Supplementary material

Distribution, Source Apportionment, and Assessment of Heavy Metal Pollution in the Yellow River Basin, Northwestern China

Cheng Ma [1, 2], Menglu Wang [1,3], Qian Li [4], Mohammadtaghi Vakili[5], Yijing Zhang [1], Shengqiang Hei [1, 2], Li Gao [1, 2], Wei Wang[6], Dengchao Liu [1, 2*]

[1] School of Geography and Planning, Ningxia University, Yinchuan 750021, China

[2] China-Arab Joint International Research Laboratory for Featured Resources and Environmental Governance in Arid Region, Yinchuan 750021, China

[3] Ningxia Technical College of Wine and Desertification Prevention, Yinchuan 750021, China

[4] Ningxia Hui Autonomous Region Food Inspection and Research Institute, Yinchuan 750000, China

[5] ORLEN UniCRE, a.s., Revoluční 1521/84, Ústí nad Labem 40001, Czech Republic

[6] State Key Laboratory of Plateau Ecology and Agriculture, Qinghai University, Xi'ning 810016, China

*Corresponding author.

E-mail: ldcxfp@nxu.edu.cn (Dr. Dengchao LIU Professor).

Text S1 Assessment of HMs in surface water and sediment

The pollution status of HMs was evaluated using the P_i and the P_N for surface water, respectively, as described below (Huang et al., 2018).

$$P_i = \frac{C_i}{S_i} \quad (1)$$

$$P_N = \sqrt{\frac{max(P_i)^2 + ave(P_i)^2}{2}} \quad (2)$$

In this equation, C_i denotes the calculated *i* concentration and S_i is the value of the EQSSW Category III threshold, max (P_i) denotes the highest Pi of all metals, aver (P_i) is the arithmetic mean of all HMs pollution indices via the P_i at each sampling point. The values of P_i and P_N were classified as shown in Table S2.

The *Igeo*, initially proposed by Muller (Aydin et al., 2023) has been extensively utilized in evaluating the level of HMs in sediments (Liao et al., 2022; Li et al., 2024). The equation follows:

$$Igeo = log_2\left[\frac{C_n}{1.5B_n}\right] \quad (3)$$

In this equation, C_n denotes the metal in sediment sample concentration (*n*), B_n represents the background value of the HM (*n*), and 1.5 is a coefficient introduced to account for natural fluctuations in background values due to lithological variations (Ke et al., 2017). The *Igeo* is categorized into six levels, as presented in Table S2.

The CF and *PLI* were utilized to evaluate the enrichment of individual metals and the extent of HM contamination in sediments, referencing the mean BSS (Xie et al., 2022). The equation are as follows:

$$CF = \frac{C_s}{Cb} \quad (4)$$

$$PLI = \sqrt[n]{CF_1 \times CF_2 \times CF_3 ... \times CF_n} \quad (5)$$

where Cs represents the surveyed level of metal (i), *Cb* indicates the BBS (Xie et al., 2022), and *n* denotes the count of metals under consideration. For pollution classification see Table S2.

Text S2 Positive matrix factorization

PMF is a mathematical receptor model grounded in factor analysis, initially proposed by Dutch scholar Paatero et al. in 1993 (Sun et al., 2020). PMF uses the least squares method to determine the principal pollution sources and their respective contribution ratios. The calculation equations are elucidated as follows:

$$X_{ij} = \sum_{k=1}^{p} \left(G_{ik} \times F_{jk}\right) + E_{ij} \tag{6}$$

where X_{ij} is the sample matrix of concentrations; G_{ik} represents the concentration within species i in source k; F_{jk} is the response of the signal source k to the contribution of the sample j; E_{ij} denotes the residual for one sample each.

