. In **S1.1**, the PMF results for each time period were described in detail. As shown in Fig. S11, the residuals of all species were basically within ±5%, indicating a good fit between the PMF analysis results and monitoring data.

There were 29 FPs calculated in the four time periods. Hierarchical cluster analysis was performed using 29 factors, and the results were shown in Fig. S12. According to the tree diagram in combination, the 29 factors were divided into seven classes. The average values of each class of factor were calculated to form seven CFPs, as shown in Fig. 7. In CFP 1, the species composition was dispersed, with the total proportion of vinyl acetate, methyl mercaptan, pentene, and acetone reaching 31.06%. In CFP 2, methyl cyclopentane was the obvious characteristic species, accounting for 30.40%. In CFP 3, the proportions of chloroethylene, dichloroethane, and styrene were 39.67%, 14.52%, and 14.22%, respectively. In CFPs 4, 5, and 6, the proportions of trichlorobenzene, tetrachloroethylene, and 1,3-butadiene accounted for 21.97%, 22.73%, and 41.26%, respectively. In CFP 7, the proportions of ethyl acetate and xylene were 19.16% and 13.39%, respectively.

In summary, 13 VOC tracer species were identified in the CFPs analyzed by PMF. Based on Fig. 6, YT had the highest average value (13.7%) for the weight percentage of methyl cyclopentane, which was obviously higher than that of other factories, as well as the most extreme values (>40%). Therefore, CFP 2 characterized by methyl cyclopentane might have represented the VOC emissions from YT. Similarly, CFP 1 characterized by pentene might have represented the VOC emissions from LKA. Additionally, CFP 7 characterized by ethyl acetate might have represented the VOC emissions from MNH. Interestingly, the weight percentage of dichloroethane from YT was also higher than that from other factories, indicating that CFP 3 could also represent the VOC emissions from YT to a certain extent.

Subsequently, we searched for the actual emission sources corresponding to these tracer species in the relevant literature. 1,3-butadiene was found to be a tracer species from power plants [35], indicating that CFP 6 might have characterized VOC emissions from a power plant. This finding was consistent with the actual distribution of a power plant, named CH, in the park. Tetrachloroethylene, which is mainly used as an organic solvent, dry cleaning agent, solvent for adhesives, degreasing solvent for metals, drying agent, and paint remover, was considered as a tracer species for sources of solvent use [36]. Therefore, CFP 5 characterized by tetrachloroethylene might represent mixed solvent volatilization from various factories in the park. However, dichloroethane is widely recognized as a tracer species for the petrochemical industry [37], which is different from the fine chemical industrial park in this study. Styrene and halogenated hydrocarbons appearing in CFPs 3 and 4 were typical tracer species from industrial processes [36], but more detailed source identification could not be found in the literature. Nevertheless, we investigated the environmental impact assessments of factories in the park and found that only JH emitted styrene, as it mainly produces fine chemical products such as fertilizer synergists and isopropanol. Moreover, only BA, which specializes in disperse dyes, produced trichlorobenzene emissions.

The above analysis revealed that direct observation of the MD_{KF} was not conductive to achieving the refined source identification of

Table 3
Identified VOC emission sources corresponding to FPs.

CFP	Tracing VOC species	Source identification			Finally identified emission sources
		Monitoring data	Literature	CAMs	
1	Vinyl acetate, methyl mercaptan, pentene, acetone	LKA	/	YT− LKA	LKA
2	Methyl cyclopentane	YT+	/	MNH− LKA− YN−	YT
3	Chloroethylene, dichloroethane, styrene	YT−	Petrochemicals− JH+	/	JH
4	Trichlorobenzene	/	Industrial processes− BA+	/	BA
5	Tetrachloroethylene	/	SV+	/	SV
6	1,3-butadiene	/	CH+	/	CH
7	Ethyl acetate, xylene	MNH+	/	YN+	YN

+ represents higher reliability of the result, − represents lower reliability of the result.

the CFPs. Essential VOC emission characteristics are hidden in monitoring data due to the impact of diffusion and chemical reactions, and need to be mined. The current literature also lacks detailed identification of characteristic species of industrial VOC emission sources. Therefore, the emission characteristics identified by the CAMs were conducive to source identification of the CFPs.

