PM2.5 and O3 in an enclosed basin, the Guanzhong Basin of Northern China: Insights into distributions, appointment sources, and transport pathways

Xiaofei Li [a,b,c,d*], Jingning Guo [a], Xuequan Wan [a], Zhen Yang [a], Lekhendra Tripathee [d], Feng Yu [a], Rui Zhang [a], Wen Yang [a], Qiyuan Wang [b,c]

[a] *School of Environmental Science and Engineering, Shaanxi University of Science and Technology, Xi'an 710021, China*

[b] *National Observation and Research Station of Regional Ecological Environment Change and Comprehensive Management in the Guanzhong Plain, Shaanxi, Xi'an 710061, China*

[c] *Key Laboratory of Aerosol Chemistry and Physics, State Key Laboratory of Loess and Quaternary Geology, Institute of Earth Environment, Chinese Academy of Sciences, Xi'an 710061, China*

[d] *State Key Laboratory of Cryospheric Science, Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences, Lanzhou 730000, China*

[*] **Corresponding Authors:**

E-mail: lixiaofei@ieecas.cn **(X.F. Li)**

This supporting information consists of 18 pages, including detailed methods, seven

figures, three tables, and literature cited.

Text S1. Before and after sampling, processing of filter membrane, and storage of samples.

PM$_{2.5}$ was continuously collected by a medium-flow sampler (HC-1010, Qingdao, China) at a flow rate of 100 L min^{-1} from 08:00 a.m. to 07:00 a.m. the next day (Beijing time). Before sampling, the quartz microfiber filters (Φ90 mm) were pre-baked at 450°C for 4 h to remove residual organics and other impurities; during sampling, masks and gloves were worn, and clean tweezers were used to clip the filter membrane; after collection, the samples were individually sealed and preserved in darkness at -20°C for further analysis. Before and after sampling, filters were weighed using an analytical balance after balancing for 48 h in a drying dish. At the beginning and end of sampling, one blank sample was collected as a control.

Text S2. Details of Desert Research Institute (DRI) Model 2015 carbon analyzer.

When the samples were being analyzed, a 0.5 cm^2 sample was extracted from each filter and loaded into the sample boat, then passed into the 100% He. The filter samples were gradually heated up to 120°C, 250°C, 450°C, and 550°C under no-oxygen conditions to volatilize organic carbon into OC1, OC2, OC3, and OC4. Then, they were passed into the mixture of 2% O$_2$ and 98% He and continued to be heated up to 550°C, 700°C, and 800°C to volatilize the elemental carbons EC1, EC2, and EC3 in the sample. The released organics were catalyzed and oxidized to CO$_2$ under the action of MnO$_2$, then reduced to CH$_4$ in the reduction furnace, and finally quantitatively detected by a flame ionization detector (FID). During the whole process of heating up, a 633 nm laser was used to irradiate the sample and measure the pyrolyzed organic carbon (OPC). OC and EC were defined as follows[25]:

$$OC = OC1 + OC2 + OC3 + OC4 + OPC \qquad (S1)$$

$$EC = EC1 + EC2 + EC3 - OPC \qquad (S2)$$

Text S3. The calculation method of primary organic carbon (POC) and secondary organic carbon (SOC).

The EC tracer method was used to estimate the concentration of SOC and POC; the formulas are expressed as follow[62, 63]:

$$SOC = OC - EC \times (OC/EC)_{min} \tag{S3}$$

$$POC = EC \times (OC/EC)_{min} \tag{S4}$$

where OC is the mass concentration of the total organic carbon, $\mu g\ m^{-3}$, and $(OC/EC)_{min}$ is the minimum value of OC/EC.

Text S4. Detailed pretreatment process for water-soluble ions (WSIIs).

First, a 2*12 mm sample was cut from the filter membrane sample and placed in a glass flask, 20 mL of ultrapure water was added, and ultrasonic extraction was used for 30 min; the temperature did not exceed 20°C during the extraction process. Thereafter, a 0.45 μm injection filter was filtered and then measured separately using an ion chromatograph (Dionex, ICS-2500; Dionex, DX-600).

Text S5. The calculation of WSIIs acidity, alkalinity, sulfur oxidation rates (SOR) and nitrogen oxidation rates (NOR).

The charge balance of the anion and cation can be used to estimate the pH of $PM_{2.5}$. The AE was defined as the anionic equivalent and the CE as the cation equivalent. With AE/CE>1, $PM_{2.5}$ is acidic, and otherwise, it is alkaline[64]. The formulas for AE and CE are defined as the following:

$$AE = \frac{F^-}{19} + \frac{Cl^-}{35.5} + \frac{NO_3^-}{62} + \frac{SO_4^{2-}}{48} \tag{S5}$$

$$CE = \frac{K^+}{39} + \frac{Ca^{2+}}{20} + \frac{Na^+}{23} + \frac{Mg^{2+}}{12} + \frac{NH_4^+}{18} \tag{S6}$$

In order to gain insight into the conversion process of SO_2 to SO_4^{2-} and NO_2 to NO_3^-, the ratio of sulfur oxidation rate (SOR) and nitrogen oxidation rate (NOR) is usually calculated to determine the degree of secondary conversion.