The effectiveness of the PMF model is assessed based on the minimized objective function Q value and the concordance between the observed concentrations of HMs and those predicted by the PMF model, thereby facilitating the determination of an optimal factor number. The calculation is expressed as follows:

$$Q = \sum_{i=1}^{n} \sum_{j=1}^{m} \left(\frac{E_{ij}}{U_{ij}}\right)^2 \tag{7}$$

$$U_{ij} = \begin{cases} \frac{5}{6} \times MDL, & (c \leq MDL) \\ \sqrt{(\delta \times c)^2 + (0.5 \times MDL)^2} & (c > MDL) \end{cases} \tag{8}$$

where Q is the target function; n denotes the count of samples; m represents the count of elements; U_{ij} is an uncertain value; δ denotes the Error Fraction; c is the value of HMs; and MDL indicates the concentration detection limit.

When executing the model, 4 to 7 factors were selected for 20 iterations of the calculation until the optimal number of factors was determined, depending on the species-proportional residuals, signal-to-noise ratio, correlation between observations and PV values, and Q-values, respectively. When the number of factors reached 5, the variance of the Q-values showed the least dispersion, with most of the residuals between -3 and 3.

The correlation analysis (Fig. S1) and PCA results unveiled the multifaceted sources of HMs pollution in surface water. While PCA can discern the sources of elements, it falls short in quantifying their relative contributions. To further explore the origins and relative proportions of HMs present in the surface water within the designated study area, the application of the PMF model facilitated source identification and analysis.

The model was executed as described in the supporting information. As a result, the investigation determined that the best factor number of PMF source resolution was 5. Throughout the wet season, Factors 1 through 5 collectively represented, on average, 63.13%, 3.47%, 26.13%, 6.19%, and 1.08% of the cumulative element concentration, respectively. Similarly, during the dry season, Factors 1 to 5 contributed 9.76%, 16.65%, 40.81%, 31.15%, and 1.63% of the total element concentrations, respectively.

Text S3 Principal component analysis

PCA is a method utilized to diminish the dimensionality of datasets, thereby enhancing interpretability while concurrently minimizing information loss (Jolliffe and Cadima, 2016). PCA employs principal factor loadings derived from chemical elements to discern source types (Wen et al., 2022). As follows:

$$Y = \sum_{h=1}^{n} a_h X_h + b \quad (9)$$

where Y represents total pollution levels, n is the count of factors; a_h signifies thecoefficient of normalized regression for h; X_h denotes the h-factor; b indicates the regression constant.

The independent and dependent variables were standardized before conducting regression analysis on the normalized equations:

$$K = \sum_{i=1}^{n} B_i X_i \quad (10)$$

where B_i is the regression coefficient. The contribution of the i^{th} source can be expressed as:

$$\mathrm{i}(\%) = B_i \ / \sum B_i \times 100 \tag{11}$$

The Kaiser-Meyer-Olkin (KMO) test was utilized to evaluate the adequacy of the data for conducting PCA. In this study, the KMO values for HMs during wet and dry seasons were 0.61 and 0.56, respectively, surpassing the threshold of 0.5. Furthermore, the significance level (p-value) below 0.05 in the Bartlett test suggests that the dataset is suitable for PCA (Xiao et al., 2021).

Table S1 Geographic information of the sampling sites in Ningxia.

Site ID	Surface water type	Sampling site	Longitude (°E)	Latitude (°N)
1	Yellow River Mainstream	Yongning	106.216056	38.137278
2	Yellow River Mainstream	Xingqing	106.577667	38.359056
3	Yuehai Lake	Jinfeng	106.193056	38.559083
4	Dianong river	Xixia-jinfeng	106.139056	38.438806
5		Jinfeng- xingqing	106.256944	38.500278
6		Jinfeng	106.215833	38.500333
7		Jinfeng-helan	106.240833	38.613472
8		Hlan- pingluo	106.322778	38.720278
9	Mingcuihu Lake	Xingqing	106.368361	38.408472

