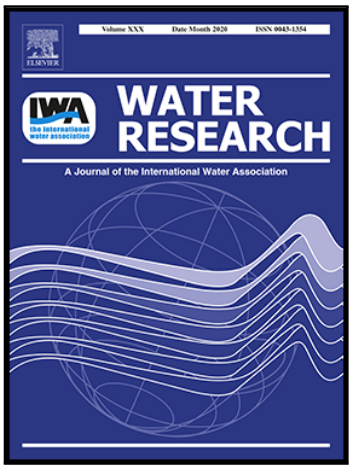
The definitive publisher version is available online at https://doi.org/10.1016/j.watres.2022.119310



Heavy metal habitat: a novel framework for mapping heavy metal contamination over large-scale catchment with a species distribution model

Jianguo Li , Zunyi Xie , Xiaocong Qiu , Qiang Yu , Jianwei Bu , Ziyong Sun , Ruijun Long , Kate J. Brandis , Jie He , Qi Feng , Daniel Ramp

PII: S0043-1354(22)01255-6
DOI: https://doi.org/10.1016/j.watres.2022.119310
Reference: WR 119310

To appear in: *Water Research*

Received date: 4 August 2022
Revised date: 12 October 2022
Accepted date: 28 October 2022

Please cite this article as: Jianguo Li , Zunyi Xie , Xiaocong Qiu , Qiang Yu , Jianwei Bu , Ziyong Sun , Ruijun Long , Kate J. Brandis , Jie He , Qi Feng , Daniel Ramp , Heavy metal habitat: a novel framework for mapping heavy metal contamination over large-scale catchment with a species distribution model, *Water Research* (2022), doi: https://doi.org/10.1016/j.watres.2022.119310

Heavy metal habitat: a novel framework for mapping heavy metal contamination over large-scale catchment with a species distribution model

Jianguo Li [a, c], Zunyi Xie [b, d] *, Xiaocong Qiu [e], Qiang Yu [f], Jianwei Bu [g], Ziyong Sun [g], Ruijun Long [a], Kate J. Brandis [h], Jie He [f], Qi Feng [i], Daniel Ramp [c]

[a] State Key Laboratory of Grassland and Agro-Ecosystems, International Centre for Tibetan Plateau Ecosystem Management, College of Ecology, Lanzhou University, Lanzhou 730000, China

[b] Key Laboratory of Geospatial Technology for the Middle and Lower Yellow River Regions, Ministry of Education, Henan University, Kaifeng 475004, China

[c] Centre for Compassionate Conservation, Faculty of Science, University of Technology Sydney, Ultimo, 2007, NSW, Australia

[d] College of Geography and Environmental Science, Henan University, Kaifeng 475004, China

[e] College of Life Sciences, Ningxia University, Yinchuan 750021, China

[f] State Key Laboratory of Soil Erosion and Dryland Farming on the Loess Plateau, Institute of Soil and Water Conservation, Northwest A&F University, Yangling 712100, China

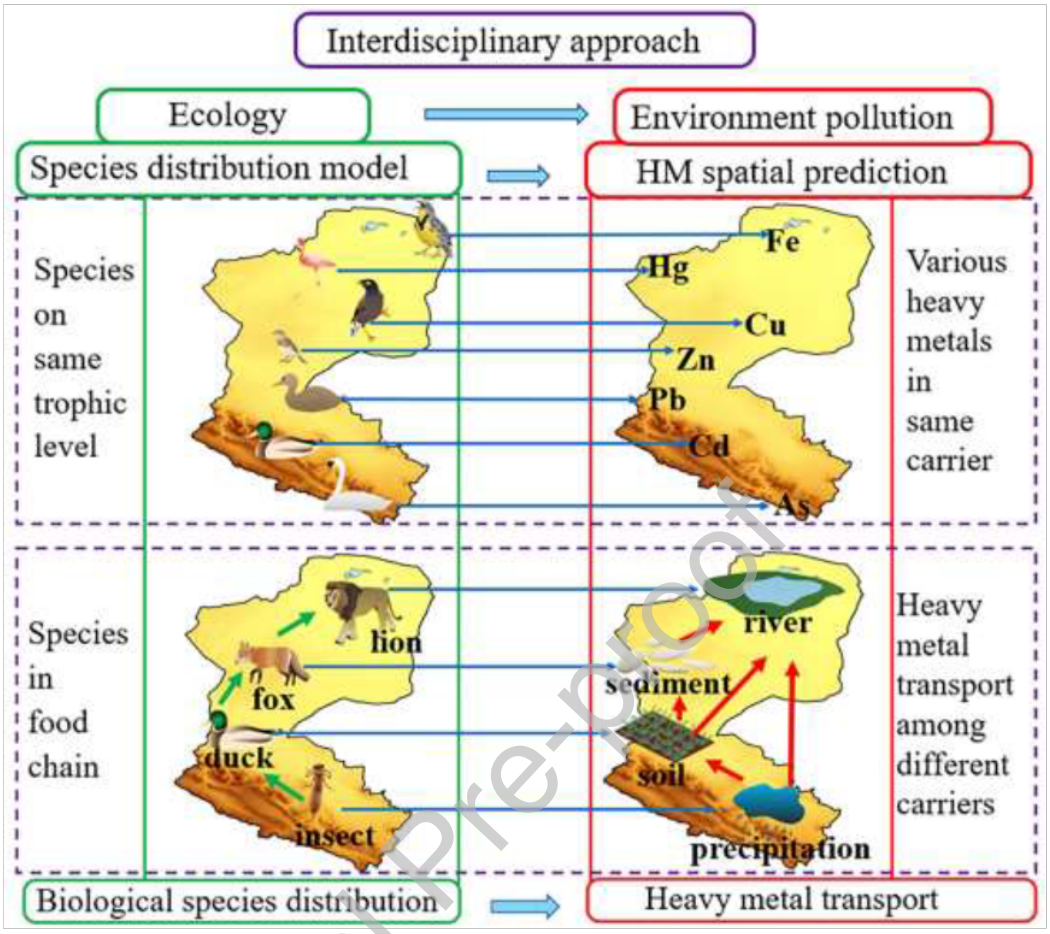
[g] Laboratory of Basin Hydrology and Wetland Eco-restoration, China University of Geosciences, Wuhan, 430074, China

[h] Centre for Ecosystem Science, School of Biological, Earth and Environmental Sciences, University of New South Wales, Kensington, 2052, NSW, Australia

[i] Key Laboratory of Ecohydrology of Inland River Basin Gansu/Hydrology and Water Resources Engineering Research Center, Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences, Lanzhou 730000, China

*Correspondence to: Zunyi Xie. E-mail: zunyixie@henu.edu.cn

Graphical Abstract



Highlights：

- Species distribution model was introduced to study heavy metal(loid) contamination
- Heavy metal(loid)s were ecologicalized as biological species to map their habitats
- The output maps of heavy metals correspond well with their determinant variables
- SDM was proved to offer new insights in the field of water pollution

Abstract

Heavy metal(loid)s (HMs) have been consistently entering the food chain, imposing great harm on environment and public health. However, previous studies on the spatial dynamics and transport mechanism of HMs have been profoundly limited by the field sampling issues, such as the uneven observations of individual carriers and their spatial mismatch, especially over large-scale catchments with complex environment. In this study, a novel methodological framework for mapping HMs at catchment scale was proposed and applied, combining a species distribution model (SDM) with physical environment and human variables. Based on the field observations, we ecologicalized HMs in different carriers as different species. This enabled the proposed framework to model the 'enrichment area' of individual HMs in the geographic space (termed as the HM 'habitat') and identify their 'hotspots' (peak value points) within the catchment. Results showed the output maps of HM habitats from secondary carriers (soil, sediment, and wet deposition) well agreed with the influence of industry contaminants, hydraulic sorting, and precipitation washout process respectively, indicating the potential of SDM in modelling the spatial distributions of the HM. The derived maps of HMs from secondary carriers, along with the human and environmental variables were then input as explanatory variables in SDM to predict the spatial patterns of the final HM accumulation in river water, which was observed to have largely improved the prediction quality. These results confirmed the value of our framework to leverage SDMs from ecology perspective to study HM contamination transport at catchment scale, offering new insights not only to map the spatial HM habitats but also help locate the HM transport chains among different carriers.

Keywords : Heavy metals; species distribution model; catchment; carriers; environmental variables; human variables

Highlights：

- Species distribution model was introduced to study heavy metal(loid) contamination
- Heavy metal(loid)s were ecologicalized as biological species to map their habitats
- The output maps of heavy metals correspond well with their determinant variables
- SDM was proved to offer new insights in the field of water pollution

1 Introduction

Heavy metal(loid)s (HMs), ubiquitous and generally persistent in the environment, tend to accumulate in natural sinks and bio-magnify in the food chain (Ali et al., 2019), imposing great potential threats to human health and environmental sustainability. HMs have been found to actively transport within multiple environmental compartments, such as the catchment (Hassellov and von der Kammer, 2008), urban system (Sarkar et al., 2021), agriculture (Meite et al., 2018), estuary (Fang et al., 2016), mining sites and public transport system etc.(Goth et al., 2019; Senduran et al., 2018; Stojic et al., 2017), spreading from local, regional, to global scales. Thus, understanding of HM transport mechanism and spatial dynamics is essential for conservation and resource management (Hassellov and von der Kammer, 2008). Studies on HM contamination within catchment have been increasingly applied but challenged by the complex environment of catchment system, which consists of multi-dimensional carriers, such as the atmospheric deposition, soil, sediment, river water and biosphere etc (Fig. 1). The river water plays as a pollutant sink for HMs within a catchment, which links various carriers through a series of physicochemical processes, including atmospheric deposition (Nickel et al., 2014), water- rock interactions from riparian and floodplain (Hassellov and von der Kammer, 2008), and soil leaching and erosion (Li et al., 2020) etc (Fig. 1). To separate the carrier types, we defined the river water as the primary carrier, while atmospheric deposition, soil and sediment as secondary carriers.