In order to understand the conversion process of SO_2 to SO_4^{2-} and NO_2^- to NO_3^-, the degree of aerosol secondary conversion is usually judged by calculating SOR and NOR. The higher the SOR and NOR values, the higher the degree of secondary conversion of aerosols[65]. The SOR and NOR were calculated using the following formulas：

$$SOR = \frac{c(SO_4^{2-})/96}{c(SO_4^{2-})/96 + c(SO_2)/64} \tag{S7}$$

$$NOR = \frac{c(NO_3^-)/62}{c(NO_3^-)/62 + c(NO_2)/46} \tag{S8}$$

Text S6. Post-processing of positive matrix factorization (PMF) model.

The positive matrix factor method (PMF) is a commonly used receptor model for source analysis of atmospheric particulate matter[66]. The model allowed the users to review the chemical concentration data, including the signal-to-noise ratio (S/N) and the specified importance of the chemicals as either "strong," "weak," or "bad"[67]. According to the model assumption that there was little change in the transport of the tracers from source to receptor, some chemicals that were less dependent on the source strength would be excluded. The concentrations of chemicals below the method detection limits (MDLs) were substituted with half MDL, and their uncertainties were calculated with the formula U = K × C, where K represents analytical uncertainty, and C is the measured chemical concentrations[68, 69]. Missing values and uncertainties were imputed by the median concentration of the component and four times the median values to decrease their influence on the results. Rotation ambiguity was controlled by the FPEAK and FKEY parameters[70].

The analytic formula is defined as the following:

$$X_{ij} = \sum_{k=1}^{p} G_{ik}F_{kj} + E_{ij} \tag{S9}$$

$$Q = \sum_{i=1}^{n} \sum_{j=1}^{m} \left[\frac{X_{ij} - \sum_{k=1}^{p} G_{ik}F_{kj}}{u_{ij}} \right]^2 \tag{S10}$$

where X_{ij} is the mass concentration of substance j in the sample i, G_{ik} is the relative contribution of the pollution source k to the sample i, F_{kj} is the content of substance j in the pollution source k, p is the total number of pollution sources, E_{ij} is the measured mass concentration of substance j in sample i and the residual difference of its analytical value, and u_{ij} is the uncertainty of the substance j in the sample i.

In the PMF model, the data were processed as follows: the average values below MDL were removed, the substance concentrations below MDL were converted to half

of the MDL, and the missing values were replaced by the median of the substance concentration.

The uncertainty calculation formula of the components below the MDL is expressed as the following:

$$Unc = 5/6 \times MDL \tag{S11}$$

The formula of the substance concentration above the MDL and the uncertainty was expressed as follows:

$$Unc = \sqrt{(EF \times c)^2 + MDL^2} \tag{S12}$$

where Unc represents the uncertainty; EF represents the measurement error ratio of each component; c represents the measured concentration of the component; and MDL represents the method detection limit.

At the same time, in order to ensure the scientific nature of the PMF model, the components participating in the simulation are screened: when the signal/noise ratio (S/N) is less than 0.2, the relevant components do not participate in the calculation; when S/N is greater than 0.2 but lower than 2, the weight of the components should be reduced to "Wake"; when S/N is greater than 2, the weight of the components participating in the regression is set to "Strong."

Text S7. Details of potential source contribution function (PSCF) model and concentration weight trajectory (CWT) model.

Meteorological data were obtained from GDAS data provided by the National Centers for Environmental Prediction (NCEP) of the National Oceanic and Atmospheric Administration (NOAA). Using the Euclidean distance method, all the trajectories reaching the hit point were clustered, and the number of clusters was determined using the total spatial variance (TSV) method. XN and QL were divided into $1° \times 1°$ grids, respectively, and the 500 m simulation height was selected as the average flow field of the boundary layer to reflect the flow characteristics of the surface airflow area and reduce the influence of ground friction[71]. The simulated trajectory lasted 48 hours with a time resolution of 1 hour. When using the PSCF model for $PM_{2.5}$,

OC, EC, O3, and WSIIs source analysis, the selected threshold was 70% of the mean.

PSCF was used to determine the contribution of each grid to the affected region by calculating the ratio of the residence time of the total trajectory in the pathway region[72]. The selected study area was divided into $i{\times}j$ grids. PSCF is defined as the ratio of the number of contaminated trajectories passing through these grids (m_{ij}) to the total number of trajectories (n_{ij}), and the formula is expressed as the following:

$$\text{PSCF} = \frac{m_{ij}}{n_{ij}} \tag{S13}$$

Since PSCF only reflects the proportion of the contaminated trajectories in the grids and not the contamination extent of the trajectory, CWT was used to weigh the trajectory of the relevant concentrations. CWT was calculated using the following equation[73]:

$$\text{CWT}_{ij} = \frac{\sum_{k=1}^{n_{ij}}(C_k \times a_{ijk})}{\sum_{k=1}^{n_{ij}} a_{ijk}} \tag{S14}$$

Where CWT_{ij} ($\mu g\ m^{-3}$) is the average pollution weight concentration of the grid (i, j), n_{ij} is the total number of traces passing through the grid (i, j), C_k is the corresponding contaminant concentration when the trajectory k passes through the grid, and a_{ijk} is the time when the trajectory k stays in the grid (i, j).