10	Drain	Lingwu East Ditch	106.339111	38.289611
11		First drain	106.240806	38.202556
12		Mesocolcus	106.346750	38.326944
13		YongQing ditch	106.344250	38.331167
14		Yonger dry ditch	106.330833	38.417222
15			106.464667	38.415583
16		Second drain	106.356111	38.484722
17			106.502278	38.549139
18		Yinxin dry ditch	106.327778	38.524444
19			106.502056	38.584778
20		Forty-two dry ditch	106.249000	38.607472
21			106.464806	38.703028
22	Yellow River Mainstream	Pingluo	106.666111	38.812944
23	Yellow River Mainstream	Huinong	106.773889	39.371389
24	Dostu River	Pingluo	106.906611	39.079528
25	Sand Lake	Pingluo	106.347750	38.822222
26	Kushui River	litong	106.205833	38.065556
27	Qingshui River	Zhongning	105.541889	37.486917
28	Kushui River	Yanchi	106.766278	37.167750
29	Hongliu Ditch	Zhongning	105.950000	37.471944
30		Zhongning	105.882078	37.563975
31	Kushui River	Sunjiatan	106.298333	37.536667
32	Qingshui River	Litong	106.160833	37.980278
33	Xiangshan Lake	Shapotou	105.200333	37.490417
34	Yellow River Mainstream	Shapotou	105.124111	37.486306
35	Yellow River Mainstream	Qingtongxia	105.926222	37.830861
36	Qingshui River	Yuanzhou	106.276194	35.924194
37		Yuanzhou	106.203167	36.201394
38		Yuanzhou	106.149194	36.322472
39		Tongxin	105.982222	36.860000
40		Hongsi Fort	105.786583	37.191694
41	Gourd River	Xiji	105.790528	35.616944
42	Jing River	Jingyuan	106.336500	35.393167
43		Jingyuan	106.444306	35.450806
44	Yu River	Longde	106.167361	35.653889
45		Longde	105.826194	35.554917
46	Ru River	Pengyang	106.345833	35.842611
47		Pengyang	106.907139	35.790500
48	Pu River	Pengyang	106.944167	36.055194
49	Hong River	Pengyang	106.809444	35.730278

Table S2 Pollution classification standards for P_i, P_N, CF, I_{geo}, and PLI.

P_i	Pollution levels	P_N	Pollution levels	CF	Degree of enrichment	Igeo	Pollution degree	PLI	Degree of enrichment
$P_i \leq 1$	No pollution (clean)	$P_N \leq 1$	No pollution (clean)	CF < 1	Low contamination	0≤ Igeo<1	Uncontaminated to moderately contaminated	PLI < 1	No contamination
$1 < P_i \leqslant 2$	Low pollution	$1 < P_N \leq 5$	Low pollution	1 ≤ CF < 3	Moderate contamination	1≤ Igeo<2	moderately contaminated	PLI ≥ 1	Contamination
$2 < P_i \leqslant 3$	Moderate pollution	$5 < P_N \leq 10$	Moderate pollution	3 ≤ CF < 6	Significant contamination	2≤ Igeo<3	moderately contaminated to heavily contaminated		
$3 < P_i$	High pollution	$10 < P_N$	High pollution	6 ≤ CF	Highly contamination	3≤ Igeo<4	heavily contaminated		
						4≤ Igeo<5	heavily contaminated to extremely contaminated		
						Igeo ≥5	extremely contaminated		

Table S3 Comparison of HMs concentrations in surface water of the world.

	Cr	Mn	Co	Ni	Cu	Zn	As	Cd	Sb	Hg	Tl	Pb	Sr	Ref
Surface water (μg/L)														
Ningxia	4.69	54.12	0.79	4.44	13.23	19.84	4.00	0.05	0.48	0.06	0.87	0.53	1384.32	This study
Upper Reaches of the Yellow River (Ningxia)	34.47	832.00	-	37.95	51.40	74.68	-	0.49	-	0.14	0.19	27.26	1015.17	(Zuo et al., 2016)
India	16.80	72.90	2.80	26.70	-	98.60	29.20	-	-	-	-	2.10	762.50	(Krishna et al., 2009)
Bangladesh	78.00	-	-	35.50	67.00	-	41.50	9.50	-	-	-	31.00	-	(Islam et al., 2015)
Greek	4.85	-	-	6.21	5.83	54.15	3.78	0.92	-	3.42	-	4.05	-	(Karaouzas et al., 2021)
Chinese lakes	11.10	73.70	-	37.90	9.20	18.70	3.20	1.00	-	0.10	-	2.20	-	(Qin and Tao, 2022)
Indonesia	0.88	165.18	1.90	-	6.30	76.05	-	-	-	-	-	3.84	-	(Wulan et al., 2020)
Northeast region of China	-	-	-	15.78-17.30	15.30-17.14	20.21-39.92	-	-	-		-	15.07-15.89	-	(Cui et al., 2022)
Henan section of the Yellow River	0.11-4.36	-	0.02-0.50	0.20-2.11	0.34-6.62	0.30-12.29	-	0.02-1.26	-	-	-	-	-	(Liu et al., 2023)

