Supporting Information

Influential Topographic Factor Identification of Soil Heavy Metals Using GeoDetector: The Effects of DEM Resolution and Pollution Sources

Yating Wu [1,2], Lingfeng Zhou [3,*], Yaobin Meng [2], Qigen Lin [4] and Yang Fei [5]

[1] Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China
[2] School of National Safety and Emergency Management, Beijing Normal University, Beijing 100875, China
[3] State Key Laboratory of Environmental Criteria and Risk Assessment, Chinese Research Academy of Environmental Sciences, Beijing 100012, China
[4] Institute for Disaster Risk Management, School of Geographical Sciences, Nanjing University of Information Science and Technology, Nanjing 210044, China
[5] Technical Centre for Soil, Agricultural and Rural Ecology and Environment, Ministry of Ecology and Environment, Beijing 100012, China
* Correspondence: zhoulf@mail.bnu.edu.cn

S1. PMF model

Table S1 Summary of EPA PMF 5.0 settings

Parameter	**CZ case**
N species	8
N samples	489
N factors	4 to 6
Treatment of missing data	No missing data
Robust mode	yes
Seed value	random
N bootstraps in BS	100
R^2 for BS	0.8
BS block size	6
DISP dQ_{max}	4, 8, 16, 32
DISP active species	all
N bootstraps; R^2 for BS in BS-DISP	100; 0.8
BS-DISP active species	all
BS-DISP dQ_{max}	0.5, 1, 2, 4

S2. PMF results

Table S2 Summary of PMF and Error estimation diagnostics of the soil HM data

Diagnostic	4 factors	5 factors	6 factors
$Q_{expected}$	1924	1427	930
Q_{true}	3324.3	1805.06	989.1
Q_{robust}	3240.3	1795.57	980.2
$Q_{true}/Q_{expected}$	1.73	1.26	1.06
R^2	Cd (0.38) Hg (0.26) As (0.98) Pb (0.85) Cr (0.87) Cu (0.79) Zn (0.73) Ni (0.90)	Cd (0.87) Hg (1) As (0.99) Pb (0.96) Cr (0.93) Cu (0.79) Zn (0.73) Ni (0.92)	Cd (1) Hg (1) As (0.99) Pb (0.97) Cr (0.98) Cu (0.82) Zn (0.82) Ni (0.93)
DISP *%dQ*	< 0.1%	< 0.1%	>0.1%
DISP swaps	0	0	1st factor 8% 2nd factor 5% 4th factor 3% 5th factor 7%
Factors with BS mapping < 100%	3nd factor 89%	4th factor 99%	all: 100%
BS-DISP % cases with swaps	0%	7%	##

With respect to the ratio $Q_{\text{true}}/Q_{\text{expected}}$, moving from four to five sources, there was a decrease in $Q_{\text{true}}/Q_{\text{expected}}$ from 1.73 to 1.26, and a smaller decrease when moving from five to six sources (1.26 to 1.06), indicating that there may be too many sources being fit, suggesting here that five sources may be the optimal solution.

As for the four-source solution, the coefficient of determination R^2 values for Cd (0.38) and Hg (0.26) being less than 0.70. And the sources with BS mapping less than 90%, indicating some instability at four sources, the four-source solution is not adequate. With five sources, results were generally stable. All sources but 4nd were mapped in 100% of BS runs (4th was mapped in 99% of runs), there were no swaps with DISP, and 93%% of the BS-DISP runs were successful. For the six-source solution, R^2 ranged from 0.82 to 1.00 and the R^2 values for all HMs increased, and there were no swaps in BS. It seems that 6 sources are appropriate, but according to the DISP results,

some swapping occurs in DISP and the dQ_{max} value decreases beyond a reasonable interval (>0.1%) and is greater than the global minimum debuggable range of 0.5%, indicating that there is significant rotational ambiguity in the six-source solution and the solution is not sufficiently robust to be used.