**Text S8. Values of W_{ij} in Xi'an (XN) and Qinling (QL).**

The value of W_{ij} was determined by the average of all junctions of all trajectories through all grids, and the average value shall be an integer greater than the calculated value. The numbers of all junctions for all trajectories and all grids were 49392 and 945 in XN, respectively, so its average value was 54, and the formula and W_{ij} in XN was expressed as (15). The numbers of all junctions for all trajectories and all grids were 37534 and 945 in QL, respectively, so its average value was 4, and the value of W_{ij} in QL was shown in formula (16).

$$W_{ij} = \begin{cases} 1 & n_{ij} > 216 \\ 0.70 & 54 < n_{ij} \leq 216 \\ 0.42 & 27 < n_{ij} \leq 54 \\ 0.17 & n_{ij} \leq 27 \end{cases} \tag{S15}$$

$$W_{ij} = \begin{cases} 1 & n_{ij} > 160 \\ 0.70 & 40 < n_{ij} \leq 160 \\ 0.42 & 20 < n_{ij} \leq 40 \\ 0.17 & n_{ij} \leq 20 \end{cases} \tag{S16}$$

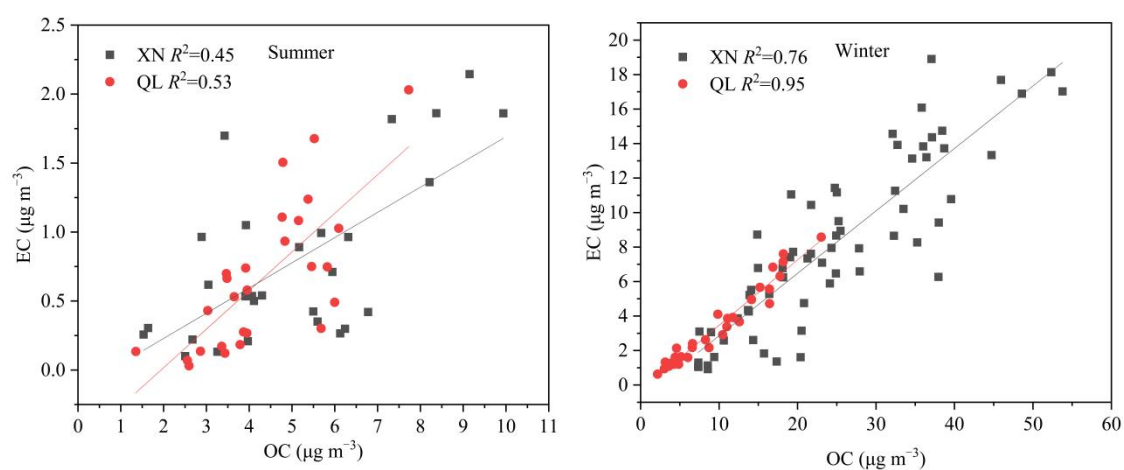$$\text{WPSCF}_{ij} = \text{PSCF} \times W_{ij} \tag{S17}$$

$$\text{WCWT}_{ij} = \text{CWT} \times W_{ij} \tag{S18}$$
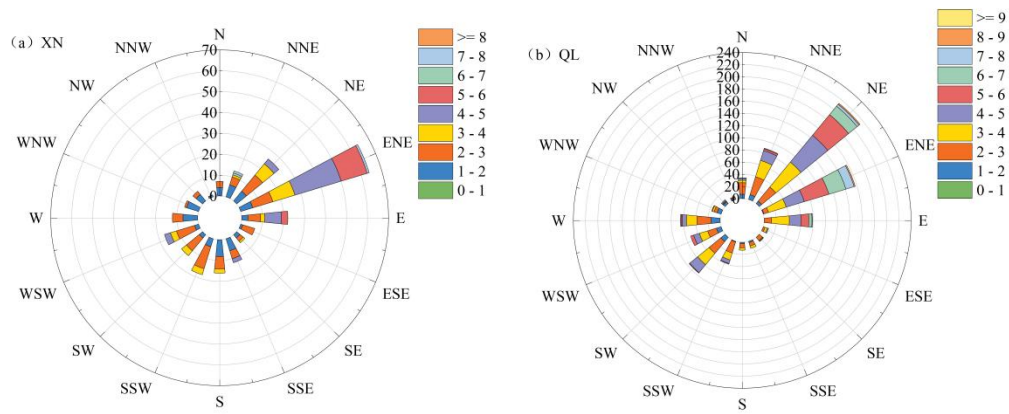
Figure S1 The wind rose diagram of XN and QL in summer.

Figure S2. Correlation of OC and EC in winter and summer in XN and QL.

Table S1. Average mass concentration of 10 water-soluble ions (WSIIs) at XN and QL and the ratio to PM$_{2.5}$.

Water-soluble ions ($\mu g\ m^{-3}$)	XN	QL
F$^-$	0.130	0.046
Cl$^-$	0.930	0.296
NO$_2^-$	0.279	0.025
SO$_4^{2-}$	2.268	3.051
NO$_3^-$	1.762	0.708
Na$^+$	0.086	4.948
NH$_4^+$	2.356	1.008
K$^+$	0.694	0.254
Mg^{2+}	0.583	0.058
Ca^{2+}	1.270	0.460
WSIIs	10.359	10.860
WSIIs/PM$_{2.5}$	19.40%	39.37%
AE/CE	0.410	0.280
SOR	0.200	0.516
NOR	0.051	0.139

Table S2. WSIIs correlation analysis in XN and QL.