Table S4 Comparison of HMs concentrations in sediments of the world.

	Cr	Mn	Co	Ni	Cu	Zn	As	Cd	Sb	Hg	Tl	Pb	Sr	Ref
Sediments (mg/kg)														
Ningxia	150.57	776.03	18.04	48.81	32.89	116.25	25.88	0.32	2.06	0.15	0.84	22.98	447.98	This study
Changjiang River,China	-	976.6	-	31.50	28.48	104.12	20.10	0.67	-	0.03	-	27.28	-	(Li et al., 2020)
Lijiang River,	43.62	565.55	8.62	22.95	31.72	129.33	18.30	0.97	-					(Xiao et al.,

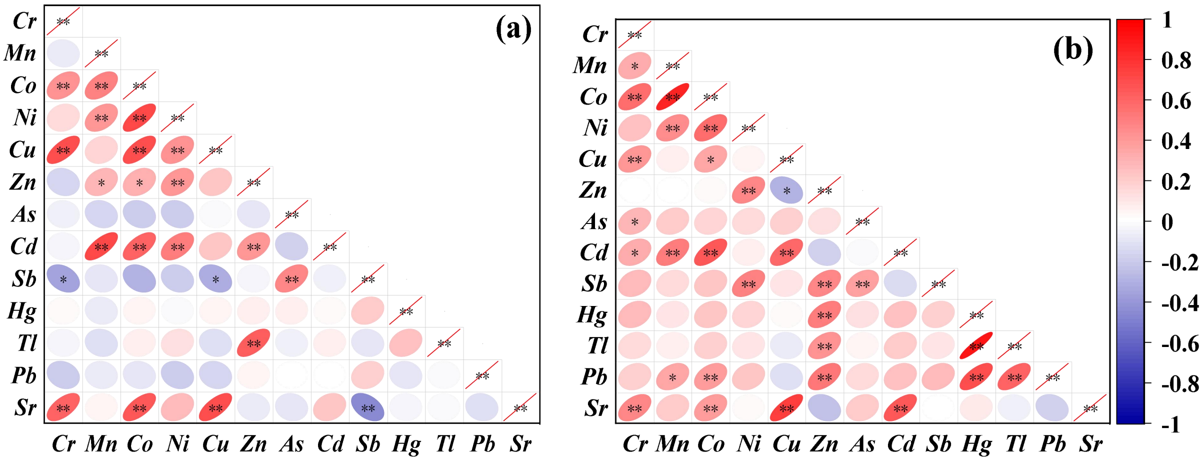
China										0.39	-	42.80	-	2021)
Odiel River, Spain	46.74	-	-	29.71	604.33	2873.64	-	0.86	-	1.65	-	765.16	-	(Bermejo et al., 2003)
Nakuvadra-Rakiraki River, Fiji	108	590	14.5	71.5	42.1	53.7	-	0.99	-	-	-	32	-	(Kumar et al., 2021)
Korotoa River, Bangladesh	109	-	-	95	76	-	25	1.2	-	-	-	58	-	(Islam et al., 2015)
Shur River, Iran	-	-	-	-	9.74	522	-	6.85	-	-	-	162	-	(Karbassi et al., 2008)
Greek	243.4	-	-	144.9	62.3	225.3	31.8	3.24	-	3.84	-	30.6	-	(Karaouzas et al., 2021)
Bangladesh	94.31		16.25	46.32	30.88	-	-	0.13	-	-	0.76	23.28	-	(Khan et al., 2020)

Table S5 Factors matrix of the HMs in the PCA.