In CFP 1, methyl mercaptan was slightly activated in the CAM of YT, while pentene was obviously activated in the CAM of LKA, indicating that this species had a large weight in the CNN classifier's correct classification of VOC samples from LKA. In CFP 2, the grid representing methyl cyclopentane was partially activated in the CAMs of MNH, LKA, and YN, indicating that this species only characterized part of the samples from the three factories. In CFPs 3–6, no species were activated in the CAMs, possibly indicating that the corresponding emissions were not generated by the eight factories. In CFP 7, ethyl acetate and xylene were obviously activated in the CAM of YN, indicating that these species significantly affected the correct classification of VOC samples from YN.

Based on the source identification results of the above three steps, the actual emission sources corresponding to the CFPs were finally determined, as shown in Table 3. For CFP 1, LKA was identified by pentene via the WPI. Therefore, CFP 1 was identified as LKA. CFPs 2, 3, and 4 were identified as YT, JH, and BA. CFP 5 was identified as solvent use (SV) in the park. CFPs 6 and 7 were identified as CH and YN.

3.4. Results of VOC source apportionment

The VOC source apportionment results are shown in Fig. 8, including the locations of main sources and receptors in the chemical industrial park and its surrounding areas (Fig. 8 (a)). Many atmospheric-sensitive spots exist around the park, including >10 villages in

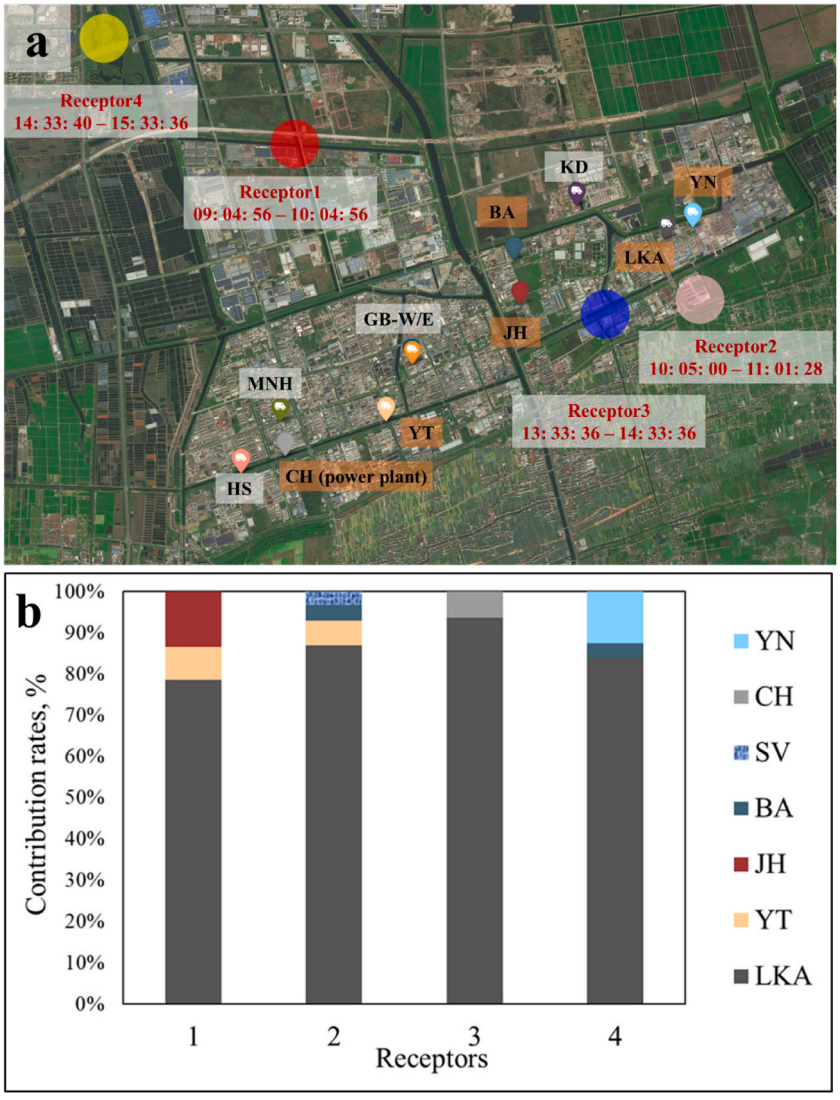


Fig. 8. VOC source apportionment results in each period. (a) Locations of sources and receptors; (b) Contribution ratios.