XN	F^-	Cl^-	NO_2^-	SO_4^{2-}	NO_3^-	Na^+	NH_4^+	K^+	Mg^{2+}	Ca^{2+}
F^-	1									
Cl^-	0.149	1								
NO_2^-	0.081	0.015	1							
SO_4^{2-}	0.020	0.061	0.016	1						
NO_3^-	0.055	0.008	0.087	0.588	1					
Na^+	0.009	0.110	0.813	0.179	0.005	1				
NH_4^+	0.035	0.037	0.064	0.222	0.305	*	1			
K^+	0.029	0.037	0.004	0.073	0.023	0.577	0.042	1		
Mg^{2+}	0.100	0.033	0.001	0.031	0.036	*	0.094	0.207	1	
Ca^{2+}	0.095	0.135	0.012	0	0.007	0.005	0.026	0.043	0.460	1

| QL	F^-	Cl^-	NO_2^-	SO_4^{2-}	NO_3^-	Na^+	NH_4^+	K^+	Mg^{2+}	Ca^{2+}
F^-	1									
Cl^-	0.011	1								
NO_2^-	0.003	0.010	1							
SO_4^{2-}	0.140	0.005	0.053	1						
NO_3^-	0.026	0.007	0.309	0.244	1					
Na^+	0.008	0.082	0.212	0.039	0.003	1				
NH_4^+	0.073	0.002	0.178	0.708	0.546	0.002	1			
K^+	0.160	0.321	0.009	0.258	0.010	0.147	0.189	1		
Mg^{2+}	0.152	0.005	0.002	0.202	0.003	0.039	0.005	0.136	1	
Ca^{2+}	0.183	0.009	0.004	0.150	0.001	0.016	0.006	0.140	0.931	1

