Electronic Supplementary Material (ESI) for Environmental Science: Processes & Impacts. This journal is © The Royal Society of Chemistry 2019

1	Electronic Supplementary Information
2	Profile and Source Apportionment of Volatile Organic Compounds from a Complex
3	Industrial Park
4	
5	Yuan Liu, ^a Qing Xie, ^a Xuehua Li, ^a Fulin Tian, ^b Xianliang Qiao, ^{a*} Jingwen Chen ^a and Wenwen
6	Ding ^a
7	^a Key Laboratory of Industrial Ecology and Environmental Engineering (MOE), School of Environmental Science
8	and Technology, Dalian University of Technology, Dalian 116024, China
9	^b Liaoning Academy of Analytical Sciences, Shenyang 110015, China
10	
11	
12	
13	
14	
15	
16	
17	
18	
19	
20	
21	

^{*}Corresponding author phone/fax: +86-411-84707844; e-mail: xlqiao@dlut.edu.cn.

Categories	Species	CAS	MW ^a	BP ^b	Ion ^c	RT ^d	k ^e	b ^f	r ^{2g}	MDL ^h	SE ⁱ
	Ethyl Acetate	141-78-6	88.1	77	43/61	10.61	0.42	0	0.998	11.6	86%
	Butyl acetate	123-86-4	116	125	43/56	15.72	0.69	0.01	0.999	8.11	90%
	1-Butanol	71-36-3	74.1	118	56/41	12.65	0.17	0	0.999	24.5	86%
	2-Ethylhexanol	104-76-7	130	185	57/41	20.39	0.83	0	0.998	17.2	86%
	Furfuryl alcohol	98-00-0	98.1	170	98/41	17.61	0.34	0.01	0.999	19.4	88%
OVOCs	2-Butanone	78-93-3	72.1	80	43/72	10.63	0.13	0	0.999	4.34	89%
	Pentaldehyde	110-62-3	86.1	103	44/41	13.21	0.21	0	0.999	11.3	88%
	Phenol	108-95-2	94.1	182	94/66/39	20.73	1.15	0	0.998	9.94	87%
	<i>m</i> -Cresol	108-39-4	108	203	108/107/79	22.23	0.82	0	0.987	4.25	79%
	<i>p</i> -Cresol	106-44-5	108	202	108/107/80	22.23	0.82	0	0.987	4.25	79%
	o-Cresol	95-48-7	108	191	108/107/81	21.79	0.93	0	0.994	8.47	84%
NVOCs	Triethylamine	121-44-8	101	90	86/58	12.1	0.04	0	0.999	6.79	88%
4.11	Methylcyclopentane	96-37-7	84.2	72	56/69	10.37	0.05	0	0.988	12.4	76%
Alkanes	3-Methylpentane	96-14-0	86.2	64	57/56	8.75	0.02	0	0.985	12.1	72%
	Benzene	71-43-2	78.1	80	78/77	11.92	1.28	0.18	0.993	4.39	88%
	Ethylbenzene	100-41-4	106	136	91/106	16.93	1.76	0.33	0.987	2.79	84%
	n-Propylbenzene	103-65-1	120	159	91/120	18.84	1.84	0	0.986	11.6	72%
	tert-Butylbenzene	98-06-6	134	169	119/91	19.59	1.91	0	0.994	2.75	81%
Aromatics	1,2,4-Trimethylbenzene	95-63-6	120	168	105/120	19.08	2.24	0	0.987	3.48	74%
	2,5-Dimethyl-p-xylen	95-93-2	134	196	119/134	21.88	2.28	0	0.989	1.48	76%
	4-Ethyl-m-xylene	874-41-9	134	186	119/134	21.3	1.91	0	0.994	4.90	84%
	Toluene	108-88-3	92.1	111	91/92	14.73	1.54	0.14	0.995	6.32	86%
	<i>m</i> -Xylene	108-38-3	106	165	91/106/105	17.09	1.54	0.37	0.988	9.39	82%