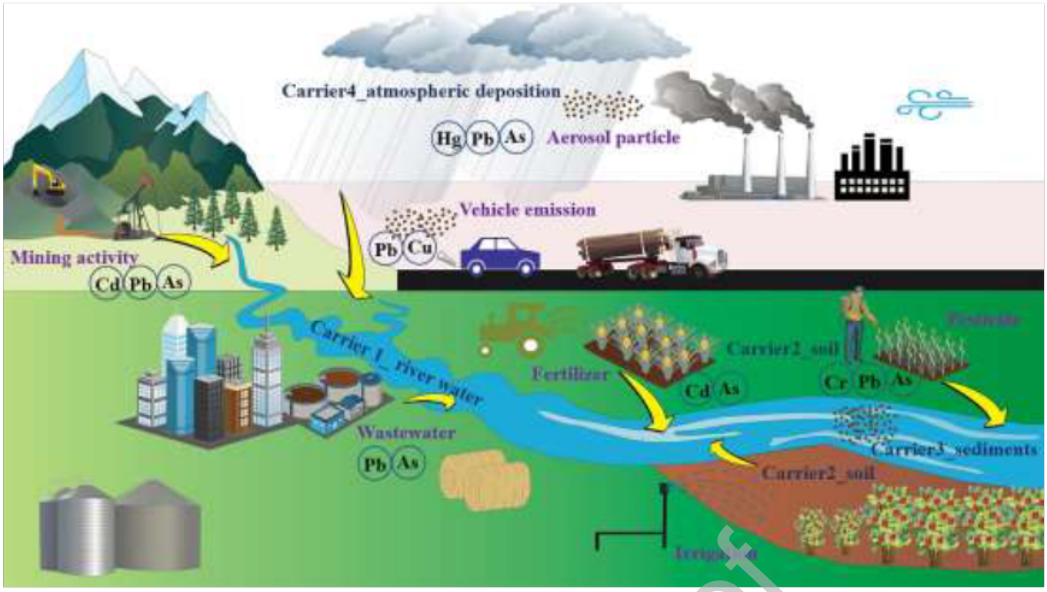


Fig. 1. The HMs transport among primary carrier of river water and secondary carriers of wet deposition, soil and sediment at catchment scale.

HM contamination and transport mechanism in catchment system have been investigated in the previous studies for addressing three main questions in the field. What are the potential explanatory variables that drive HM transport? Where are the HM enrichment areas, and should the recovery efforts be focused? And how do these various carriers interact to spread HMs? Key papers including their strength and limitations for these essential research questions are summarized in Table 1. For example, the Pb isotopes approach and source apportionment methods have been widely used to study the water supply systems and identify HM sources within catchment (Cable and Deng, 2018; Gassama et al., 2021; Lv, 2019). Multivariate receptor models were then introduced to quantify the relative contributions of different pollutant sources (Taiwo et al., 2014). Recent development of nanotechnology may offer potential solution for accurately identifying the relationships between HMs and explanatory variables, which yet is too expensive and not suitable for large-scale study areas (Gajbhiye et al., 2016; Zhou et al., 2020). Additionally, modern statistical methods including simple linear models and advanced machine learning, as well as novel statistical method, such as a flag element ratio approach (Hong et al., 2018) have been increasingly proposed to identify the relationships between HMs and explanatory variables (Hu et al., 2020). As a result, although these previous studies have made significant contributions (details in the literature review of Table 1), there are three important 'gaps' in the current knowledge remain to be solved, which are largely due to the spatial inconsistence and scarcity of the field observations. (i) Insufficient explanatory variables of HMs were often selected to correspond with a specific carrier studied in previous research. Therefore, comprehensive explanatory

variables from separated environment and human factors are needed for a better understanding of HM contamination and transport mechanism. (ii) Current mapping methods of HMs are dependent on the spatial interpolation like Kriging, and the results are greatly sensitive to the intensity and spatial distribution of field data sampling. (iii) Only single HM carrier was investigated in most of the previous studies because of the spatial heterogeneous sampling sites among different carriers, profoundly hindering our understanding of the interactions from a multi-dimension-carrier perspective, which play an important role in transporting the HMs.

Therefore, a comprehensive framework of novel methods and datasets is needed to fill these knowledge gaps for better mapping and assessing HM contamination in large-scale catchment. Species distribution models (SDM) have been widely applied in research areas such as conservation biology, ecology, and evolution. SDM is a model of a general format, which can predict the suitability of the potential habitats of different species with presence location sample data and environmental features (e.g., climate, land use, soil etc.). In other words, SDMs of various algorithms can map the past, current and future spatial distributions of different species, which however were all designed according to the same core rule - the characteristics of species. Thus, if significant research similarities between species and certain research targets (e.g, HMs) exist, there would be potential in putting the targets as entries to SDMs instead of species to map the spatial distributions of targets. This rationality has been evidenced by the success of several previous studies, such as geographical planning and design (Clemente et al., 2019), virus spread monitoring (Harrigan et al., 2014), and spatial distribution of human emotion (Li et al., 2021). However, SDMs and HM mapping have only been developed in parallel, and this study showed that SDM has potential in mapping habitats of HMs due to the important research similarities between species and HMs:

(i) Fuzziness: the spatial distributions of these species and HMs can be represented by specific boundaries, such as administrative regions, geomorphic units etc (Fig. 4).

(ii) Complexity: both the species and HMs are being influenced by multi-dimensional factors, such as the environment and human variables (Fig. 4).

(iii) Spatial coexistence: multiple biological species on the same trophic level can share the same habitat; similarly, multiple HMs (e.g., Pb, Hg, As, Cd etc.) in each carrier discussed in this study often co-exist in the catchment (Fig. 5).

(iv) Dynamics: interactions always occur among multiple biological species on food chains, likewise, HM transport and accumulate among various carriers (Fig. 5).

In this study, we developed a novel framework for mapping HM contamination at large catchment scale combining a typical species distribution model method (Guisan and Zimmermann, 2000; Hijmans and Elith, 2013),with environmental and human variables. With different HMs (e.g., Pb, Hg, As, Cd, etc.) being considered as different 'species', we were able to model the 'habitat' of each HM enrichment area in the geographic space (termed as the HM 'habitat') along with HM 'hotspots' (HM peak value points). In particular, we aimed to: (i) separate impacts from natural environment and human activities on secondary carriers (e.g., soil, sediment, and atmospheric deposition); (ii) investigate the spatial 'habitats' and 'hotspots' of HMs where mitigation efforts should be given within various carriers; (iii) assess spatiotemporal HM transmission patterns of river water (the primary carrier) across the entire catchment, with modelled spatially consistent data. This study proposed to leverage SDM models from ecology to map HM contamination, which can effectively build the environment-human-HM relationships to fill the current important knowledge gaps over large-scale catchment. This will offer new future perspectives and directions to the field of water pollution.

Table 1. Summary of relevant studies conducted previously for mapping HM contamination and transport mechanism at catchment scale (the current paper is added for completeness). Three components in the 'Key results' column are identified by the code:(1) Interactions among multiple carriers;(2) Explanatory variables used;(3) Spatial maps of HM distributions. NA represents 'not applicable' in the relevant research.

Study	Field sampling design/ requirements	Analytical methods	Key results
Lv (2019)	Prior designs with sampling density of 2km × 2km at a 1138 km^2 County	Receptor models, APCS/MLR and PMF, Kriging interpolation	(1) Only one carrier of soil was analysed (2) NA (3) Kriging maps with the factors resulting from receptor models
Nickel et al. (2014)	Sampling network throughout Norway is required	GLM, geo-statistics multivariate regression, Kriging interpolation	(1) NA (2) A set of potential predictors were classified based on literature review (3) Regression and residual maps with grid size of 5 km × 5 km
Hu et al. (2020)	Proximity	Three machine learning methods GBM, RF, GLM	(1) Soil and crop ecosystems (2) Soil properties and land use types. (3) NA
(Gajbhiye et al., 2016)	Proximity, nearby roadside shoulders	SEM	(1) soil, road dust, plant leaf, foliar dust

	along the paved roads		(2) only human activity (3) NA
(Wijesiri et al., 2019)	Proximity, three water samples and three sediment samples were collected at each location	An innovative conceptual model	(1) water and sediments (2) human activity around an urban river (3) NA
Liang et al. (2017b)	Proximity	Curvilinear regression analysis	(1) Atmospheric deposition and soil (2) Only land use types were analysed (3) NA
(Zhou et al., 2020)	Proximity, collected in ten villages around the artisanal zinc smelting area	SEM	(1) soil, smelting waste particles (sediment) (2) industrial activities, smelting activity (3) NA
(Hong et al., 2018)	Proximity, primarily alongside the traffic lanes	An innovative flag element ratio approach	(1) stormwater, dust, soil (2) traffic activities, such as gasoline emission and vehicle exhaust attached to the urban road (3) NA
Zang et al. (2021)	Sampling station and synchronous equipment is necessary	Receptor models, PCA, PMF	(1) Only one carrier of precipitation (2) Natural environmental factors and human influence (3) NA
Liang et al. (2017a)	High sampling density of seven sites/ km^2	PMF, Kriging interpolation	(1) Only one carrier of soil (2) Only land use types (3) Kriging interpolation maps
Li et al. (2017b)	Grid distribution point method	Correlation analysis, PCA, Kriging interpolation	(1) NA (2) Only the land use types close to piles of mine tailings were analysed (3) Kriging interpolation maps
This study	Covering entire catchment; each dimension of carrier has their own sampling density	SDMs; RDA; GLM; GAM; model selection and model averaging	(1) Multiple carriers of river water, atmospheric deposition, soil, and sediment were analysed (2) Systematic explanatory variables to represent environmental and human influential factors, auxiliary variables of secondary carriers (3) Spatial maps of HM 'habitat', 'hotspot' and river transmission across the entire catchment modelled by a SDM

Proximity is a sampling strategy that chose a short distance apart between different dimensions of carriers at the same location. Abbreviations of the analytical methods used in Table 1 include absolute principal component score/multiple linear regression (APCS/MLR), positive matrix factorization (PMF), generalised linear models (GLM), gradient boosted machine (GBM), random forest (RF), scanning electron microscopy (SEM), principal component analysis (PCA), species distribution models (SDMs), redundancy analysis (RDA) and generalised additive models (GAM).

2 Study area

We applied our proposed methodological framework to the Heihe River Basin (HRB), which is a large-scale catchment that has always been considered as a natural laboratory in many previous research (Cheng et al., 2014). The HRB originates from the Qilian Mountains along the northeastern margin of the 'Third Pole of the World' - Qinghai-Tibet Plateau (Fig. 2A). The main river flows 821 km, covering an area of approximately 14.3×10^4 km^2, with a mean annual runoff of 1.588×10^9 m^3 (Ge et al., 2013). It is separated into upper, middle, and lower stream areas by the Yingluoxia and Zhengyixia hydrological stations (green triangles in Fig.2A).