Heavy metals	Wet season				dry season			
	PCA component				PCA component			
	1	2	3	4	1	2	3	4
Co	**0.165**	-0.095	-0.109	-0.103	**0.836**	-0.188	0.164	-0.424
Cr	0.124	-0.074	-0.076	**0.462**	**0.669**	-0.220	0.088	0.213
Cd	0.051	**0.268**	-0.349	-0.053	**0.667**	-0.450	-0.396	-0.233
Mn	**0.128**	0.056	0.082	-0.240	**0.666**	-0.117	0.240	-0.584
Hg	0.013	**0.299**	-0.34	0.117	0.576	**0.566**	-0.501	0.186
Zn	0.126	0.086	**0.156**	-0.334	0.301	**0.773**	0.162	0.150
Sr	0.140	-0.140	-0.080	**0.297**	**0.533**	-0.666	-0.153	0.307
Cu	0.155	-0.111	-0.035	**0.226**	**0.483**	-0.642	-0.121	0.380
Pb	0.055	0.191	**0.182**	0.116	0.578	**0.600**	-0.212	-0.157
Tl	0.096	**0.254**	-0.054	-0.019	0.450	**0.594**	-0.555	0.074
Sb	0.023	0.205	**0.370**	0.078	0.409	0.330	**0.583**	0.365
Ni	**0.149**	-0.065	0.018	-0.310	0.535	0.243	**0.564**	-0.196
As	0.079	0.146	**0.322**	0.304	0.362	0.019	0.344	**0.486**
Eigenvalue	5.258	2.531	1.751	1.277	4.949	2.965	1.694	1.364
Variance/%	37.557	18.080	12.506	9.124	35.352	21.18	12.099	9.744
Cumulative/%	37.557	55.637	68.143	77.267	35.352	56.532	68.632	78.375

* The black font indicates the main influence component of each principal component

The correlation coefficient is extensively utilized in scientific research to assess the correlation between variables. High correlations among HMs often indicate a common pollution source for these elements. In the wet season, as illustrated in Fig. S1a, Sr exhibited positive correlations ($P<0.01$) with Cr, Co, and Cu; Cd showed a positive correlation ($P<0.01$) with Mn, Co, and Ni; Cu demonstrated a positive correlation ($P<0.01$) with Co; Ni displayed a positive correlation ($P<0.01$) with Co; Co had a positive correlation ($P<0.01$) with Mn; and Tl-Zn and Sb-As exhibited positive correlations ($P<0.01$). Conversely, during the dry season (Fig. S1b), Pb was positively correlated ($P<0.01$) with Zn, Hg, and Tl; Sb demonstrated positive correlations ($P<0.01$) with Ni and Zn; As only showed weak correlations with other HMs. These results collectively suggest a common pollution source for the studied elements.



* p<=0.05 ** p<=0.01

Fig. S1 Correlation coefficient figure (a) wet season and (b) dry season.