the southeast with a total of >30,000 residents and several schools in the northwest. The circles represent the average coordinates of the MD_{SR} for the four time periods, marked in red, pink, blue, and yellow in chronological order. The eight key factories in the MD_{KF} and three other identified sources are also marked with different colors. The source contribution ratios calculated by PMF are shown in Fig. 8 (b). In each time period, the contribution of LKA's emissions accounted for the majority, with 79%, 87%, 94%, and 84%, respectively, which first increased and then decreased from morning to afternoon. The wind speed and direction in the chemical park were recorded during the monitoring period, as shown in Fig. S13. The average wind speed was 2.7 $m \cdot s^{-1}$, with a maximum of 9.7 $m \cdot s^{-1}$. The prevailing wind direction was 249°, and moments with wind speeds below 5 $m \cdot s^{-1}$ accounted for 89.1% of the time. Overall, wind speeds remained relatively low, corresponding to Beaufort scale levels 0 to 3, indicating minimal impact on pollutant dispersion. LKA was significantly closer to receptors 2 and 3 than to receptors 1 and 4, thus exerting a greater impact and resulting in a relatively higher contribution ratio. However, although LKA was closer to receptor 1 than to receptor 4, its contribution ratio was lower, indicating that LKA's VOC emissions in the afternoon were higher than those in the morning, which to some extent reflected the intermittent production methods employed in the park.

In the school area northwest of the park, except for LKA, 8% and 13% of morning VOC emissions originated from YT and JH, respectively. In the afternoon, 4% and 12% of VOCs in the environment originated from BA and YN. In the residential areas southeast of the park, except for LKA, 6% and 4% of VOCs in the morning environment originated from YT and BA, while a clear source was not identified for another 3%, which was speculated to be generated by solvent use somewhere in the park. In the afternoon, the contribution of VOCs in the environment from LKA reached a maximum at 94%, indicating that LKA-derived emissions reached their maximum during the day. The remaining 6% originated from CH, a power plant.

4. Conclusions

A novel approach for identifying VOC emission characteristics based on mobile MD and deep learning was proposed that meets the requirements of rapid, refined, and accurate VOC source apportionment in chemical industrial parks. In most cases, the characteristic species generally corresponded to species with high weight percentages. However, species with high weight percentages from different emission sources were similar in some cases, so the characteristic species beneficial to source classification were mined. Based on the identified VOC emission characteristics, the main pollution sources and their contributions to surrounding schools and residential areas were determined. The major pollution sources in the park were accurately positioned and temporal variations were discovered, reflecting intermittent production methods employed in the park. According to our research, the management of VOC emissions from LKA should be strengthen, especially the pentene and acetaldehyde emissions, which require in-depth analysis of the production process and identification of key emission units.

However, this study still has certain limitations. In terms of data collection, limited by actual park management requirements, it is difficult for vehicles to enter the factory area for long-term monitoring. Moreover, due to the technological limitations of mobile monitoring platform sensors, the detected VOC species are not yet abundant enough. Long-term monitoring of emission sources and larger data set will help to improve model performance and identify more refined emission characteristics. The eight factories represented the dye and pharmaceutical industries in the park in terms of production types, and covered the western, central, and eastern regions of the park, supporting that the data selection was representative. Although the numbers of samples for each factory were not strictly equivalent due to varying sizes of the factories and differences in mobile monitoring times, the difference in sample numbers among different classes was not significant and a validation accuracy of up to 99% indicated that classifier training was not affected by class imbalance.

As for model selection, CNN is currently a widely used deep learning model, which has effective performance in classifying images (two-dimensional data) based on their pixel features, which has been illustrated in many researches [38]. The CNN is applied since it is essentially similar to a typical classification question of accurately classifying the pollution sources based on VOC emission characteristics. The technical feasibility of visualizing image features using CAM has also been validated in literature. The Keras used to build CNN models is currently the most popular open-source deep learning framework. Researchers can easily make appropriate adjustments to the network architecture based on actual situations. Therefore, the approach for identifying VOC emission characteristics of this study is expected to be promoted.

For the emission characteristics identification results, we divided the CAMs into 6 × 8 grids, and identified the species represented by the clearly activated grid as characteristics species. However, in some factories, the activation area was not strictly concentrated within the designated grid, but rather on the boundary line. It meant that what affected CNN's classification accuracy might not be individual species, but rather the relative relationships between species, such as the ratio between two species. Therefore, when researchers encounter such situations, they can try to identify emission characteristics based on relative relationships between species by combining their own research reality.