There are distinct environmental backgrounds within the HRB. The elevation ranges from 5,000 meters above the sea level (m asl) in the headwater area to nearly 900 m asl in the terminal lakes (Fig. 2B). Topographically, the basin consists of the Qilian Mountains in the south, the Hexi Corridor Plain in the middle segment and the foothills and Alxa plain in lower reaches where partly borders the Badain Jaran and Tengger deserts. The cold mountain zone climate is dominant over the upper reaches area, with the mean temperature varying from –5°C (December) to 4°C (July). The average annual total precipitation of the upper reach is over 250 mm, which ranges from 600 to 700 mm at higher altitudes. Most of rain there occurs during summer between June and September(Li et al., 2017a). The middle and lower reaches areas are typical arid temperate zone and extremely arid temperate zone, with the mean annual temperature being around 6 and 8 °C, and the annual total precipitation less than 200 mm and 50 mm, respectively (Yang et al., 2019). The main geomorphological compartments of HRB include the glacier, alpine ice-snow and permafrost, water conservation forest, piedmont oasis, and desert oasis (Li et al., 2020). The upper reaches of the HRB owns 88% of the total annual water resources in the basin, but the mainstream in the lower reaches becomes wide and sluggish, or even dried up in some sections due to the extremely arid climate and over exploited water resource. This strongly hinders the comprehensive water sampling in the field. While the natural setting of the HRB provides a perfect research platform for the HM transport among various carriers, the complexity of the basin also makes field sampling and observation rather challenging. Thus, it is hard to use the conventional design and approach to conduct a spatially consistent sampling of carriers within this catchment.

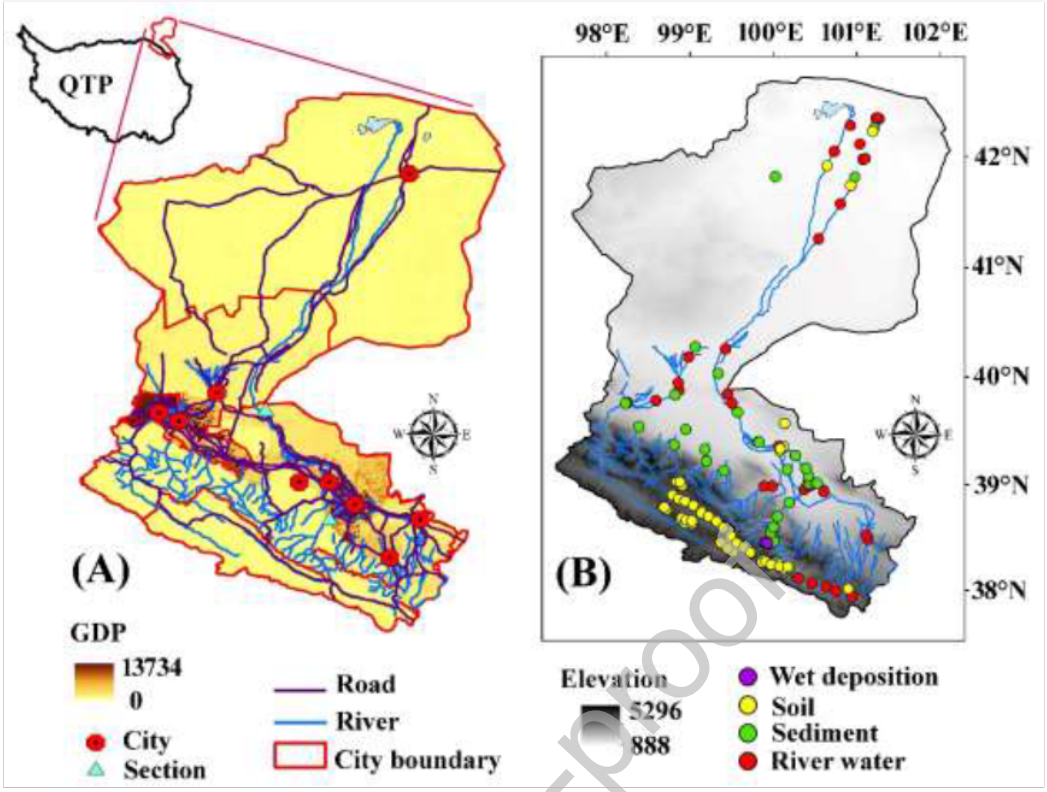


Fig. 2. Study site showing: (A) the Heihe River Basin (HRB) originating in the northeastern margin of the Qinghai-Tibet Plateau, as well as the main cities, hydrological sections, road, and GDP (10^4 RMB km^{-2}) distributions; and (B) the sampling sites of different carriers within the catchment.

The regional social-economic characteristics are also distinguishable from upper to lower stream areas. The upstream area is dominated by animal husbandry, coupled with small-scale mining. These mines had been placed into operation since the 1980s for the special metallogenic conditions and resources (Li et al., 2020). Although the government had adopted the conservation policy of Qilian Mountain National Park in June 2017 to ban mining activities (Yan and Ding, 2020), the subsequent effects of mining can last for decades. The middle reaches, accounting for 95% of the population and 88.7% of the economy contributions (Fig. 2A), is known for its development of agriculture, nonferrous metals, steel, and petrochemical resources (Zhang et al., 2020b). The downstream district is a pastoral area, but the tourist population had increased from 0.03 million in 2000 to 1.1 million in 2015, and the tourism income accounted for two-thirds of the tertiary industry in 2015 (Lu et al., 2021). Therefore, the diverse social-economic conditions within the HRB provide an ideal site for extracting typical human activity patterns.

3 Framework design

There were 4 main steps in this research (Fig. 3): (1) Create spatial distribution layers of secondary carriers via SDM. In order to separate human and environmental explanatory variables, soil sampling sites were collected intentionally close to the human activity locations, while sediment and wet deposition sites were far away from human activities (see 3.2.1). The importance analysis of explanatory variables based on RDA and model selection will be then performed, and important variables will be used for prediction. (2) Select systematic explanatory variables for river water, including human activity variables and environmental variables, as well as the HM layers of secondary carriers that were created in Step 1. (3) Analyze the relationships between HMs in river water and explanatory variables by generalized linear models (glm) and generalized additive models (gam). (4) Predict the spatial distributions of HMs in river by SDM, considering all the potential variables.

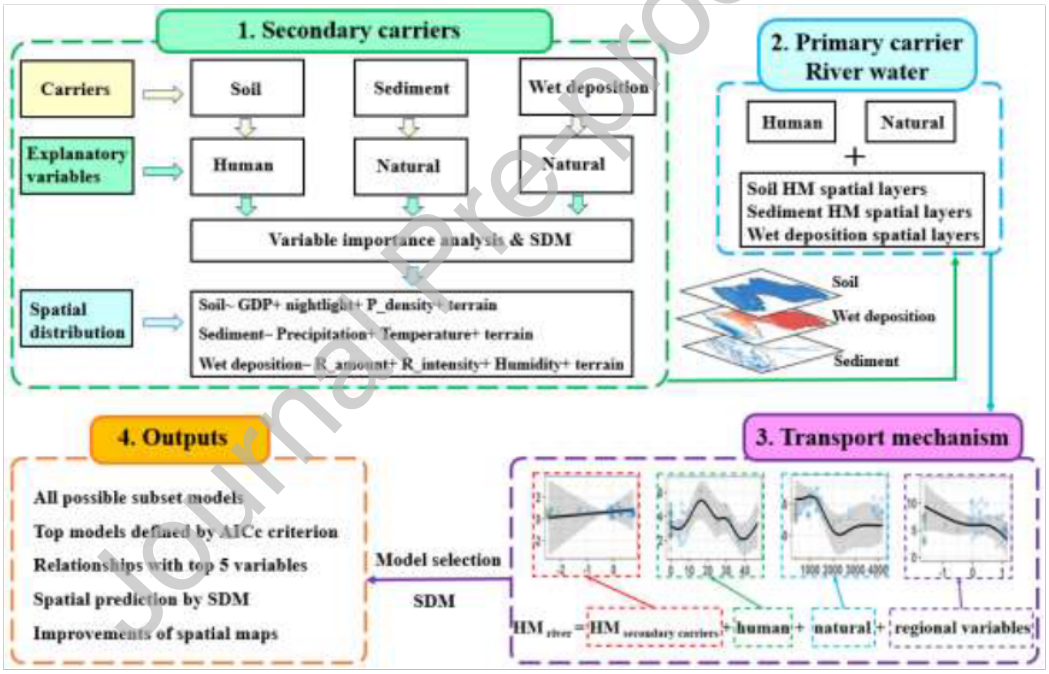


Fig. 3. Flowchart of the approach framework with specific information for each step.

The model building process of HM mapping in this study was mainly based on a typical SDM framework (Guisan and Zimmermann, 2000; Hijmans and Elith, 2013), including (i) model conceptualization, (ii) data preparation; (iii) model fitting, (iv) performance assessment and outputs.

3.1 Model conceptualization

In ecology, the SDMs use species presence-only data and explanatory variables to assess species' niches and predict their potential habitats (Fig. 4). In other research fields, the various entities are regarded as different 'species', and the areas showing the positive relations with these research objectives were defined as 'habitat', such as the 'emotional habitat' (Li et al., 2021). In this study, we ecologicalized HMs in these carriers as different species, which enabled the proposed framework to model the 'enrichment area' of individual HMs in the geographic space (termed as the HM 'habitat') and identify their 'hotspots' (peak value points) in the catchment. Each HM, like As, Cd, Pb, Zn in the geographic space was regarded as the 'species', and the spaces where show high concentrations were regarded as the 'Heavy metal habitat' (Fig. 4). Additionally, the HM transport among different carriers were converted into the food chain relationships in the SDM (Fig. 5).