Table S3 Factor Profiles (conc. of HMs) from Base Run of the minimum dQ_{max} (5-factors solution)

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Cd	0.0100	0.0092	0.0402	0.0000	0.1409
Hg	0.1470	0.0000	0.0083	0.0087	0.0000
As	0.1082	0.1576	0.0042	15.4630	3.3263
Pb	1.6969	4.5094	25.9140	1.0638	6.3417
Cr	2.5764	47.8400	4.1986	2.4882	4.6212
Cu	0.8858	13.6480	1.1209	1.3804	9.2593
Zn	3.7568	23.7650	24.1830	0.0229	23.6950
Ni	0.8261	16.7670	0.9494	0.5616	3.5765

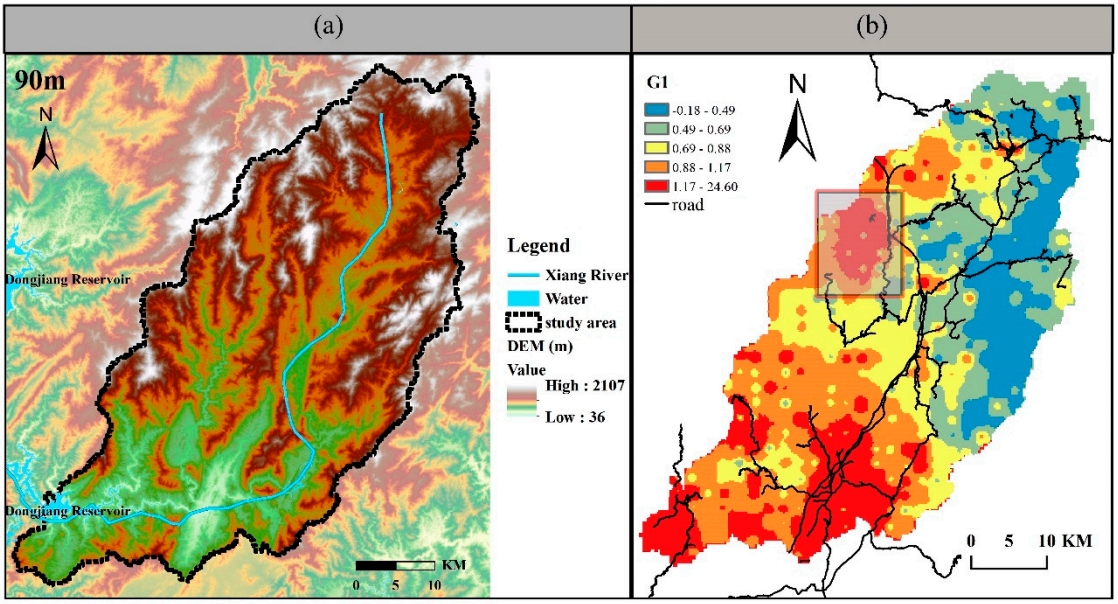


Figure S1 (a) The DEM distribution (90m) and (b) Factor contribution of source 1 (identified as atmospheric emissions and subsequent deposition and transport).

Please note that the high concentration zone on the northwest edge (the area shown with gray shading in Fig. S1(b)) has high DEM values and is not a low-lying area prone to deposition processes. After the field study, we found the area is a town called Shatian. Therefore, this area is considered to be the Hg pollution caused by fossil fuel combustion, such as coal burning, etc.