* Denoted missing values.
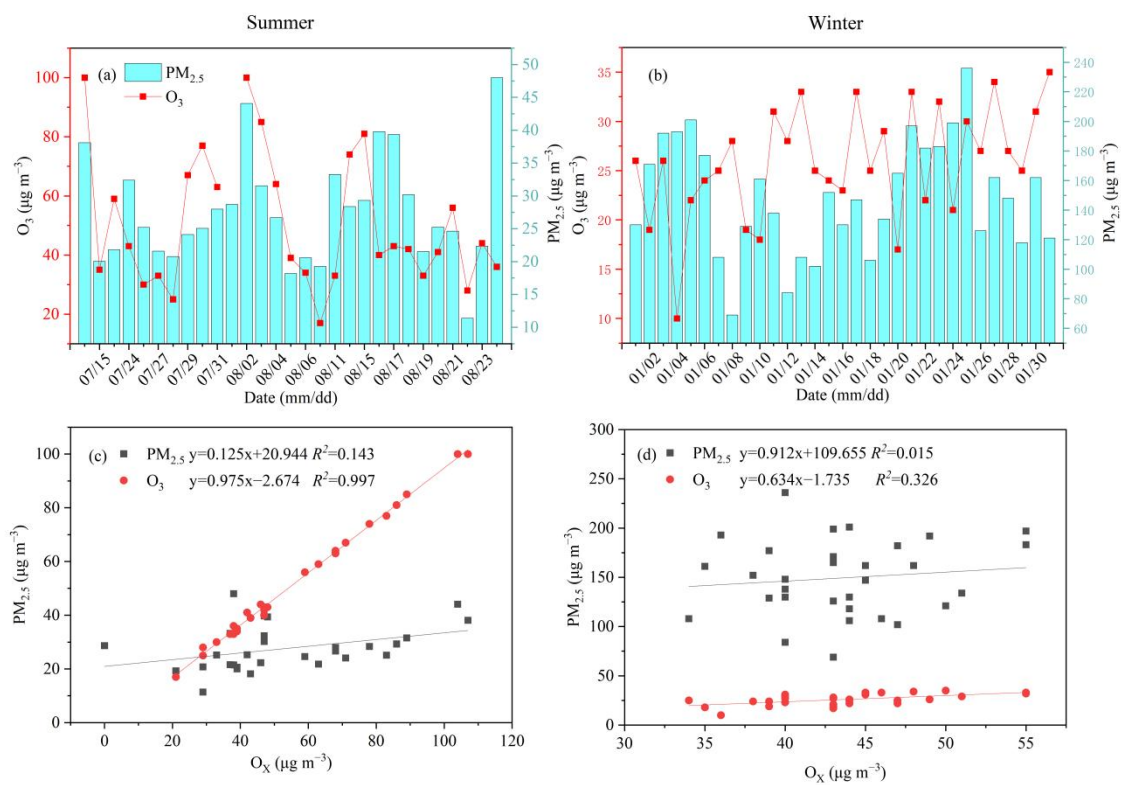
Figure S3. The relationship between O_3 and $PM_{2.5}$ in QL.
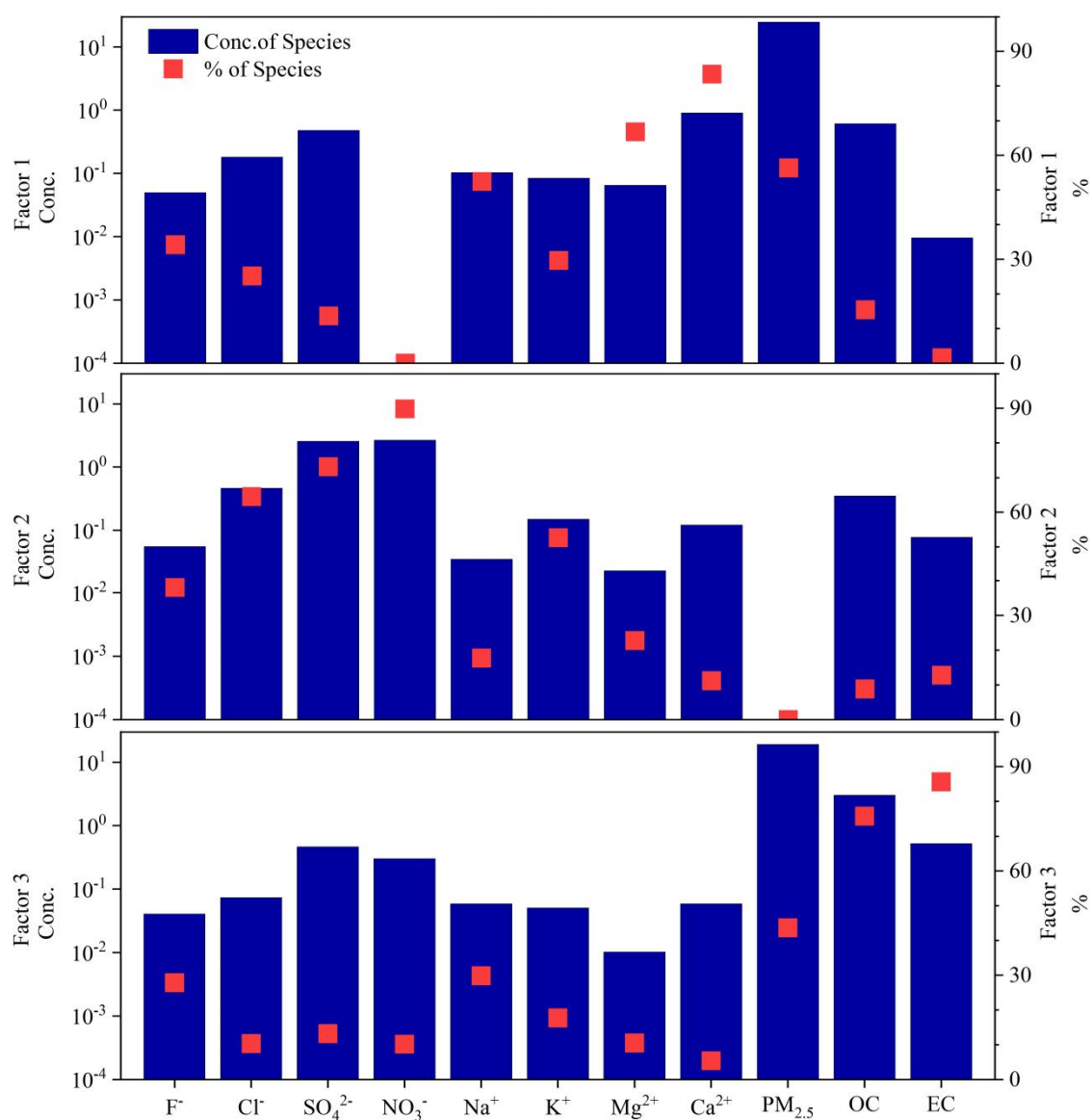
Figure S4. PMF source resolution map