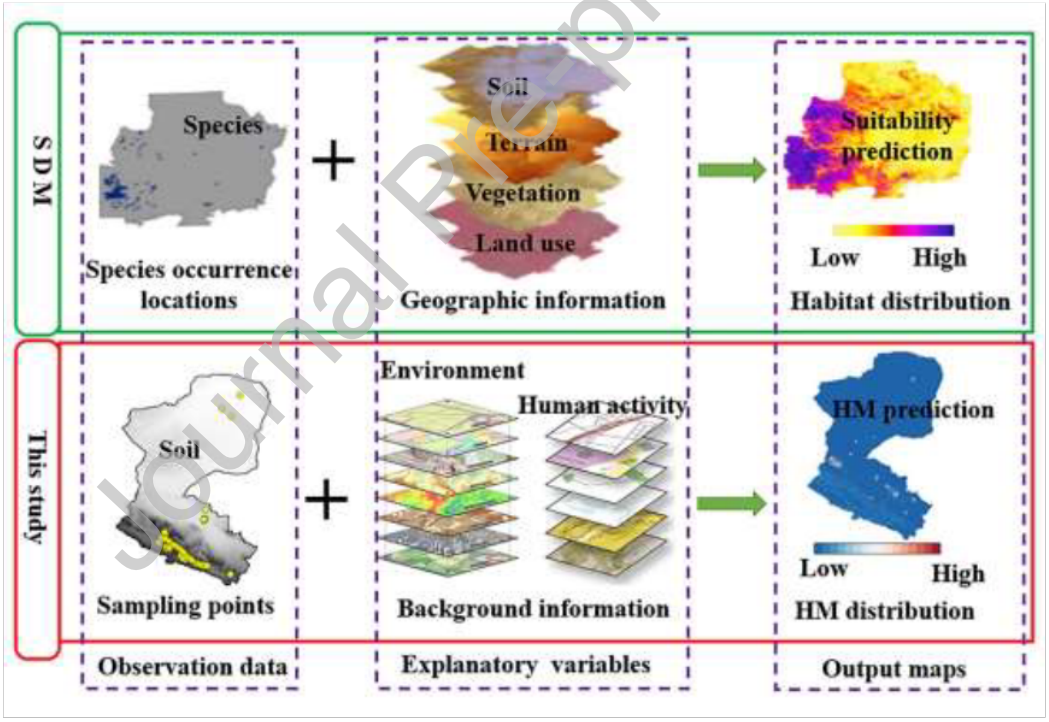


Fig. 4. Comparison between the framework of the species distribution model (SDM)[1] and this study.

[1] Figures of SDM were adapted from https://www.natureserve.org/products/species-distribution-modeling

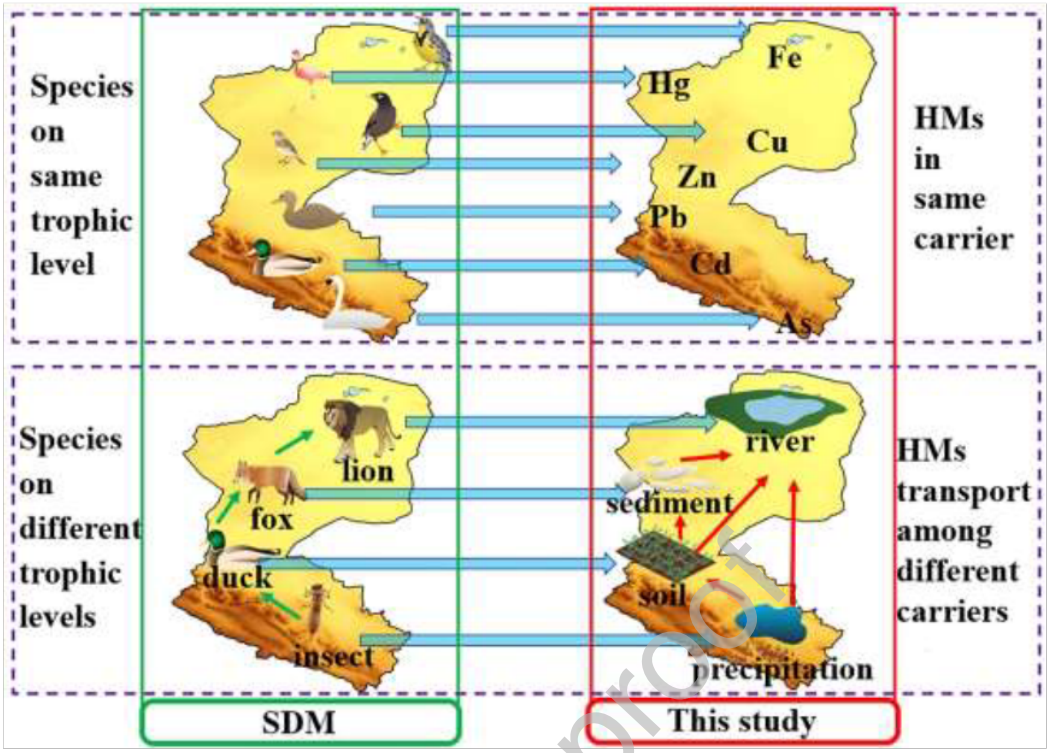


Fig. 5. Comparison of species on different trophic levels in the species distribution model (SDM) and HM transport in multi-dimensional carriers in this study

3.2 Data preparation

3.2.1 Field sampling of various carriers

HMs were sampled respectively in the primary carrier of river water, as well as in the secondary carriers such as soil, modern fluvial sediment, and atmospheric deposition within the HRB (Fig. 2B). Although SDM does not require sampling density and spatial consistency for field observations, we tried to cover the entire catchment when collecting river water. A set of river water samples were collected along the mainstream and tributaries of the HRB both in winter (December 2014 & January 2015) and summer (July 2014) periods at the same locations. All river water samples were collected in the surface stream into pre-cleaned polythene bottles. To further separate the influence from natural (e.g., sediment, atmospheric deposition) and anthropogenic (e.g., soil) variables, the HMs in secondary carriers were sampled with intentions around the same hydrological year. The atmospheric depositions were collected at the forest sites, with an elevation of 2850 and 3050 m asl respectively in the upstream of the catchment. The samples were collected after each precipitation event and then stored in polyethylene bottles that had been thoroughly cleaned with deionized water. The

ultrapure HNO_3 were added to acidify both river water and wet deposition samples, and all the liquid samples were stored at 4 °C until being shipped to the laboratory. In terms of the solid samples, the sites for collecting the soils were in typical land use types, for example the mining soils in the upstream, agricultural and industry soils in the middle stream, and the livestock field soils in the downstream areas (Fig. 2B). To obtain representative samples, approximately 1 kg of surface layer fresh soil (0–20 cm in depth) was collected using a precleaned plastic dustpan and brush. In contrast, the modern fluvial sediment samples were collected from the subsurface below 10 cm depth at riverbed and floodplain by a pre-cleaned spade. All the sediment sampling sites were far from urban areas and farmland to minimize the influence of human activities (Fig. 2B). Both the soil and sediment samples were stored in plastic bags and then shipped to the laboratory for further HM concentration analysis.

Prior to HM concentration analysis, the river water and atmospheric deposition samples were digested with 1% HNO_3. Both soil and sediment samples were air dried at ~20 °C; sorted through a 2 mm plastic sieve to remove any clasts; homogenized by using an agate mortar; and finally passed through a 200-mesh sieve (Zhang et al., 2020a). Took 0.1 g of each solid sample, and mixed 2 mL of concentrated HNO_3 and 1 mL of $HClO_4$ in a polypropylene vessel. The solution was heated for nearly 4 h, and then the residue was re-dissolved in a plastic vessel with 2 mL of 4 mol L^{-1} HCl and diluted to 10 mL with deionized water. The blanks and standard reference materials (GBW07408 (GSS-8)) were then processed. The concentrations of HMs of river water, wet deposition and soil were measured by the inductively coupled plasma mass spectrometer (ICP-MS, 7500a, Agilent, Santa Clara, CA, USA), sediments by (ICP-MS, X-7, Thermo-elemental, USA). For most of the HMs, the corresponding relative standard deviation was below 5%. Finally, a valid field dataset was ready for subsequent analyses, including 94 samples from river water (47 in summer and 47 in winter), 8 wet deposition samples, 65 soil samples and 37 sediment samples.

3.2.2 Groups of explanatory variables

We proposed two types of adequate and systematic explanatory variables for the primary carrier- river water. One type was the main explanatory variables, including the environmental and human variables, the other was the corresponding HM compositions in the secondary carriers. The sets of explanatory variables were only selected in the models if: (i) they were able to stand for typical natural and economic conditions in this catchment, such as the social- economy status; (ii) they were closely related to HMs transport process, for example the frequent and strong erosions; and (iii) they had been proven by previous studies to have a profound impact on the HMs, such as the terrains around the mining sites. Gross

domestic product (GDP) (Luo et al., 2021), night light (Levin et al., 2020), population density (Trujillo-González et al., 2016), land use and transport (Hong et al., 2018) datasets were selected as proxies of human activities (e.g., urbanization and/or industrialization) that could directly or indirectly affect the HM abundance. All these datasets were extracted from the social, statistical, geographical information system, and remote sensing products as below:

The explanatory variables were presented in Table 2. Land use, terrain, and meteorological data were obtained from the resources and environment data center of the Chinese Academy of Sciences (RESDC) (http://www.resdc.cn). Socio-economic data, also downloaded from the RESDC, include spatial distribution data of a 1 km grid GDP, nightlight, population density, and transport data (railways, expressways, national highways, and provincial roads). In total, five environmental and eleven human activity variables were extracted.

Table 2. Potential natural and human explanatory variables for HMs.

Variables	Comments	Resolution	Unit
natural variables			
NDVI	Normalized difference vegetation index	1 km	-
DEM	elevation	30 m	m asl
erosion	soil erosion classification	1 km	-
temperature		500 m	°C
precipitation		500 m	10^{-1}mm
human variables			
GDP	gross domestic product	1 km	10^4 RMB km^{-2}
P_ density	population density	1 km	inhabitants km^{-2}
d_ road	distance to road	NA	m
night light	Range from 0 to 63	1 km	-
land_1	forests	30m	m^2
land_2	grassland	30m	m^2
land_3	water	30m	m^2
land_4	unexploited land	30m	m^2
land_5.1	agriculture and plantations	30m	m^2
land_5.2	urban and rural residential	30m	m^2
land_5.3	industry, mining	30m	m^2

The corresponding HM concentrations in secondary carriers were considered as the auxiliary variables to predict the HM in the primary carrier. Each carrier has its own sampling density during the filed collection process. We first calculated the respective determinant variables for

each secondary carrier according to the SDM model, and then predicted their spatial distributions based on these determinant variables by SDM. Once the spatial distribution rasters of each carrier were created within the catchment, we extracted all raster values by the river water locations. By doing so, all the HM values of secondary carriers can be interpolated to river water sampling sites.

3.3 Model fitting

The choice of adequate modelling algorithms and desired model complexity should be guided by the research objective and by hypotheses regarding the specific entities-variables relationship (Elith and Franklin, 2013). Based on the dataset characteristics, we used redundancy analysis (RDA), glm and gam as parts of model fitting process.

3.3.1 Redundancy analysis

RDA was introduced here to deal with multicollinearity and determine which variables should be included in the model. RDA is a powerful methodology to produce an ordination that regresses the impact of a matrix of multiple explanatory variables. The objects with similar variable values were ordinated closer together, while different values were projected apart. Before analysis, the datasets were normalized to reduce the influence of different units. The qualitative variables were recorded as dummy variables (Legendre and Anderson, 1999), and they were different HMs in this study. Then a permutation test with 1000 was conducted to examine the null hypothesis that no linear relationship exists between the response and explanatory variables. We then examined the significance of constraining variables by the permutation test with 100000, and the determinant variables with higher significant value were remained when collinearity. The proportion of variance explained by each RDA axis was presented by an eigenvalue. Considering the extreme complexity of natural environment (Badry et al., 2019), the sum of the first two axes' eigenvalues above 10% was acceptable in this study.