Table S1. The specific properties and quantitative parameters of VOCs in this study

<i>p</i> -Xylene	106-42-3	106	139	91/106	17.09	1.54	0.37	0.988	9.39	82%
o-Xylene	95-47-6	106	144	91/106	17.7	1.65	0.33	0.991	4.44	84%
1,3,5-Trimethylbenzene	108-67-8	120	165	105/120	19.66	2.12	0	0.988	5.38	79%
<i>m</i> -Ethyltoluene	620-14-4	120	158	105/120	18.94	1.97	0	0.994	8.03	83%
<i>p</i> -Ethyltoluene	622-96-8	120	162	105/120	18.94	2.68	0	0.994	8.03	83%
2-Ethyl-p-xylene	1758-88-9	134	187	119/134	21.78	1.91	0	0.989	6.43	77%
2,5-Dimethylfuran	625-86-5	96.1	93	96/95	12.88	0.26	0	0.997	9.46	87%

23 *a*, molecular weight; *b*, boiling point, °C; *c*, quantitative ion in GC/MS; *d*, retention time, min; $e \sim g$, parameters of calibration curve; *h*, Method of detection limits, ng m⁻³. *i*, sampling efficiency.

24 The procedures described by Zhu *et al.* were followed¹. Sample tubes were spiked at 300-900 ng per compound and flushed afterward with 12 L of high purity nitrogen to evaluate sampling

25 efficiency.

28

Table S2. The specific properties and quantitative parameters of internal standard

Internal standards	CAS	MW ^a	BP ^b	Ion ^c	\mathbf{RT}^{d}
Chloroform- <i>d</i> ₅	865-49-6	120.4	70	84/86	10.33
Toluene- <i>d</i> ₈	2037-26-5	100.2	110	98/100	13.95
Phenol- <i>d</i> ₅	4165-62-2	99.1	182	99/71/42	20.7
Nitrobenzene-d ₅	4165-60-0	128.1	88	82/54/128	22.3

29 *a*, molecular weight; *b*, boiling point, °C; *c*, quantitative ion in GC/MS; *d*, retention time, min;

All of this standard compounds were purchased from J&K scientific Ltd. They were dissolved in
methanol (Sigma-Aldrich Co., Ltd.) as stock solutions. The air pump (KC-60D) was purchased
from Laoshan Mountain Electronic Instrument Factory Co., Ltd. The Tenax tube (6 mm × 4 mm ×
200 mm) was purchased from Beijing Huayisanpu Yiqi Co., Ltd.

34

The main application of FA-NNC is to reduce the numbers of variables and detect the structure in the relationships between variables, that is, to classify variables. Compared with conventional factor analysis (FA), which always inevitably gets negative values causing physical meaning loss in practice, FA-NNC ensures factor axes to become less orthogonal and makes factor scores and loadings more interpretable due to fewer negative values.^{2,3} The fundamental equation underlying the principal component analysis based model is

$$43 \quad D = C \quad \times \quad R \tag{S1}$$

44
$$(m \times r)$$
 $(m \times n)$ $(n \times r)$

45 where the species concentration matrix (D) is fractioned into factor loading matrix (C) and factor 46 score matrix (R), representing source profiles and source contributions, respectively. *m*, *n*, and *r*

³⁵

³⁶ FA-NNC model

47 stand for the number of samples, sources, and species, respectively.

The procedures for FA-NNC described by Rachdawong *et al.*⁴ and Tian *et al.*³ were followed. In short, the data matrix D was averagely scaled to reduce the effects of bias from species with high concentrations. For species with concentration below the detection limit or not detected, the values were replaced by half of the method detection limit. Then a covariance matrix was calculated with dimensions of m × m using the scaled data matrix, and factor loading matrix C and factor score matrix R were determined based on the computation method described by Imamoglu et al.⁵

The reduced *C* and *R* matrices were rotated with non-negative constraints until the sum of square of the negative entries in the back scaled *C* matrix was below 0.0001 or the changing of the sum of squares of the *C* matrix was below 0.001, the iteration process was complete. The coefficient of determination, cumulative percent variance, and Exner function are used to determine the number of sources. The values of VOCs from different sampling sites of receptor region are considered the species concentration matrix (*D*). Calculated steps are applied by Matlab 9.0 software.