of the environmental elements in XN during the sampling periods.
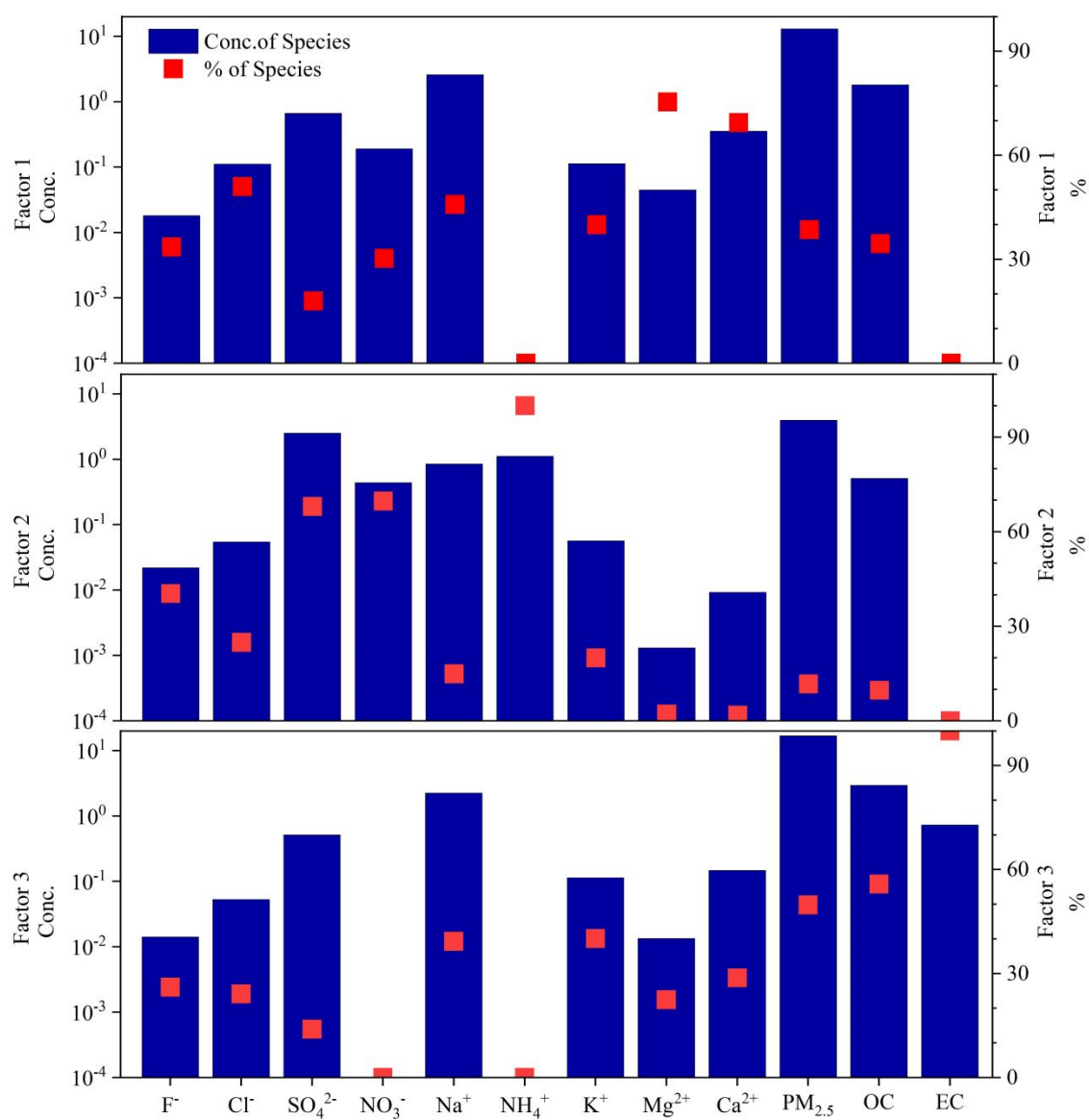
Figure S5. PMF source resolution map of the environmental elements in QL during the sampling periods.