3.3.2 Generalized linear models and generalized additive models

Generalized linear models (glm) allowing the response variables to have error distribution models, has been widely used as an appropriate theoretic approach. Glm was applied here to examine the ability of a reasonable combination of explanatory variables to explain variation of the HMs. After that it was applied to predict the spatial distributions of HMs based on the determinant variables. For the wet deposition, the glm was only used for predicting spatial

patterns, and the important variables were selected according to previous research (Stankwitz et al., 2012; Zang et al., 2021).

Table 3. The generalized linear models used in this study.

Determinant variables selection	
Soil	land use+ GDP+ Population density+ nightlight+ distance to road
Sediment_1	NDVI+ precipitation+ temperature+ erosion+ grain size
Sediment_2	Zircon+ Apatite+ Rutile+ Garnet+ Tourmaline+ Ilmenite+ Magnetite+ Amphibole+ Epidote+ Pyroxene+ Limonite+ Sphene
River water	human variables (GDP+ P_density+ nightlight+ d_road+ Land use) +Natural variables (NDVI+ precipitation+ temperature+ erosion) +Secondary carriers (wet deposition+ soil+ sediment)
Spatial distribution prediction	
Soil	[GDP+ nightlight+ P_density] + [terrain (dem+ slope+ tpi+ rough+ tri)]
Sediment	[precipitation+ temperature] + [terrain (flowdir+ dem+ slope+ tpi+ rough+ tri)]
Deposition	[precipitation+ rainfall intensity+ humidity] + [terrain (dem+ slope+ tpi+ rough+ tri)]
River water	[GDP+ d_road+ nightlight+P_density] + [precipitation+temperature] + [Secondary carriers] + [terrain (flowdir+ dem+ slope+ tpi+ rough+ tri)]

Generalized additive models (gam), where the response variable is not restricted to be linear in the explanatory variables, were introduced to further estimate the smooth components of the glm models using smooth functions. According to the characteristics of our dataset, only five important variables were selected in the final models to depict the relationships between river water HMs and their explanatory variables. Considering some of the HM concentrations were between 0 to 1 μg/L(mg/kg), we kept the negative prediction values during the gam analysis. Prior to model fitting procedure, environmental and human variables were scaled by subtracting the mean and dividing by the standard deviation to enhance comparability of effect sizes across variables measured in different scales. The collinearity between the explanatory variables was considered and only the most significant variables were retained for further analysis.

3.4 Model assessment and outputs

The model fitting should avoid overfitting or underfitting and achieve a low generalization error that characterizes its prediction performance. In order to build such model, model selection and model averaging were used here to decide which model to select from candidate model families based on performance evaluations.

3.4.1 Model selection

When multiple model algorithms or candidate models were fitted, model selection was conducted to select the top model from a set of best models. The top model is the most parsimonious combination of explanatory variables using the cross- validation process. There are many selection criterions for model selection, and Akiake Information Criterion (AIC) is the most commonly used criteria, which is a measure of the goodness fit of an estimated statistical model. The best models were then ordered by the criteria of AIC. The AIC is a measure of fit that penalizes for the number of variables. And AICc works better when the sample sizes are small:

$$\mathrm{AICc} = \mathrm{AIC} + \frac{2K(K+1)}{n-K-1} \quad (1)$$

Smaller values indicate better fit and thus the AICc can be used to compare models. Thus, models with the lowest AICc are considered to be the top models, highlighting the included variables and model support relative to other models within 2 AICc.

3.4.2 Model averaging

The model averaging was proposed for addressing the issues of uncertainty in the choice of probability distribution functions and the biased regression parameters during the model selection process (Burnham et al., 2011). With dataset and the smooth functions used in this study, the selection of variables was based on: (a)whether corresponding to the research objectives well; (b) whether consistent with the related results in section 3.1 and 3.2; (c) sum of weights; and (d) whether collinear to each other. In this study, only the most important five explanatory variables were retained in the final model set to predict the relationships for each HM element in river water.

Once the relationships between individual HMs and explanatory variables were established, we created the spatial maps of HMs in river water by projecting the model onto environmental layers. The data preparation, analyses, and mapping in this study were

achieved in ArcGIS 10.5 (Environmental Systems Research Institute Inc.) and R v3.5.1 (R Core Team, 2018).

4 Results

4. 1 Importance of explanatory variables for secondary carriers

4.1.1 Human driven variables via soil carrier

All the peak values of HM concentrations in soil were observed in the middle reach areas, where the average concentration of As in middle areas was as twenty times as that in the upstream areas (Table 4). According to the results of RDA, the longer arrows mean this variable plays an important role in the response matrix. The GDP and 'distance to road' strongly drove the variation along the first axis, while other explanatory variables along the second axis (Fig. 6). The elements of As, Zn and Pb significantly correlated with the GDP, especially in the middle stream areas. In contrast, the 'distance to road' was related to the HMs in the downstream area. Cd centered in the coordination system, showing a correlation with most of the explanatory variables. The land use types of forest (land_1) and mining (land_5.3) played an important role in the HM compositions in the upstream area, which consisted with the environmental and socio-economic conditions in this area.

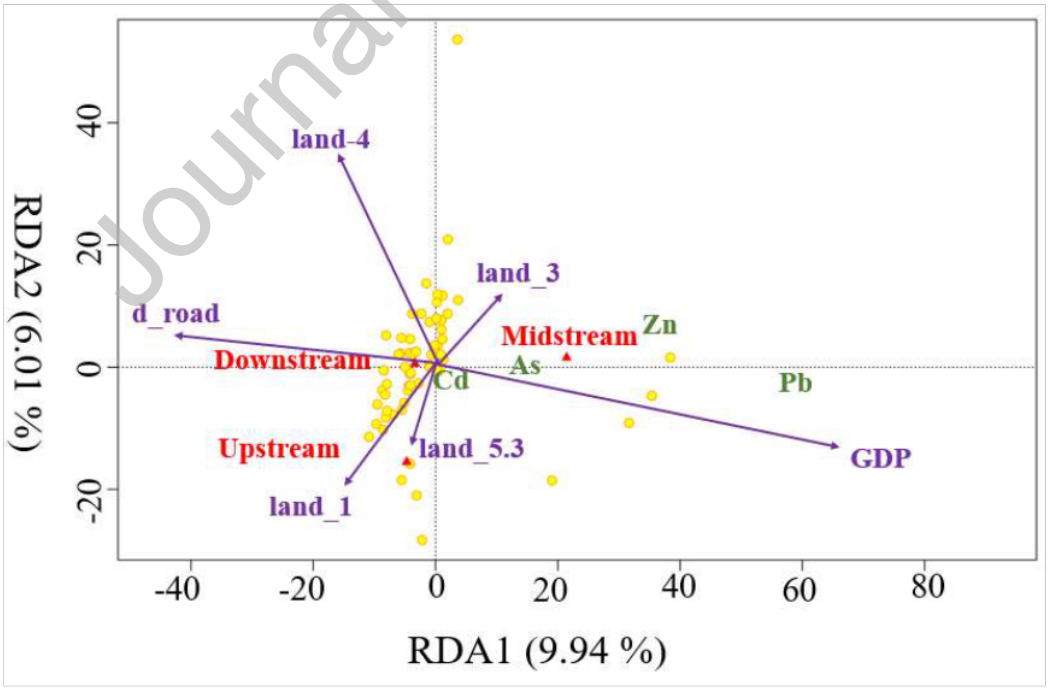


Fig. 6. Triplot of redundancy analysis (RDA) showing the effects of human activity variables on soil heavy metal(loid)s (R square=0.60). The yellow circles are the soil sampling sites, purple vectors represent the effect of explanatory variables, the red triangles are the catchment segments, and the green characters are plotted as the response variables.

Table 4. Average heavy metal(loid) concentrations of different carriers during summer period

Carrier	Segment	As	Cd	Cr	Cu	Hg	Pb	Zn	unit
water	up	0.53	0.08	3.98	2.93	0.11	5.37	38.98	μg/L
	middle	0.68	0.31	4.67	3.54	0.33	6.98	32.42	
	down	0.56	0.25	3.62	3.57	0.25	5.74	49.17	
soil	up	15.36	1.64	35.44	34.87	0.04	24.72	107.62	mg/kg
	middle	320.55	2.70	91.68	46.03	0.06	1456.15	1030.23	
	down	1.45	0.21	29.12	14.88	0.02	4.72	55.57	
wet deposition		0.22	0.37	2.59	6.87	0.04	4.20	-	μg/L
sediments		-	-	255.17	-		28.79	-	mg/kg

Top models with low AICc values were listed in Table 5. We hypothesized that the soil HM compositions might be dominated by different land use types, but only the grassland showed correlation with Cd from the top models. While 'distance to road', GDP and nightlight were more likely than land use types to be correlated with soil HMs.

Table 5. Top models identified through model selection based on AICc. df= Degrees of freedom. Weight= Akaike weight.

HM	smoothed variables	df	loglik	AICc	weight
soil					
As	d_road+ GDP+ nightlight	5	-388.05	787.12	0.33
	d_road+ GDP+ land5.3+ nightlight	6	-387.58	788.6	0.16
Cd	d_road+ GDP+ land2+ P_density	6	-119.68	252.81	0.11
	d_road+ GDP+ land2	5	-121.03	253.07	0.1
Pb	d_road+ GDP+ nightlight	5	-495.92	1002.86	0.43
	d_road+ GDP+land_3+ nightlight	6	-495.46	1004.36	0.2
Zn	d_road+ GDP+ nightlight	5	-471.43	953.87	0.11
	d_road+ GDP+ land_5.2+ nightlight	6	-470.28	954.01	0.1
sediment					
Cr	2000μm+ mean size+ temperature	5	-185.93	384.16	0.14

	Garnet+ Zircon	4	-152.16	314.32	0.23
Ni	63μm+ 500μm+ mean size	5	-140.64	293.58	0.24
	Garnet+ Zircon	4	-123.19	256.37	0.32
Pb	land_3+ precipitation	4	-108.89	227.27	0.09
	Magnetite+ Sphene	4	-79.28	168.55	0.66
water					
As	d_road+ precipitation	4	-9.12	27.19	0.09
Cr	Cr.sedi+ Cr.wet	4	-77.48	163.92	0.22
Cu	d_road+ land_2+ land_5+ nightlight	8	-57.93	135.65	0.16
Hg	d_road+Hg.wet+land2+land3+land4+land5+nightlight	11	44.7	-59.86	0.04
Pb	d_road+GDP+land5.1+P_density+Pb.sedi+Pb.wet	8	-92.16	204.1	0.09
Zn	d_road+GDP+land1+land2+land4+land5.1	8	-169.17	358.13	0.39

Symbols for explanatory variables in this table are distance to road (d_road), population density (P_density), the related element concentration in sediment (element.sedi), and the related element concentration in wet deposition (element.wet).