62

63 Monte Carlo simulation

64
$$X_{ij} = F_{ij} + C_{ij}F_{ij} \left[\sqrt{2} \operatorname{erf}^{-1}\left(2R_{ij} - 1\right)\right]$$
 (S2)

65 The procedures for Monte Carlo simulation described by Bzdusek *et al.*² and Rachdawong *et al.*⁴
66 were followed. Detailed description of parameters in eqn (S2) is the same as eqn (1) presented in
67 the manuscript. The idea in eqn (S2) is that all VOC species from a given source will vary together
68 which is affected by the same random number. For another source, there will be a different

69 random number and so forth. The variations are caused by random variations in transport paths.
70 The sample concentrations are then the sum of concentrations from all sources. It is assumed that
71 the source profiles do not change in relative weight percentage from source to receptor site. In
72 addition, it is assumed that the variation of each species is proportional to its weight percentage
73 contribution to the VOC profile, *e.g.* constant coefficient of variation for all VOC species in a
74 given source profile.

75 In this study, four *in situ* monitored emission sources and uniformly distributed random contributions were combined to give a simulated VOC data matrix of thirty species and 30 76 samples. Because of P3 and P4 as in situ monitored emission sources from refractory materials 77 production in this park have similar source profiles, the average values of them are used in Monte 78 Carlo simulation. Since each profile is generated using a different random number, the simulated 79 VOC data matrix should reflect a linear combination of VOC sources which contribute 80 independently to the concentrations of VOCs at the sample site. The second term in eqn S2 is a 81 variation term that is characteristic of a normal distribution in the form of the inverse Gaussian 82 error function. The data matrix was subsequently used as input for the receptor model. The factor 83 loadings obtained from the model were compared with the *in situ* monitored emission sources, and 84 the ability of the model to make a prediction in terms of source profiles was evaluated.⁴ 85

- 86
- 87
- 88
- 89
- 90

91 Table S3. Coefficient of determination, cumulative variance (%), and Exner function values for

Spaigas	Coefficient of determination					
Sperces	1	2	3	4		
Ethyl acetate	0.52	0.76	0.82	0.86		
Butyl acetate	0.74	0.87	0.89	0.97		
1-Butanol	0.13	0.16	0.25	0.27		
2-Ethylhexanol	0.19	0.50	0.52	0.61		
Furfuryl alcohol	0.9	0.90	0.91	0.96		
2-Butanone	0.23	0.26	0.41	0.42		
Pentaldehyde	0.27	0.66	0.67	0.67		
Phenol	0.83	0.95	0.95	0.95		
m/p-Cresol	0.78	0.96	0.96	0.98		
o-Cresol	0.77	0.78	0.79	0.97		
Triethylamine	0.96	0.96	0.97	0.98		
Methylcyclopentane	0.38	0.41	0.50	0.52		
3-Methylpentane	0.52	0.52	0.91	0.92		
Benzene	0.38	0.15	0.21	0.24		
Ethylbenzene	0.35	0.50	0.73	0.74		
n-Propylbenzene	0.76	0.76	0.76	0.92		
tert-Butylbenzene	0.50	0.52	0.64	0.64		
1,2,4-Trimethylbenzene	0.10	0.06	0.02	0.03		
2,5-Dimethyl-p-xylen	0.34	0.34	0.40	0.40		
4-Ethyl-m-xylene	0.10	0.66	0.80	0.81		
Toluene	0.62	0.63	0.68	0.77		
<i>m</i> / <i>p</i> -Xylene	0.77	0.79	0.82	0.82		
o-Xylene	0.90	0.93	0.93	0.94		
1,3,5-Trimethylbenzene	0.18	0.56	0.65	0.70		
<i>m/p</i> -Ethyltoluene	0.57	0.76	0.78	0.80		
2-Ethyl-p-xylene	0.77	0.81	0.81	0.98		
2,5-Dimethylfuran	0.72	0.73	0.77	0.88		
Cumulative variance (%)	82.6	86.94	90.23	92.4		
Exner function	0.32	0.24	0.19	0.14		

0	2
ч	1
-	~

different factors from FA-NNC model.