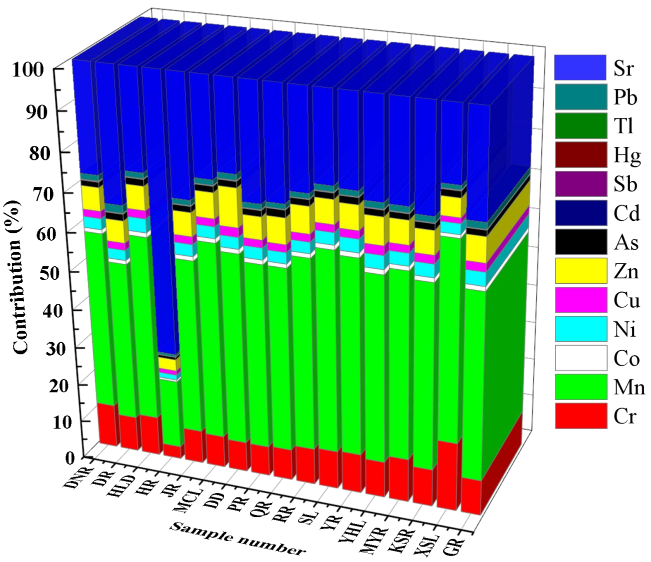
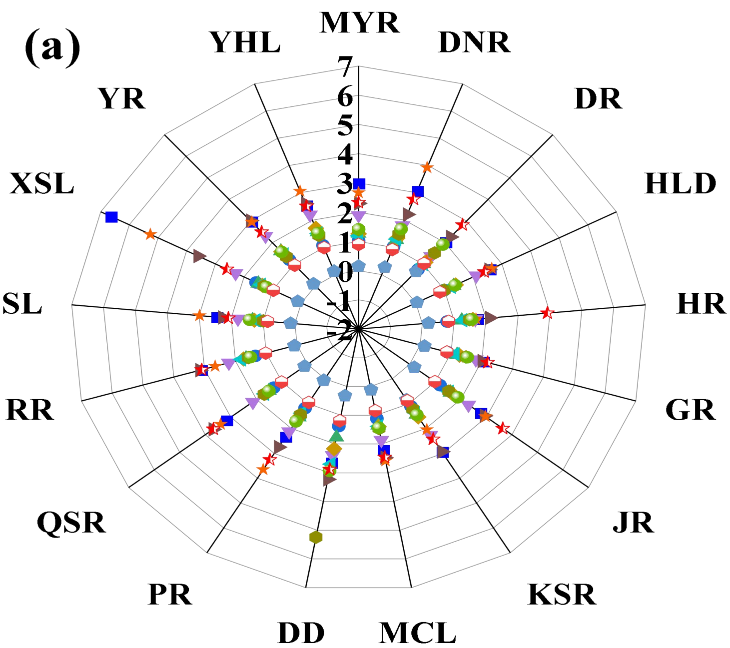


Fig. S2 Contribution profiles (%) of HMs in sediments samples (DNR: Diannong River, DR: Dostu River, HLD: Hongliu Ditch, HR: Hong River, JR: Jing River, MCL: Mingcui Lake, DD: Drain Ditch, PR: Pu River, QSR: Qingshui River, RR: Ru River, SL: Sand Lake, YR: Yu River, YHL: Yuehai Lake, MYR: Mainstream of the Yellow River, KSR: Kushui River, XSL: Xiangshan Lake, GR: Gourd River).



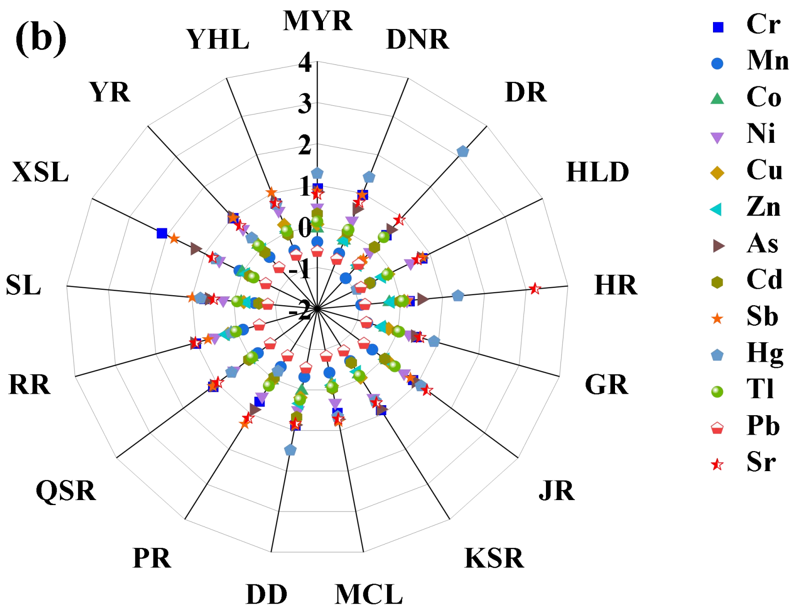


Fig. S3 The Radar of CF and *Igeo* for HMs in Ningxia sediments:(a) CF; (b) *Igeo*.