4.1.2 Environment driven variables via sediment carrier

The triplot of Fig. 7A showed that the grain sizes of 0-63, 250-500 μm explained together 54.08% of the total variance of the sediment HM data. The 0-63 μm size showed significant correlations with the sediment HMs in the downstream area, while the upstream area was related to the precipitation amounts. For the lithology RDA (Fig. 7B), the drivers of Ilmenite, Garnet, and Magnetite were distributed along the first axis (36.54%). The HM variations in the upstream showed correlative relationships with the Epidote minerals, while downstream was correlated to the lithologies of Sphene and Tourmaline.

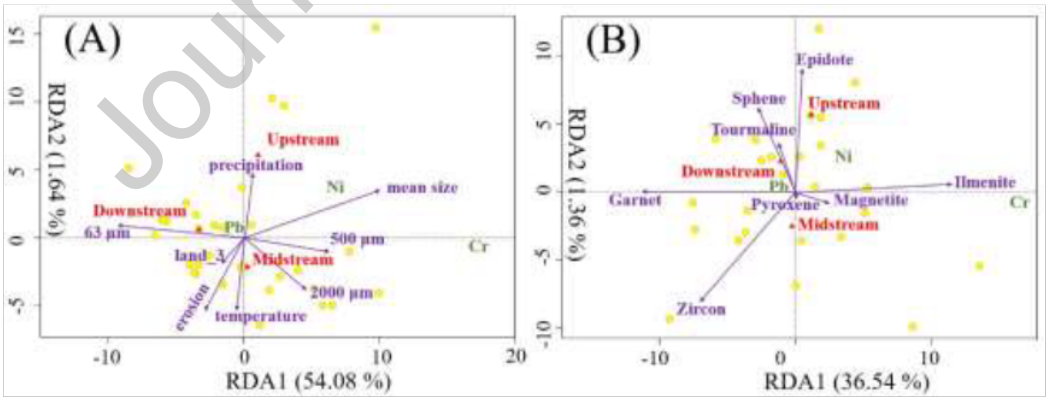


Fig. 7. Triplots of Redundancy analysis (RDA) showing the effects of environmental variables on sediment heavy metal(loid)s (R square=0.56 and 0.38, respectively). The yellow circles are the sediment sampling sites, purple vectors represent the effect of environmental variables, the red triangles are the catchment segments, and the green characters are plotted as the response variables. Symbols for explanatory variables are the grain size of 0-63 μm

(63μm), 250-500 μm (500 μm), 500-2000 μm (2000 μm), the mean size of all the sediments (mean size) and the soil erosion classification (erosion).

The '500-2000μm grain size', 'mean size', and 'temperature' were the best explanatory variables for Cr in the sediments (Table 5). And the 'Garnet' and 'Zircon' lithology strongly correlated to the Cr. Additionally, the '0-63 and 250-500 grain sizes' showed strongest correlation with Ni, and it had the same best lithology variables as Cr. In contrast, the grain sizes had no significant impacts on Pb, but the 'land use type of water' and 'precipitation' showed great correlations with the accumulation of Pb in sediment. Unlike Cr and Ni, the lithology type of Magnetite and Sphene were found to predict the Pb best (Table 5).

4. 2 Spatial HM habitats in secondary carriers

The explanatory variables we used to predict the HM distribution in soil were listed in Table 3, including GDP, nightlight, and population density, as well as the terrain characteristic factors. The concentrations of As, Pb, and Zn showed peak values around the big cities in middle areas, especially Jiuquan and Jiayuguan where the typical industrial cities are located in northwestern China (Fig. 8).

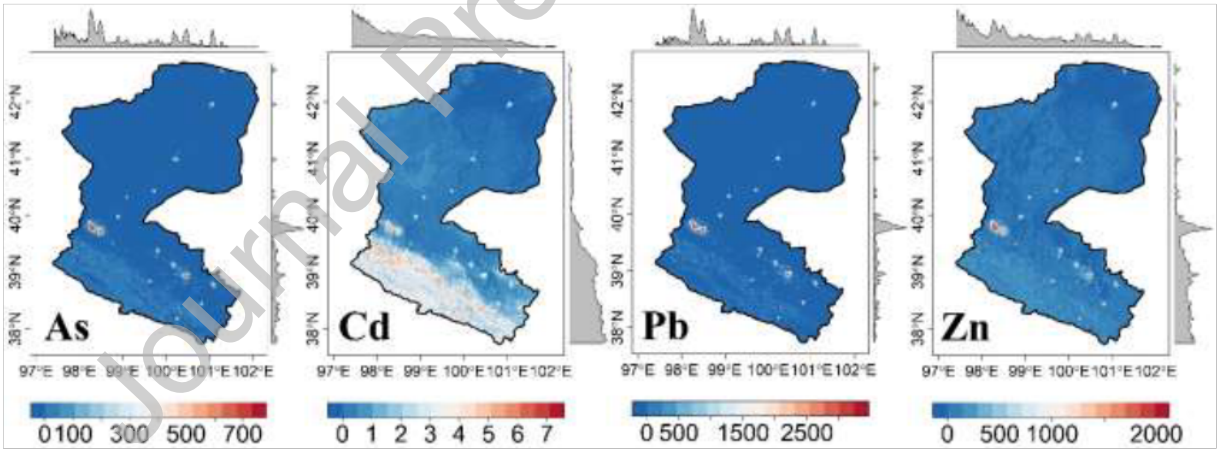


Fig. 8. Predicted maps of soil heavy metal(loid) habitats (Unit: mg/kg).

According to the results of glm in Table 5, the precipitation amount and temperature were selected as the explanatory variables to predict the spatial distribution of HMs in the sediment, and the terrain characteristics were also considered in the SDM. The output maps indicated that concentrations of all the HMs showed an increasing trend from upper reaches to lower reaches, with Pb showing the most obvious increase in the downstream areas (Fig. 9). However, high values of each element in the upstream spread out spatially, reflecting the influence of the hydraulic sorting process.

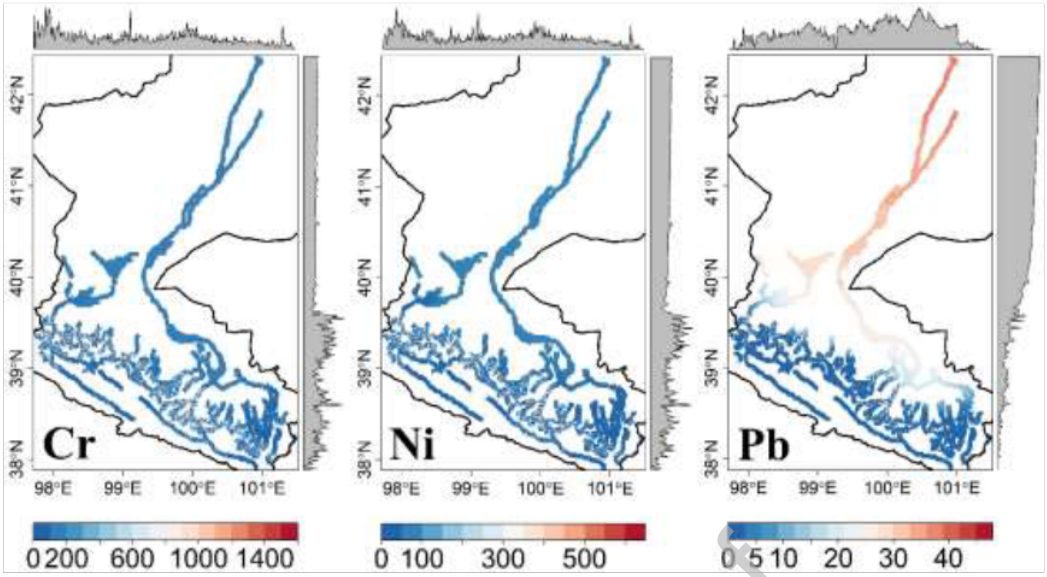


Fig. 9. Predicted distributions of sediment heavy metal habitats (Unit: mg/kg).

Precipitation amount, rainfall intensity, humidity, and terrain factors were selected as the explanatory variables to predict the spatial distributions of HMs in atmospheric deposition (Table 3). Output maps showed that the high values of HMs in wet deposition were distributed in the downstream areas, but relatively lower in the upstream and middle stream segments (Fig. 10). This pattern could be explained by the scavenging effect and washout process of precipitation. In downstream areas where precipitation is scarce, HMs will be continuously accumulated in atmospheric particles, so high concentrations of HMs were observed during the limited precipitation events. In contrast, the large amount and high frequency of precipitation in the upstream areas made the concentrations of HMs relatively lower.