93

94

95



98 Five production processes were labeled as P1 (furan no-bake mold system from foundry production), P2 (cold-box
99 system from foundry production), P3 (burned bricks with about 1600 °C from refractory materials production), P4
100 (unburned bricks with about 200 °C from refractory materials production), and P5 (printing production). The
101 percentages of VOC species from different workshops were calculated by average concentrations of samples.
102 Fig. S1. The percentages of VOC species detected in the complex industrial park
103
104

106 Cosine similarity

107 The similarity value (cos φ coefficient) was calculated quantitatively to estimate the similarity 108 between the two vectors. The one of the various factors obtained from FA-NNC or PMF models 109 was regarded as vector x. The one of the measured profiles of VOCs from the complex industrial 110 park was regarded as vector y. The cos φ value was calculated as following, 111

$$\cos \varphi = \frac{\sum_{k=1}^{m} x_k y_k}{\sqrt{\sum_{k=1}^{m} x_k^2} \sqrt{\sum_{k=1}^{m} y_k^2}}$$
(S3)

112 where x_k and y_k were the factor loadings and percentages of individual VOC species from 1 to k,

113 respectively. Generally, the value larger than 0.7 suggested they had high similarity.⁶







123

Fig. S3. Comparison of the factor loading results from *in situ* monitored receptor matrices and artificial receptor matrices created by Monte Carlo simulation. Error bars representing the standard deviation of the mean for ten runs of FA-NNC model (a) and PMF (b) model using matrices created by Monte Carlo simulation.

129 References

130	1 J. Zhu, R. Newhook, L. Marro and C. C. Chan, Selected volatile organic compounds in residential air in the city
131	of Ottawa, Canada, Environ. Sci. Technol., 2005, 39, 3964-3971, DOI: 10.1021/es050173u.

132 2 P. A. Bzdusek, E. R. Christensen, A. Li and Q. Zou, Source apportionment of sediment PAHs in Lake Calumet,
133 Chicago: Application of factor analysis with nonnegative constraints, *Environ. Sci. Technol.*, 2004, 38,
134 97-103, DOI: 10.1021/es034842k.

135 3 F. Tian, J. Chen, X. Qiao, Z. Wang, P. Yang, D. Wang and L. Ge, Sources and seasonal variation of atmospheric
136 polycyclic aromatic hydrocarbons in Dalian, China: Factor analysis with non-negative constraints
137 combined with local source fingerprints, *Atmos. Environ.*, 2009, 43, 2747-2753, DOI:

138 10.1016/j.atmosenv.2009.02.037.

4 P. Rachdawong and E. R. Christensen, Determination of PCB sources by a principal component method with
 nonnegative constraints, *Environ. Sci. Technol.*, 1997, **31**, 2686-2691, DOI: 10.1021/es970107v.

141 5 I. Imamoglu, K. Li and E. R. Christensen, Modeling polychlorinated biphenyl congener patterns and

- dechlorination in dated sediments from the Ashtabula River, Ohio, USA, *Environ.Toxicol. Chem.*, 2002,
 21, 2283-2291, DOI: 10.1002/etc.5620211105.
- 6 M. Cai, Y. Lin, M. Chen, W. Yang, H. Du, Y. Xu, S. Cheng, F. Xu, J. Hong, M. Chen and H. Ke, Improved
 source apportionment of PAHs and Pb by integrating Pb stable isotopes and positive matrix factorization
 application (PAHs): A historical record case study from the northern South China Sea, *Sci. Total.*
- 147 *Environ.*, 2017, 609, 577-586, DOI: 10.1016/j.scitotenv.2017.07.190.