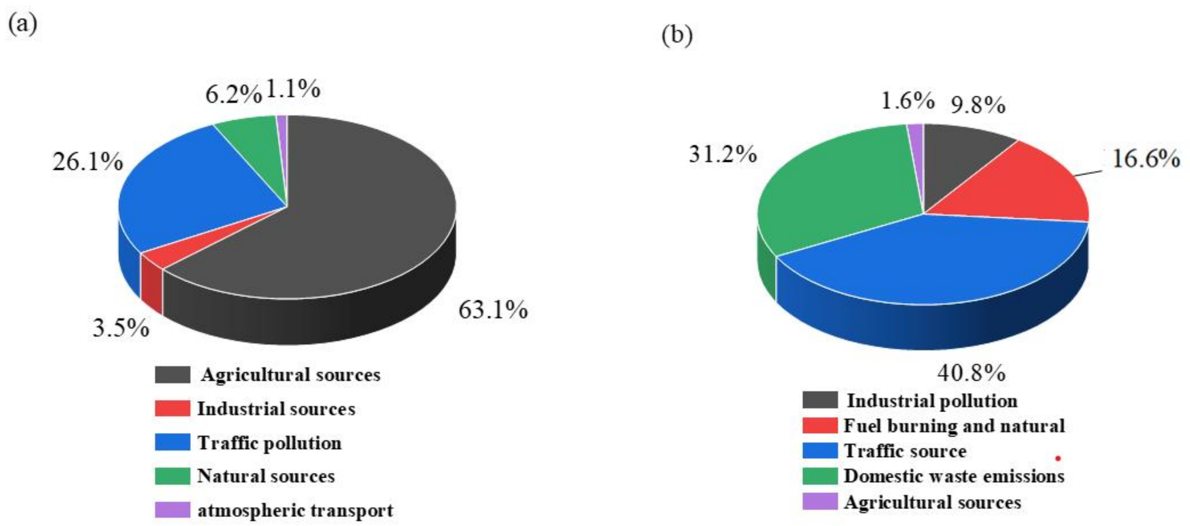


Fig. S4 Contribution of the potential sources of parameters, (a) wet season and (b) dry season.

References

Aydin H, Tepe Y, Ustaoglu F (2023). A holistic approach to the eco-geochemical risk assessment of trace elements in the estuarine sediments of the Southeastern Black Sea. Marine Pollution Bulletin, 189: 9

Bermejo J C S, Beltrán R, Ariza J L G (2003). Spatial variations of heavy metals contamination in sediments from Odiel river (Southwest Spain). Environment International, 29(1): 69-77

Cui Y, Bai L, Li C, He Z, Liu X (2022). Assessment of heavy metal contamination levels and health risks in environmental media in the northeast region. Sustainable Cities and Society, 80: 103796

Huang Y, Chen Q Q, Deng M H, Japenga J, Li T Q, Yang X E, He Z (2018). Heavy metal pollution and health risk assessment of agricultural soils in a typical peri-urban area in southeast China. Journal of Environmental Management, 207: 159-168

Islam M S, Ahmed M K, Raknuzzaman M, Habibullah-Al-Mamun M, Islam M K (2015). Heavy metal pollution in surface water and sediment: A preliminary assessment of an urban river in a developing country. Ecological Indicators, 48: 282-291

Jolliffe I T, Cadima J (2016). Principal component analysis: a review and recent developments. Philosophical Transactions of the Royal Society a-Mathematical Physical and Engineering Sciences, 374(2065): 16

Karaouzas I, Kapetanaki N, Mentzafou A, Kanellopoulos T D, Skoulikidis N (2021). Heavy metal contamination status in Greek surface waters: A review with application and evaluation of pollution indices. Chemosphere, 263: 15

Karbassi A R, Monavari S M, Bidhendi G R N, Nouri J, Nematpour K (2008). Metal pollution assessment of sediment and water in the Shur River. Environmental Monitoring and Assessment, 147(1-3): 107-116