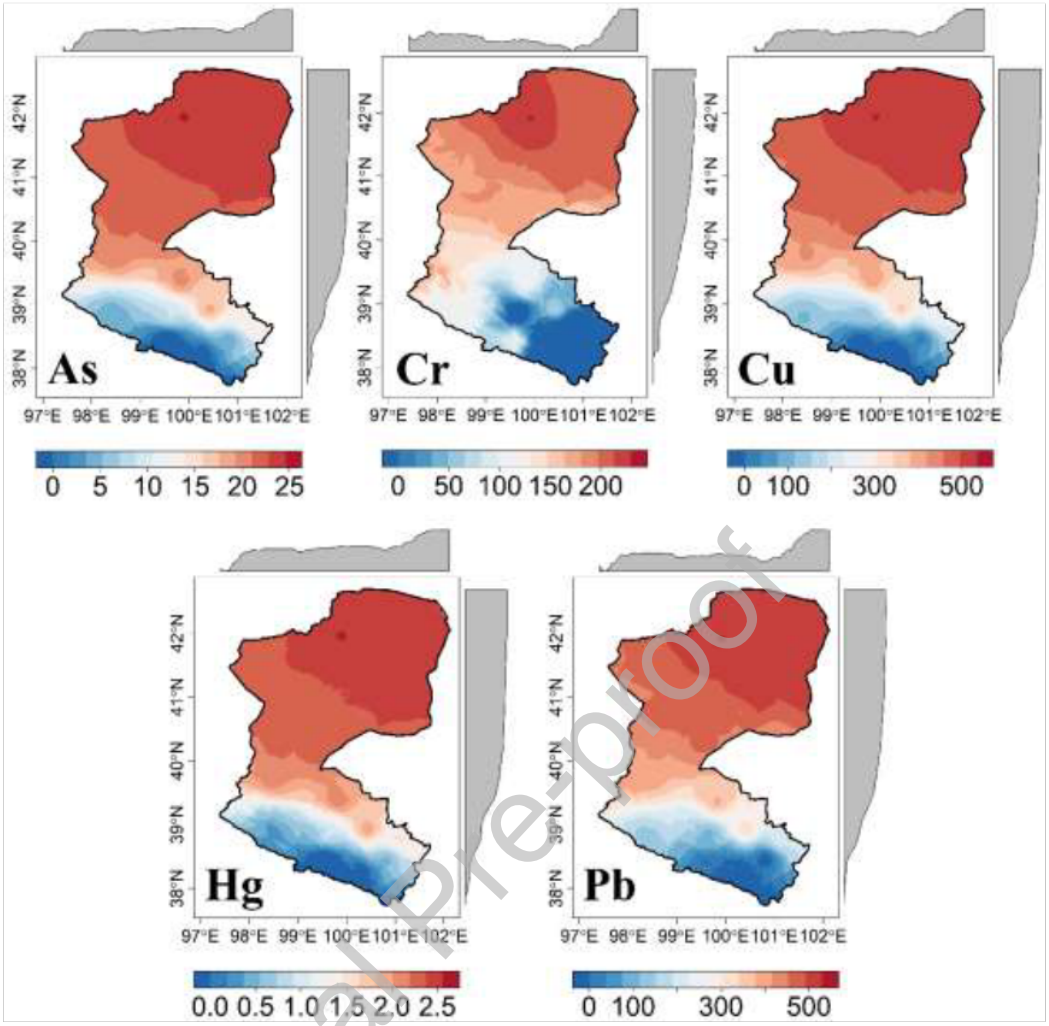


Fig. 10. Predicted distributions of wet deposition heavy metal(loid) habitats (Unit: μg/L).

4.3 HM transmission at large scale catchment

In summer period, the highest values of As, Cd, Cr, Hg and Pb in river water were all in the middle stream areas, while the upstream areas had the lowest values for most of the HMs (Table 4).

4.3.1 Spatiotemporal distributions of HMs in river water

Except for Cu and Zn, the other four elements in the middle reach were significantly higher than those in other sections. As and Cr in the middle reach were significantly higher in winter than in summer, while Cu and Hg showed an opposite trend (Fig. 11).

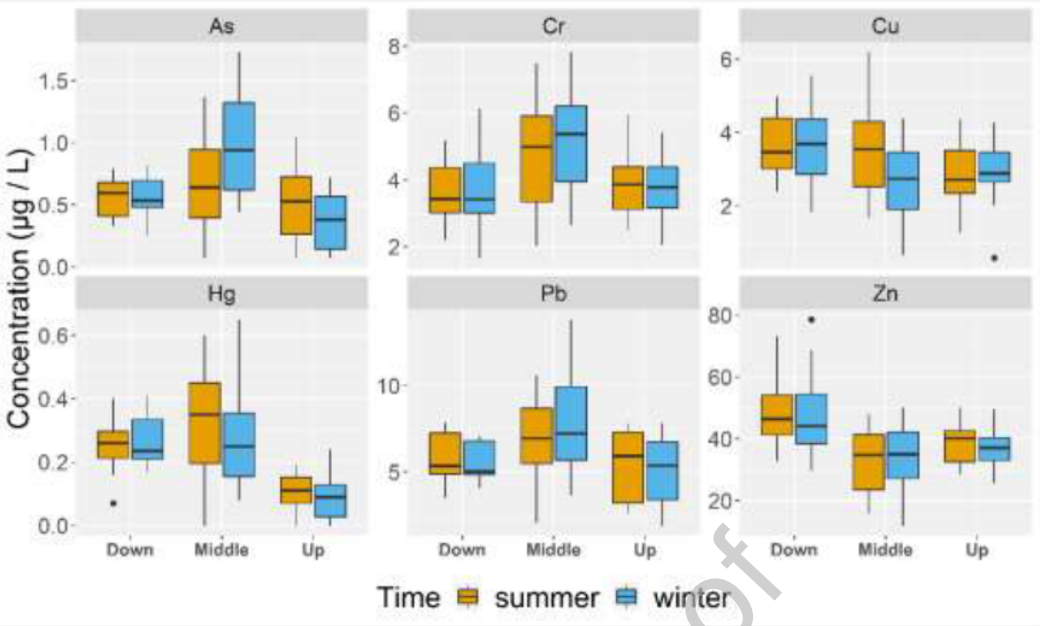


Fig. 11. The temporal and spatial distributional characteristics of water heavy metal(loid)s. (Down stands for downstream, Middle for middle stream, and Up for upstream).

4.3.2 Importance of variables for HMs in river water

According to the top models of HMs in river water (Table 5), As showed correlations with 'distance to road' and precipitation amount. The top model of Cr was explained by the corresponding Cr compositions from secondary carriers, including the sediment and wet deposition. The variation of Hg was influenced by the 'distance to road', 'night light', most of the land use types, as well as the Hg accumulation process from the wet deposition (Table 5). Additionally, the HM transport and accumulation processes from sediments and wet deposition played the most important role in the river water Pb compositions, followed by the economic parameters, including 'distance to road', GDP, 'land use type of agriculture' and 'population density'.

4.3.3 Relationships between HMs in river with all explanatory variables

Considering the importance of the explanatory variables and our study objectives, the top five significant variables were selected as the final variables to explain the HM compositions in river water (Table 6). The 'distance to road' showed negative relationships with As, Cu, Hg and Pb, but a positive influence on Zn, and a non-linear relationship with Cr (Fig. 12). The GDP had positive relationship with Pb, but negative with Cr and Zn. While for As, Cu and Hg,

the GDP showed a negative trend at first until the GDP increased to 500, then turned into positive relationships. Precipitation had positive relationships with As and Cu, but roughly negative relationships with Cr and Hg. The influence from wet deposition, showed positive relationships with As, Cu and Pb, but negative with Cr, and non-linear relationship with Hg. HMs from soil carrier greatly contributed to the accumulation of As, Pb, and Zn. It firstly showed a positive trend, but then turned into negative trend when the concentration of As and Pb were around 7.5 ppm.

Table 6. Generalized Additive Models describing the response of heavy metal(loid)s in river water to the major explanatory variables.

HM	variables	Edf	ref.df	F	P	weights	N	DE
As	d_road	1.00	1.00	7.23	0.01	1	5	31.30%
	GDP	2.29	2.77	1.14	0.28	0.42	2	
	precipitation	1.34	1.57	0.10	0.12	0.13	1	
	As.wet	1.00	1.00	0.07	0.79	0.13	1	
	As.soil	1.96	2.35	1.59	0.18	0.12	1	
Cr	d_road	1.57	1.94	0.65	0.07	1	5	48%
	GDP	1.00	1.00	4.62	0.04	0.54	2	
	precipitation	5.73	6.75	2.65	0.03	0.87	4	
	Cr.wet	1.00	1.00	3.05	0.09	0.63	3	
	Cr.sedi	1.00	1.00	0.29	0.59	0.13	1	
Cu	d_road	1.00	1.00	2.02	0.16	0.63	8	41.50%
	GDP	1.49	1.78	0.49	0.61	0.5	7	
	precipitation	1.00	1.00	2.49	0.12	0.38	5	
	Cu.wet	1.00	1.00	3.02	0.09	0.24	3	
	nightlight	5.68	6.57	1.71	0.19	0.24	4	
Hg	d_road	1.00	1.00	0.54	0.47	-	-	76.50%
	GDP	2.19	2.74	1.19	0.30	-	-	
	precipitation	2.27	2.80	5.54	0.01	1	2	
	Hg.wet	7.37	8.24	3.03	0.01	-	-	
	land_5.1	1.00	1.00	0.93	0.34	0.59	1	
Pb	d_road	1.00	1.00	1.46	0.23	0.51	4	46.10%
	GDP	1.00	1.00	1.93	0.17	0.39	3	
	Pb.wet	1.47	1.76	4.33	0.08	0.58	4	
	Pb.sedi	2.91	3.55	1.61	0.25	0.48	3	
	Pb.soil	2.35	2.80	3.35	0.02	1	7	
Zn	d_road	1.00	1.00	3.99	0.05	0.33	2	41.30%
	GDP	1.00	1.00	2.26	0.14	0.67	4	
	Zn.soil	4.10	4.83	2.46	0.06	0.67	4	
	land_1	1.00	1.00	0.83	0.37	0.42	2	
	land_5.1	1.66	2.06	0.69	0.53	0.14	1	

Symbols in this table are Degrees of freedom (Edf), Degrees of freedom estimated to waste (Ref.df), Significance of smoothed terms (F), P values (P), Deviance explained

(DE), sum of weights (weights), and the number of times that each explanatory variable was included in the total models (N).

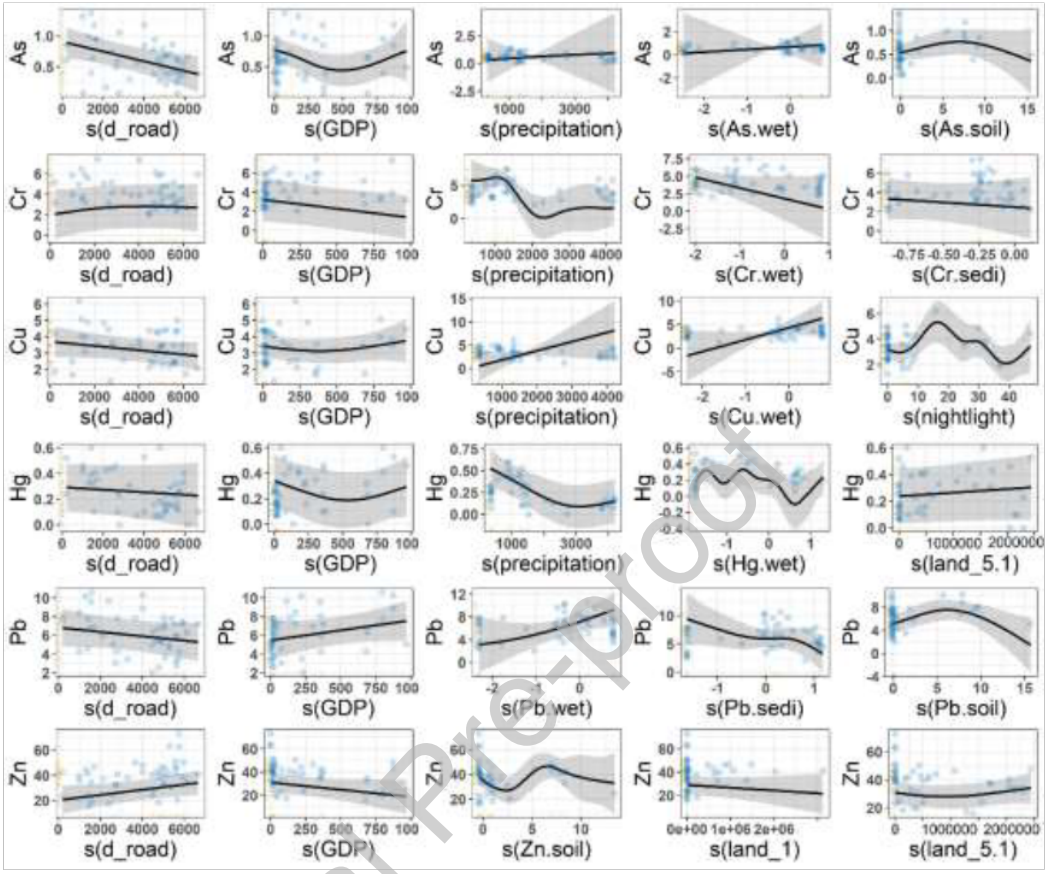


Fig. 12. Smoothed fits of relationships between heavy metal(loid)s in river water and the final five explanatory variables.