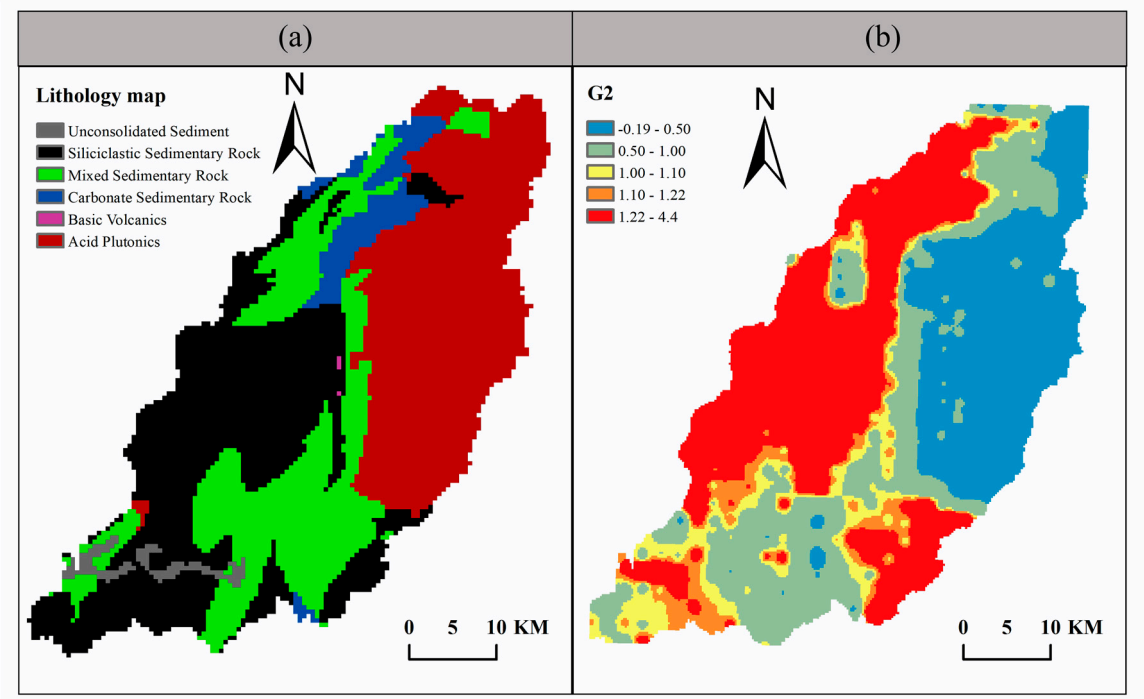


Figure S2 A comparison between (a) the lithology map (Hartmann and Moosdorf, 2012) and (b) factor contribution of source 2 (identified as natural source of parent material).

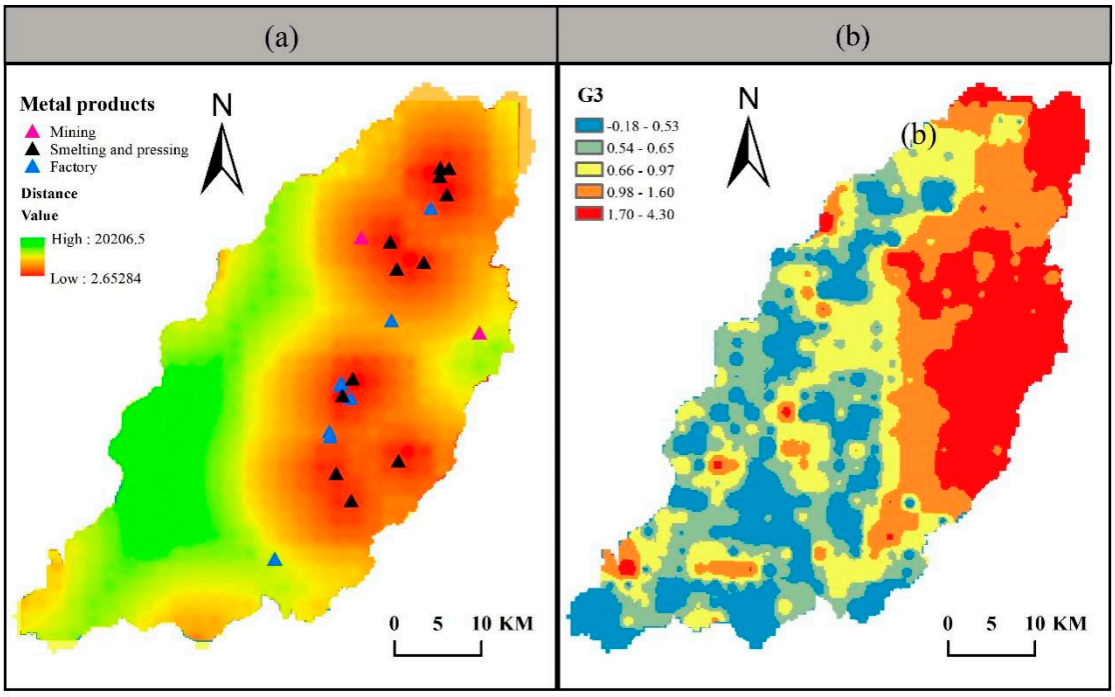


Figure S3 A comparison between (a) distance map of the industrial sites and (b) factor contribution of source 3 (identified as pollution by industrial activities).