Table S3 Calculation results of PMF models in XN and QL during the sampling periods.

QL	Factor Profiles (conc. of species)			Factor Profiles (% of species sum)		
Species	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 3
F^-	0.049933	0.04073	0.055551	34.15062853	27.85642962	37.99294185
Cl^-	0.18044	0.073885	0.46224	25.18124664	10.31099761	64.50775575
SO_4^{2-}	0.47965	0.46341	2.5685	13.65917142	13.1966989	73.14412967
NO_3^-	0	0.30234	2.6666	0	10.18343247	89.81656753
Na^+	0.10286	0.058562	0.034765	52.42956975	29.850092	17.72033825
K^+	0.084382	0.050448	0.14979	29.64724896	17.72468555	52.62806549
Mg^{2+}	0.065053	0.010207	0.022148	66.78404238	10.47860545	22.73735217
Ca^{2+}	0.90573	0.058613	0.12142	83.41875713	5.398323575	11.18291929
$PM_{2.5}$	24.663	19.101	0	56.35453798	43.64546202	0
OC	0.61336	3.0126	0.35057	15.42450327	75.75951898	8.815977749
EC	0.0095946	0.52202	0.077896	1.574148177	85.64576235	12.78008947
QL	Factor Profiles (conc. of species)			Factor Profiles (% of species sum)		
Species	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 3
F^-	0.018146	0.021822	0.014076	33.5763452	40.37821035	26.04544445
Cl^-	0.11122	0.054291	0.052556	51.00267349	24.89647677	24.10084974
SO_4^{2-}	0.66223	2.5046	0.51302	17.99611397	68.0625569	13.94132913
NO_3^-	0.18989	0.4376	0	30.26183684	69.73816316	0
Na^+	2.5925	0.84314	2.2235	45.81084758	14.8987302	39.29042222
NH_4^+	0	1.1092	0	0	100	0
K^+	0.11266	0.056288	0.11295	39.96480997	19.96750598	40.06768406
Mg^{2+}	0.044789	0.0012963	0.013331	75.38167136	2.181724544	22.4366041
Ca^{2+}	0.35419	0.0092232	0.14671	69.43224696	1.808033824	28.75971922
$PM_{2.5}$	13.005	3.9393	16.803	38.53641625	11.67293383	49.79064992
OC	1.814	0.50916	2.9339	34.50597863	9.685261344	55.80876003
EC	0	0	0.72777	0	0	100
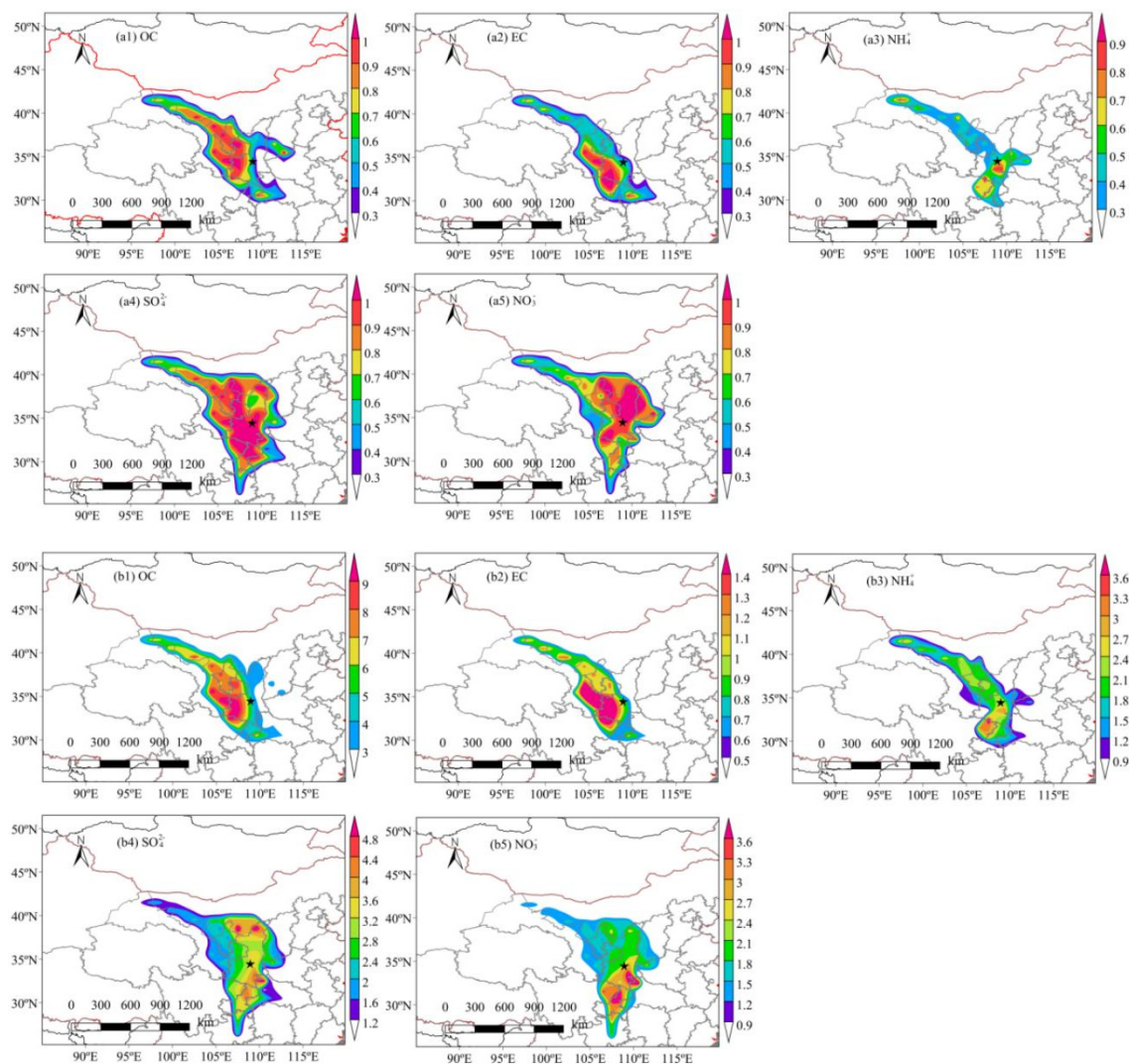
Figure S6 (a) The WPSCF and (b) WCWT of carbon components (OC and EC) and major WSIIs during the sampling periods in XN.
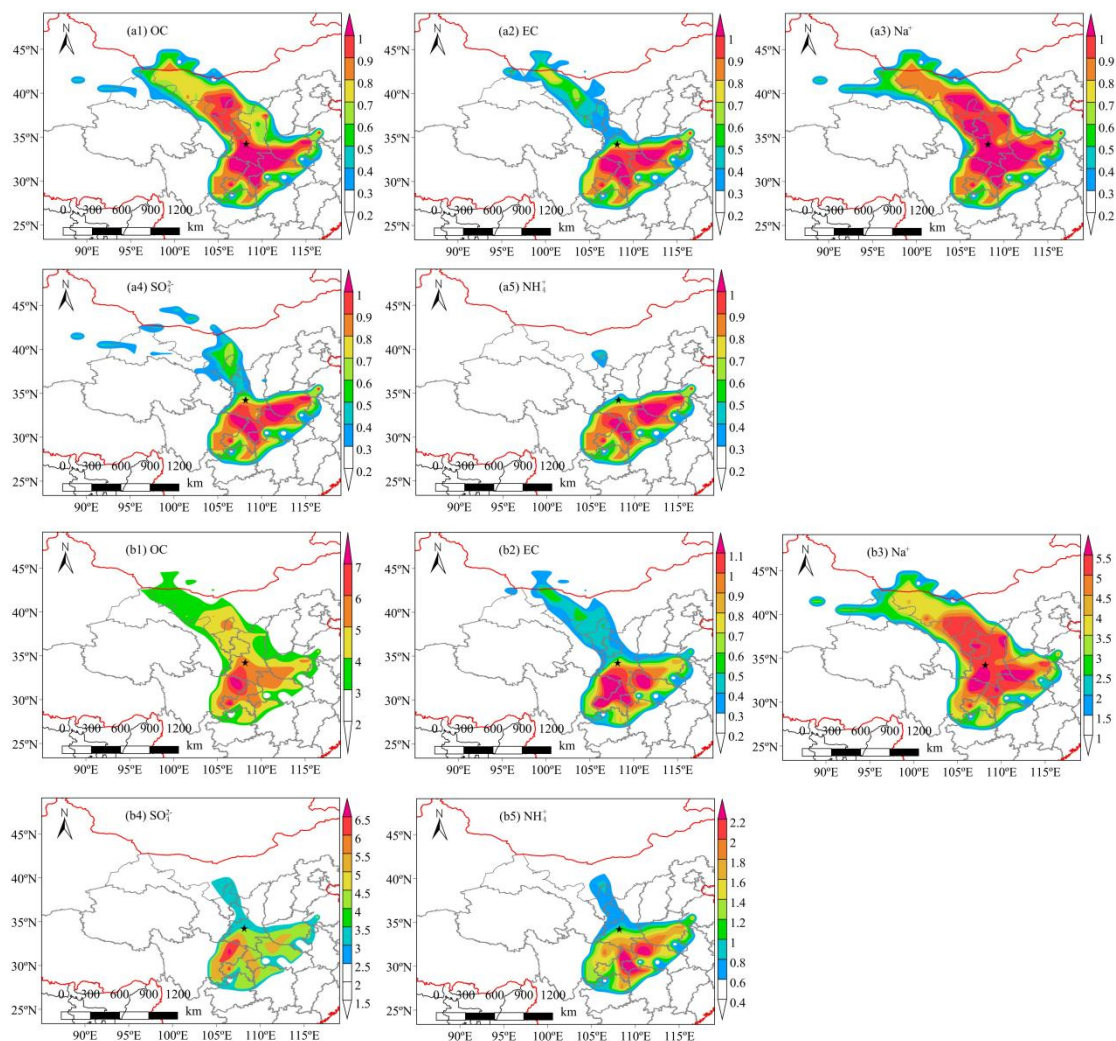
Figure S7 (a) The WPSCF and (b) WCWT of carbon components (OC and EC) and major WSIIs (Na$^+$, SO$_4^{2-}$, and NH$_4^+$) during the sampling periods in QL.