In terms of interpreting the source apportionment results, we comprehensively considered CAM results, mobile monitoring data analysis, and existing literature to identify the emission sources comprehensively. However, the emission sources identified by these three methods still had biases in some cases. As for analyzing source contribution ratios, PMF, as a representative of receptor models, it is difficult to consider the influence of meteorological factors such as wind speed and direction in PMF. In this study, due to the small wind speed and stable wind direction, the influence of meteorological conditions was not considered, resulting in a certain deviation between the traceability results and the actual situation. However, in situations where conditions are sufficient, such as source profiles and emission inventories with high quality in parks, researchers can try to combine them with CMB models or air dispersion models. By combining the methods of this study with various traditional models and conducting multiple comparisons, the interpretation of the

results can be more realistic and reasonable.

CRediT authorship contribution statement

Deji Jing: Writing – review & editing, Writing – original draft, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Kexuan Yang:** Writing – review & editing, Investigation, Formal analysis. **Zhanhong Shi:** Investigation, Formal analysis. **Xingnong Cai:** Investigation, Formal analysis. **Sujing Li:** Investigation, Formal analysis. **Wei Li:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Qiaoli Wang:** Writing – review & editing, Methodology, Investigation, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was funded by National Key Research and Development Program of China (2022YFC3702002).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e29077.