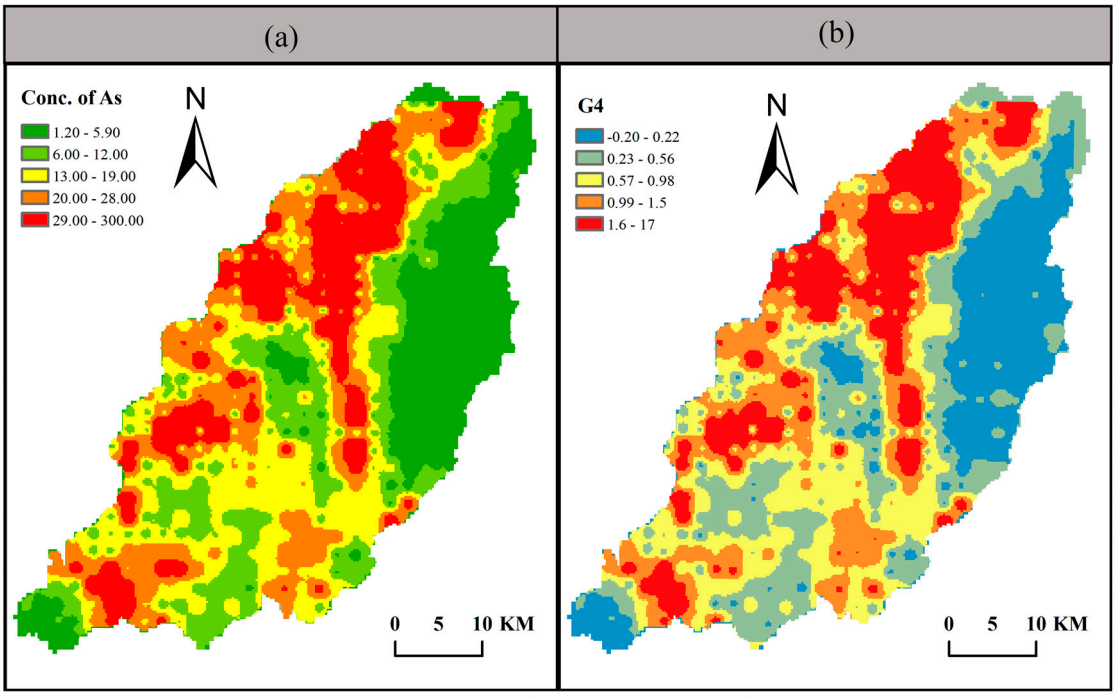


Figure S4 A comparison between (a) concentration of As and (b) factor contribution of source 4 (identified as historical anthropogenic As contamination).