References

[25] Chow, J. C.; Watson, J. G.; Chen, L. W. A.; Chang, M. O.; Robinson, N. F.; Trimble, D.; Kohl, S., The IMPROVE_A temperature protocol for thermal/optical carbon analysis: maintaining consistency with a long-term database. *Journal of the Air & Waste Management Association* **2007,** 57, (9), 1014-1023.

[62] Wu, C.; Wu, D.; Yu, J. Z., Estimation and uncertainty analysis of secondary organic carbon using 1 year of hourly organic and elemental carbon data. *Journal of Geophysical Research: Atmospheres* **2019,** 124, (5), 2774-2795.

[63] Wu, C.; Yu, J. Z., Determination of primary combustion source organic carbon-to-elemental carbon (OC/EC) ratio using ambient OC and EC measurements: secondary OC-EC correlation minimization method. *Atmospheric Chemistry and Physics* **2016,** 16, (8), 5453-5465.

[64] Xu, L.; Chen, X.; Chen, J.; Zhang, F.; He, C.; Zhao, J.; Yin, L., Seasonal variations and chemical compositions of $PM_{2.5}$ aerosol in the urban area of Fuzhou, China. *Atmospheric Research* **2012,** 104, 264-272.

[65] Shon, Z. H.; Ghosh, S.; Kim, K. H.; Song, S. K.; Jung, K.; Kim, N. J., Analysis of water-soluble ions and their precursor gases over diurnal cycle. *Atmospheric Research* **2013,** 132, 309-321.

[66] Paatero, P., The multilinear engine—a table-driven, least squares program for solving multilinear problems, including the n-way parallel factor analysis model. *Journal of Computational and Graphical Statistics* **1999,** 8, (4), 854-888.

[67] Jaeckels, J. M.; Bae, M. S.; Schauer, J. J., Positive matrix factorization (PMF) analysis of molecular marker measurements to quantify the sources of organic aerosols. *Environmental Science & Technology.* **2007,** 41, (16), 5763-5769.

[68] Prendes, P.; Andrade, J. M.; López-Mahía, P.; Prada, D., Source apportionment of inorganic ions in airborne urban particles from Coruña city (N.W. of Spain) using positive matrix factorization. *Talanta* **1999,** 49, (1), 165-178.

[69] Wang, Y.; Li, S.; Wang, M.; Sun, H.; Mu, Z.; Zhang, L.; Li, Y.; Chen, Q., Source apportionment of environmentally persistent free radicals (EPFRs) in $PM_{2.5}$ over Xi'an, China. *Science of the Total Environment.* **2019,** 689, 193-202.

[70] Paatero, P.; Hopke, P. K.; Song, X. H.; Ramadan, Z., Understanding and controlling rotations in factor analytic models. *Chemometrics and Intelligent Laboratory Systems* **2002,** 60, (1), 253-264.

[71] Polissar, A. V.; Hopke, P. K.; Harris, J. M., Source regions for atmospheric aerosol measured at Barrow, Alaska. *Environmental Science & Technology* **2001,** 35, (21), 4214-4226.

[72] Sulaymon, I. D.; Mei, X.; Yang, S.; Chen, S.; Zhang, Y.; Hopke, P. K.; Schauer, J. J.; Zhang, Y., $PM_{2.5}$ in Abuja, Nigeria: Chemical characterization, source apportionment, temporal variations, transport pathways and the health risks assessment. *Atmospheric Research* **2020,** 237, 104833.

[73] Hsu, Y. K.; Holsen, T. M.; Hopke, P. K., Comparison of hybrid receptor models to locate PCB sources in Chicago. *Atmospheric Environment* **2003,** 37, (4), 545-562.