References

[1] W. Xu, Y. Kuang, Y. Bian, L. Liu, F. Li, Y. Wang, B. Xue, B. Luo, S. Huang, B. Yuan, P. Zhao, M. Shao, Current challenges in visibility improvement in southern China, Environ. Sci. Technol. Lett. (2020) 395–401, https://doi.org/10.1021/acs.estlett.0c00274.
[2] K. Li, D.J. Jacob, H. Liao, L. Shen, Q. Zhang, K.H. Bates, Anthropogenic drivers of 2013–2017 trends in summer surface ozone in China. Proceedings of the Na-Tional Academy of Sciences, 2019, pp. 422–427, https://doi.org/10.1073/pnas.1812168116.
[3] C. Zheng, J. Shen, Y. Zhang, X. Zhu, X. Wu, L. Chen, X. Gao, Atmospheric emission characteristics and control policies of anthropogenic VOCs from industrial sources in Yangtze river Delta region, China, Aerosol Air Qual. Res. (2017) 2263–2275, https://doi.org/10.4209/aaqr.2016.06.0234.
[4] W. Sun, M. Shao, C. Granier, Y. Liu, C.S. Ye, J.Y. Zheng, Long-term trends of anthropogenic SO_2, NO_X, CO, and NMVOCs emissions in China, Earth's Future (2018) 1112–1133, https://doi.org/10.1029/2018EF000822.
[5] C. Zheng, J. Shen, Y. Zhang, W. Huang, X. Zhu, X. Wu, L. Chen, X. Gao, K. Cen, Quantitative assessment of industrial VOC emissions in China: historical trend, spatial distribution, uncertainties, and projection, Atmos. Environ. (2017) 116–125, https://doi.org/10.1016/j.atmosenv.2016.11.023.
[6] X. Liang, X. Chen, J. Zhang, T. Shi, X. Sun, L. Fan, L. Wang, D. Ye, Reactivity-based industrial volatile organic compounds emission inventory and its implications for ozone control strategies in China, Atmos. Environ. (2017) 115–126, https://doi.org/10.1016/j.atmosenv.2017.04.036.
[7] M. He, H. Xiao, X. Chen, X. Liang, L. Fan, D. Ye, Emission characteristics of volatile organic compounds in chemical industry park based on emission links, China Environ. Sci. 37 (2021) 38–48. China Petroleum and Chemical Federation. 2021from, http://www.cpcif.org.cn/.
[9] Y. Li, S. Yin, S. Yu, M. Yuan, Z. Dong, D. Zhang, L. Yang, R. Zhang, Characteristics, source apportionment and health risks of ambient VOCs during high ozone period at an urban site in central plain, China, Chemosphere (2020) 126283, https://doi.org/10.1016/j.chemosphere.2020.126283.
[10] J. Duan, J. Tan, L. Yang, S. Wu, J. Hao, Concentration, sources and ozone formation potential of volatile organic compounds (VOCs) during ozone episode in Beijing, Atmos. Res. (2008) 25–35, https://doi.org/10.1016/j.atmosres.2007.09.004.
[11] Y. Gu, B. Liu, Y. Li, Y. Zhang, X. Bi, J. Wu, C. Song, Q. Dai, Y. Han, G. Ren, Y. Feng, Multi-scale volatile organic compound (VOC) source apportionment in Tianjin, China, using a receptor model coupled with 1-hr resolution data, Environ. Pollut. (2020) 115023, https://doi.org/10.1016/j.envpol.2020.115023.
[12] B. Barletta, S. Meinardi, F.S. Rowland, C. Chan, X. Wang, S. Zou, L.Y. Chan, D.R. Blake, Volatile organic compounds in 43 Chinese cities, Atmos. Environ. (2005) 5979–5990, https://doi.org/10.1016/j.atmosenv.2005.06.029.
[13] S.L. Miller, M.J. Anderson, E.P. Daly, J.B. Milford, Source apportionment of exposures to volatile organic compounds. I. Evaluation of receptor models using simulated exposure data, Atmos. Environ. (2002) 3629–3641, https://doi.org/10.1016/S1352-2310(02)00279-0.
[14] M. Song, X. Liu, Y. Zhang, M. Shao, K. Lu, Q. Tan, M. Feng, Y. Qu, Sources and abatement mechanisms of VOCs in southern China, Atmos. Environ. (2019) 28–40, https://doi.org/10.1016/j.atmosenv.2018.12.019.
[15] Y. Huang, G. Xiu, Y. Lu, S. Gao, L. Li, L. Chen, Q. Huang, Y. Yang, X. Che, X. Chen, Q. Fu, Application of an emission profile-based method to trace the sources of volatile organic com-pounds in a chemical industrial park, Sci. Total Environ. 768 (2021) 144694, https://doi.org/10.1016/j.scitotenv.
[16] D. Wang, Fine source tracing of typical VOCs pollution episodes around industrial park, China Environ. Sci. 42 (2022) 585–592, https://doi.org/10.19674/j.cnki.issn1000-6923.20210930.001.
[17] D. Feng, C. Wang, B. Chen, Q. Wang, S. Zhang, J. Chen, W. Li, Characteristics and source analysis of VOCs in a fine chemical industrial park, Chin. J. Chem. Eng. 35 (2021) 935–942.
[18] B. Yuan, M. Shao, S. Lu, B. Wang, Source profiles of volatile organic compounds associated with solvent use in Beijing, China, Atmos. Environ. (2010) 1919–1926, https://doi.org/10.1016/j.atmosenv.2010.02.014.
[19] L. Shen, P. Xiang, S. Liang, W. Chen, M. Wang, S. Lu, Z. Wang, Sources profiles of volatile organic compounds (VOCs) measured in a typical industrial process in Wuhan, Central China, Atmosphere-Basel (2018) 297, https://doi.org/10.3390/atmos9080297.
[20] N. Cheng, D. Jing, C. Zhang, Z. Chen, W. Li, S. Li, Q. Wang, Process-based VOCs source profiles and contributions to ozone formation and carcinogenic risk in a typical chemical synthesis pharmaceutical industry in China, Sci. Total Environ. (2021) 141899, https://doi.org/10.1016/j.scitotenv.2020.141899.
[21] H. Wang, Y. Gao, S. Jing, S. Lou, X. Hu, J. An, Y. Wu, W. Gao, L. Zhu, C. Huang, Characterization of volatile organic compounds (VOCs) using mobile monitoring around the industrial parks in the Yangzte river Delta region of China, Environmental Science 42 (2021) 1298–1305.
[22] D. Ma, Z. Zhang, Contaminant dispersion prediction and source estimation with integrated Gaussian-machine learning network model for point source emission in atmosphere, J. Hazard Mater. (2016) 237–245, https://doi.org/10.1016/j.jhazmat.2016.03.022.
[23] P.E. Bieringer, G.S. Young, L.M. Rodriguez, A.J. Annunzio, F. Vanden-berghe, S.E. Haupt, Paradigms and commonalities in atmospheric source term estimation methods, Atmos. Environ. (2017) 102–112, https://doi.org/10.1016/j.atmosenv.2017.02.011.