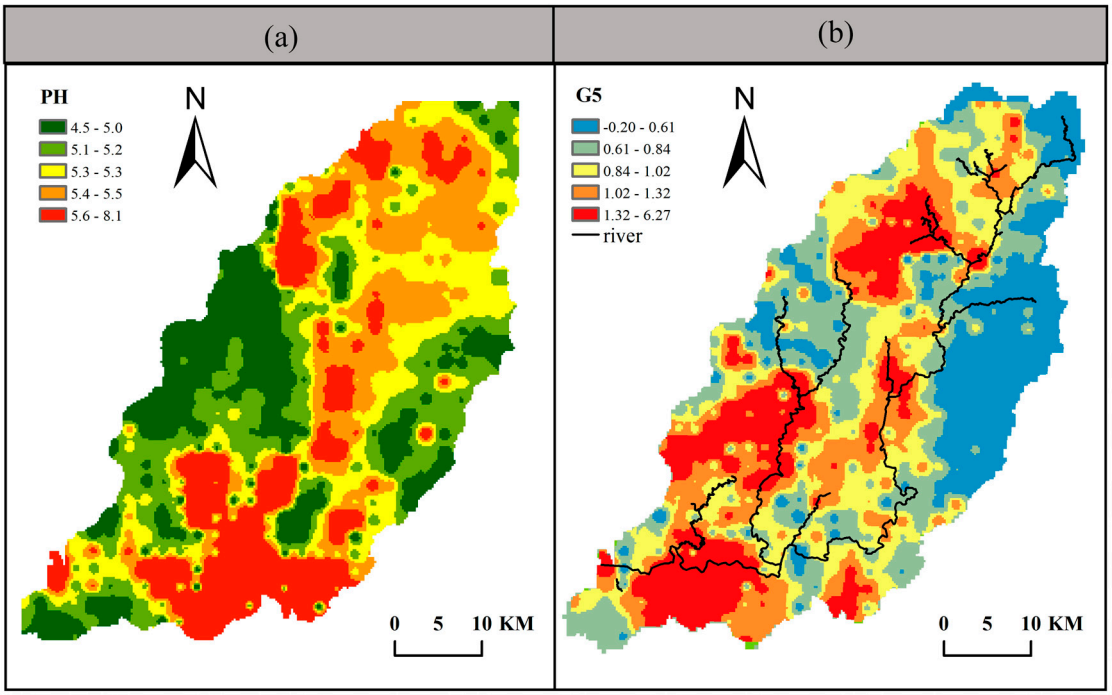
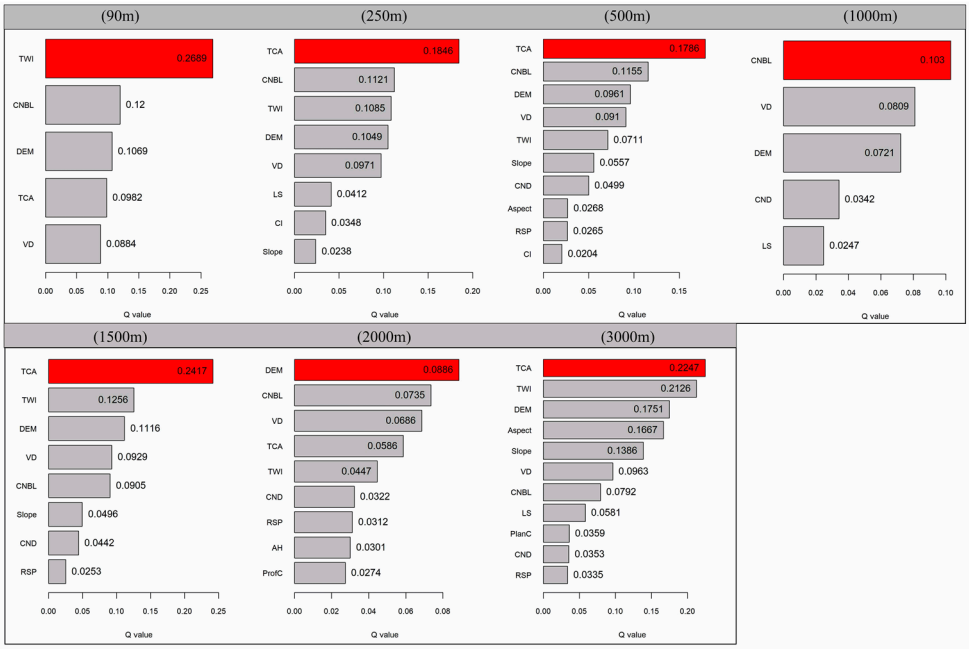