Ke X, Gui S F, Huang H, Zhang H J, Wang C Y, Guo W (2017). Ecological risk assessment and source identification for heavy metals in surface sediment from the Liaohe River protected area, China. Chemosphere, 175: 473-481

Khan M H R, Liu J G, Liu S F, Li J R, Cao L, Rahman A (2020). Anthropogenic effect on heavy metal contents in surface sediments of the Bengal Basin river system, Bangladesh. Environmental Science and Pollution Research, 27(16): 19688-19702

Krishna A K, Satyanarayanan M, Govil P K (2009). Assessment of heavy metal pollution in water using multivariate statistical techniques in an industrial area: a case study from Patancheru, Medak District, Andhra Pradesh, India. J Hazard Mater, 167(1-3): 366-373

Kumar S, Islam A M T, Hasanuzzaman M, Salam R, Khan R, Islam M S (2021). Preliminary assessment of heavy metals in surface water and sediment in Nakuvadra-Rakiraki River, Fiji using indexical and chemometric approaches. Journal of Environmental Management, 298: 14

Li B J, Song J X, Guan M C, Chen Z Y, Tang B, Long Y Q, Mao R C, Zhao J W, Xu W J, Zhang Y T (2024). With spatial distribution, risk evaluation of heavy metals and microplastics to emphasize the composite mechanism in hyporheic sediments of Beiluo River. Journal of Hazardous Materials, 462: 13

Li R, Tang X Q, Guo W J, Lin L, Zhao L Y, Hu Y, Liu M (2020). Spatiotemporal distribution dynamics of heavy metals in water, sediment, and zoobenthos in mainstream sections of the middle and lower Changjiang River. Science of The Total Environment, 714: 14

Liao J B, Cui X Y, Feng H, Yan S (2022). Environmental Background Values and Ecological Risk Assessment of Heavy Metals in Watershed Sediments: A Comparison of Assessment Methods. Water, 14(1): 17

Liu S T, Yu F R, Lang T, Ji Y K, Fu Y, Zhang J N, Ge C (2023). Spatial distribution of heavy metal contaminants: The effects of water-sediment regulation in the Henan section of the Yellow River. Science of the Total Environment, 892: 12

Qin Y, Tao Y (2022). Pollution status of heavy metals and metalloids in Chinese lakes: Distribution, bioaccumulation and risk assessment. Ecotoxicol Environ Saf, 248: 114293

Sun X, Wang H Q, Guo Z G, Lu P L, Song F Z, Liu L, Liu J X, Rose N L, Wang F W (2020). Positive matrix factorization on source apportionment for typical pollutants in different environmental media: a review. Environmental Science-Processes & Impacts, 22(2): 239-255

Wen C, Zhu S J, Li N H, Luo X (2022). Source apportionment and risk assessment of metal pollution in natural biofilms and surface water along the Lancang River, China. Science of The Total Environment, 843: 11

Wulan D R, Marganingrum D, Yoneda M (2020). Distribution, source identification, and assessment of heavy metal pollution in the surface and pore waters of Cipeles River, West Java, Indonesia. Environ Sci Pollut Res Int, 27(31): 39123-39134

Xiao H, Shahab A, Xi B D, Chang Q X, You S H, Li J Y, Sun X J, Huang H W, Li X K (2021). Heavy metal pollution, ecological risk, spatial distribution, and source identification in sediments of the Lijiang River, China. Environmental Pollution, 269: 10

Xie F Y, Yu M C, Yuan Q K, Meng Y, Qie Y K, Shang Z M, Luan F B, Zhang D L (2022). Spatial distribution, pollution assessment, and source identification of heavy metals in the Yellow River. Journal of Hazardous Materials, 436: 12

Zuo H, Ma X L, Yang K, Chen Y Z, Chen J H, Guo Y, Zhao J Y, Wang R M, Fang F, Liu Y (2016). Distribution and Risk Assessment of Metals in Surface Water and Sediment in the Upper Reaches of the Yellow River, China. Soil & Sediment Contamination, 25(8): 917-940