[24] M. Hutchinson, H. Oh, W. Chen, A review of source term estimation methods for atmospheric dispersion events using static or mobile sensors, Inform. Fusion (2017) 130–148, https://doi.org/10.1016/j.inffus.2016.11.010.
[25] X. Zhao, K. Cheng, W. Zhou, Y. Cao, S. Yang, J. Chen, Source term estimation with deficient sensors: a temporal augment approach, Process Saf. Environ. (2022) 131–139, https://doi.org/10.1016/j.psep.2021.10.042.
[26] B. Wang, B. Chen, J. Zhao, The real-time estimation of hazardous gas dispersion by the integration of gas detectors, neural network and gas dispersion models, J. Hazard Mater. (2015) 433–442, https://doi.org/10.1016/j.jhazmat.2015.07.028.
[27] D. Ma, W. Tan, Z. Zhang, J. Hu, Parameter identification for continuous point emission source based on Tikhonov regularization method coupled with particle swarm optimization algorithm, J. Hazard Mater. (2017) 239–250, https://doi.org/10.1016/j.jhazmat.2016.11.071.
[28] D. Jing, N. Cheng, C. Zhang, Z. Chen, X. Cai, S. Li, J. Zhao, Q. Wang, W. Li, A novel approach for VOC source apportionment combining characteristic factor and pattern recognition technology in a Chinese industrial area, J. Environ. Sci.-China (2022) 25–37, https://doi.org/10.1016/j.jes.2021.08.056.
[29] Y. Liu, W. Yao, F. Qin, L. Zhou, Y. Zheng, Spectral classification of large-scale blended (Micro)Plastics using FT-IR raw spectra and image-based machine learning, Environ. Sci. Technol. (2023) 6656–6663, https://doi.org/10.1021/acs.est.2c08952.
[30] W. Cheng, C.A. Ng, Using machine learning to classify bioactivity for 3486 per- and polyfluoroalkyl substances (PFASs) from the OECD list, Environ. Sci. Technol. (2019) 13970–13980, https://doi.org/10.1021/acs.est.9b04833.
[31] Q. Di, I. Kloog, P. Koutrakis, A. Lyapustin, Y. Wang, J. Schwartz, Assessing $PM_{2.5}$ exposures with high spatiotemporal resolution across the continental United States, Environ. Sci. Technol. (2016) 4712–4721, https://doi.org/10.1021/acs.est.5b06121.
[32] X. Hu, J.H. Belle, X. Meng, A. Wildani, L.A. Waller, M.J. Strickland, Y. Liu, Estimating $PM_{2.5}$ concentrations in the conterminous United States using the random forest approach, Environ. Sci. Technol. (2017) 6936–6944, https://doi.org/10.1021/acs.est.7b01210.
[33] J. Xu, M. Zhang, A. Ganji, K. Mallinen, A. Wang, M. Lloyd, A. Venuta, L. Simon, J. Kang, J. Gong, Y. Zamel, S. Weichenthal, M. Hatzopoulou, Prediction of short-term ultrafine particle exposures using real-time street-level images paired with air quality measurements, Environ. Sci. Technol. (2022) 12886–12897, https://doi.org/10.1021/acs.est.2c03193.
[34] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-CAM: visual explanations from deep networks via gradient-based localization, Int. J. Comput. Vis. (2020) 336–359, https://doi.org/10.1007/s11263-019-01228-7.
[35] Q. Wang, D. Sheng, C. Wu, D. Jing, N. Cheng, X. Cai, S. Li, J. Zhao, W. Li, J. Chen, A supplementary assessment system of AQI-V for comprehensive management and control of air quality in chemical industrial parks, J. Environ. Sci.-China (2023) 114–125, https://doi.org/10.1016/j.jes.2022.06.037.
[36] Q. He, Y. Yan, H. Li, Y. Zhang, L. Chen, Y. Wang, Characteristics and reactivity of volatile organic compounds from non-coal emission sources in China, Atmos. Environ. (2015) 153–162, https://doi.org/10.1016/j.atmosenv.2015.05.066.
[37] X. Han, S. Ma, W. Wan, C. Song, Y. Liu, Research progress of volatile organic compounds source profiles in petrochemical industry in China, Pet. Process. Petrochem. 53 (2022) 9–16.
[38] Y. Sun, B. Xue, M. Zhang, G.G. Yen, Evolving deep convolutional neural networks for image classification, IEEE T. Evolut. Comput. (2020) 394–407, https://doi.org/10.1109/TEVC.2019.2916183.