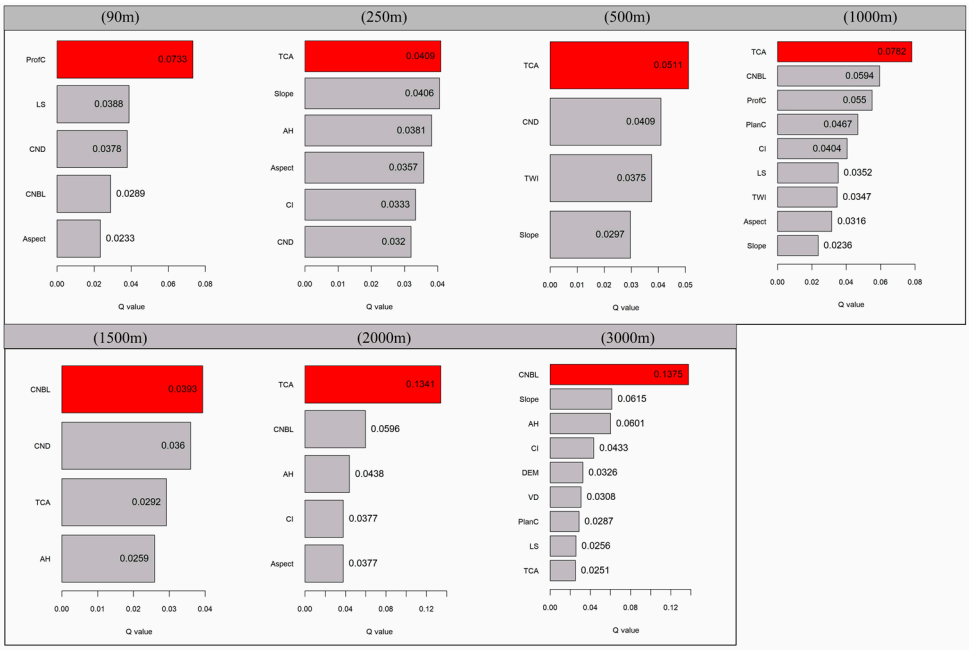
Figure S5 A comparison between (a) The pH distribution and (b) factor contribution of source 5 and River Network (identified as dissolved HM soil hydrological transports).

S3. Factor detector of topographic factors on source-specific heavy metal concentration investigated by the GeoDetector

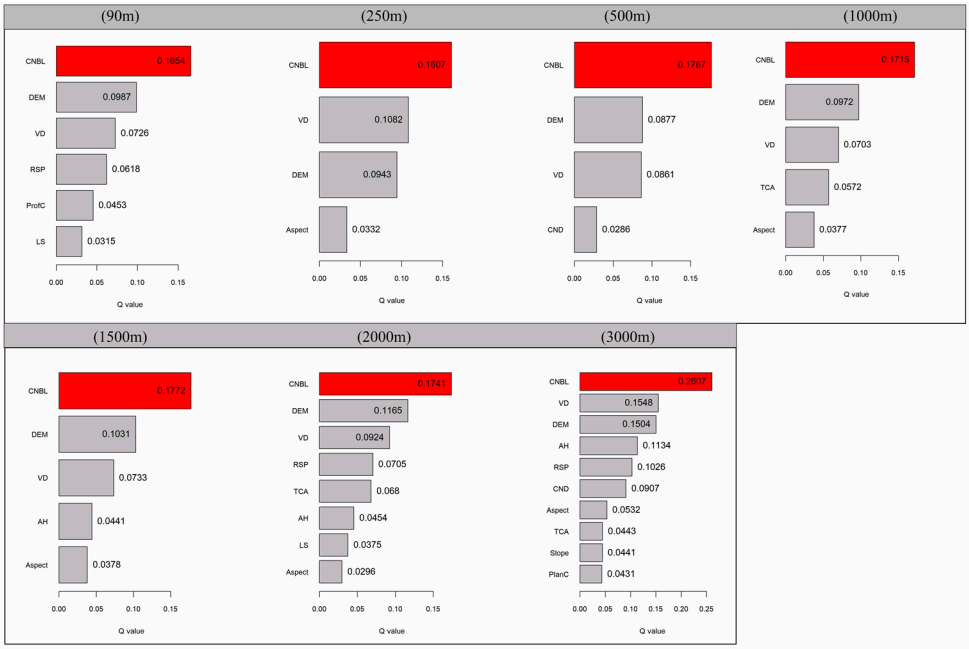
Source 1:



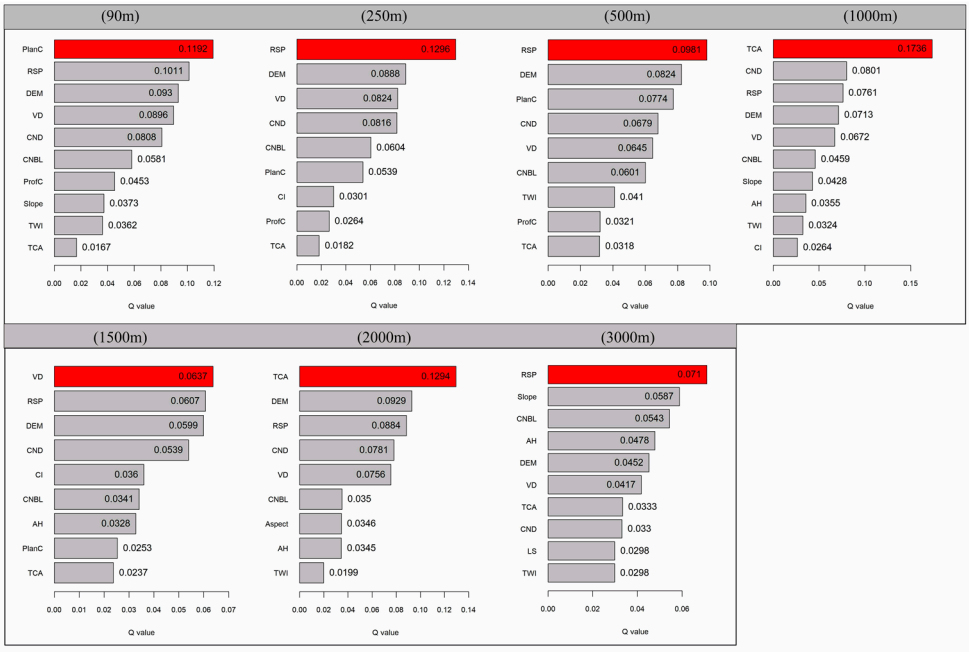
Source 2:



Source 3:



Source 4:



Source 5:

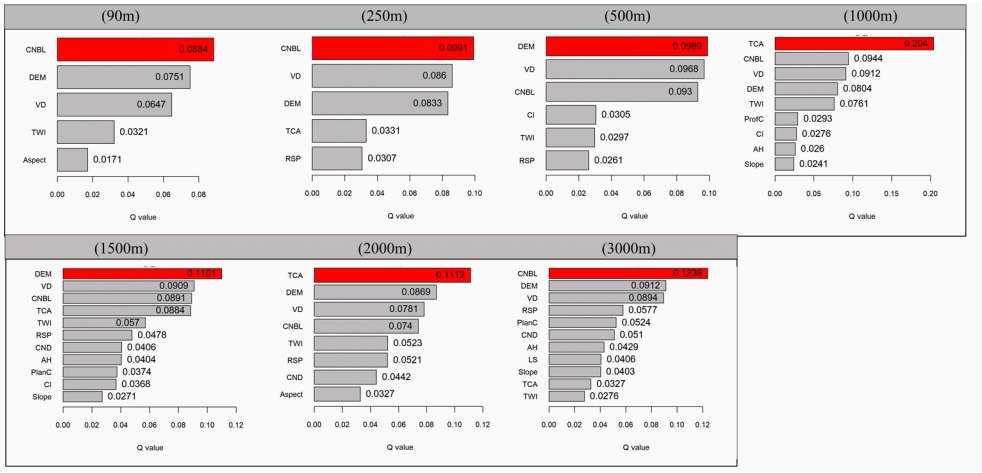


Figure S6 The factor detector results of HMs concentration after PMF source separation: (a) Source 1, atmospheric emissions and subsequent deposition and transport, (b) Source 2, natural source of parent material, (c) Source 3, pollution by industrial activities, (d) Source 4, historical anthropogenic As contamination (e) Source 5, dissolved HM soil hydrological transports.