4.3.4 HM habitats and hotspots in river water

The human activity variables (including GDP, population density, nightlight, distance to road, and land use types) and environmental variables (including NDVI, precipitation, temperature, and erosion classification) were chosen as the predictors in SDM to predict the HMs in river water. In order not to discount the influence of HM transport and accumulation processes among different carriers, the corresponding HMs from secondary carriers were also considered in our study. And the output maps with considering the HMs in secondary carriers were presented in Fig. 13.

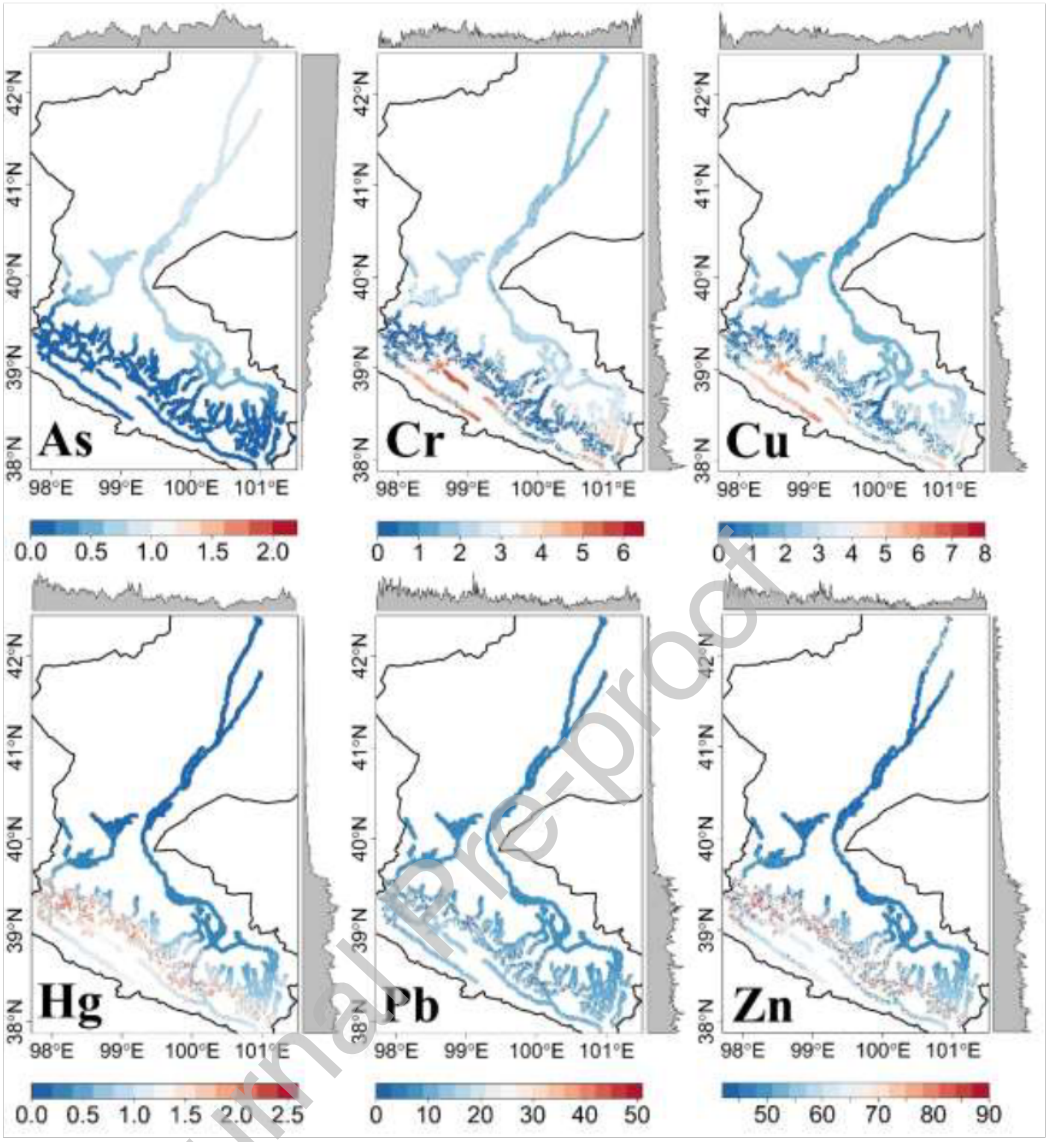


Fig. 13. Predicted distributions of heavy metal(loid)s in river water, with considering the related heavy metal(loid)s in secondary carriers (Unit: μg/L).

Compared with the output maps without considering the HMs in secondary carriers, the prediction accuracy of maps in Fig.13 had been significantly improved. For instance, there were more hotspots of As around the cities if we considered the contributions of As accumulation from soil and wet deposition (Fig. 14A). Additionally, more hotspots of As and Pb were predicted in the upstream areas when we added the HM compositions of secondary carriers in the SDM (Fig. 14). The predicted intervals of Pb compositions showed an increased trend when we considered the Pb accumulations from soil, sediment and wet deposition (Fig. 14B).

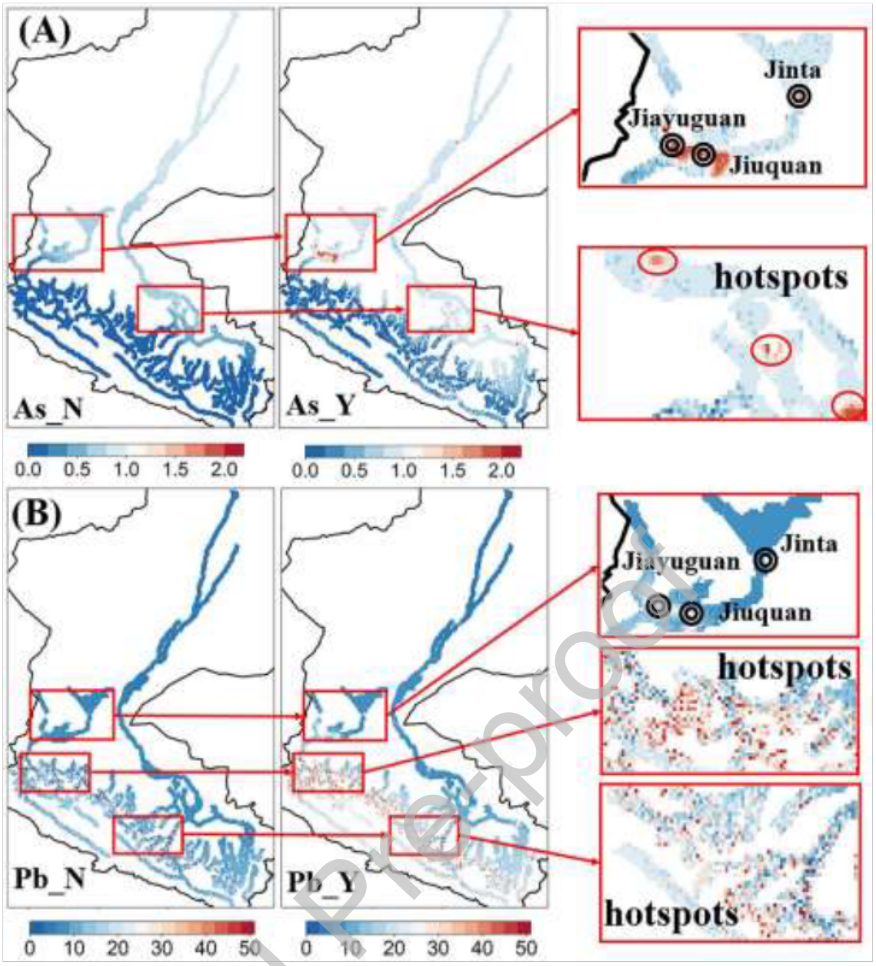


Fig. 14. Comparisons of improved prediction accuracy with considering the corresponding HM in secondary carriers. As_N and Pb_N stand for 'without considering the secondary carriers', while As_Y and Pb_Y means considering the HM in secondary carriers.

5 Discussion

5.1 Creation of comprehensive and consistent explanatory variables for better prediction of HMs

Insufficient explanatory variables will greatly hinder the completeness and accuracy of investigating the HM transport and accumulation processes. Thus, the selection of appropriate indirect and direct potential explanatory variables has been shown as a key step for identifying the HM sources and depicting the transport processes (Nickel et al., 2014). Estimating the explanatory variables has previously been undertaken using three main conventional approaches. (i) Collect samples in typical land use areas to be considered as the

prevailing explanatory factor (Li et al., 2017b; Liang et al., 2017a; Liang et al., 2017b; Sharma et al., 2008). Such systematic reconnaissance survey requires the evaluation of the land use types or other information before sampling (Sharma et al., 2008). (ii) Sample the point source pollutions around typical locations (Hochella Jr et al., 2005; Liang et al., 2017b; Quinton and Catt, 2007). The mining operations are important point sources in previous research, like the waste rock and smelter waste (Liang et al., 2017b). (iii) Monitor a series of synchronous factors by some equipment in the field. For example, Zang et al. (2021) arranged an equipment in the field to collect the synchronous rainfall amount, intensity, duration, and wind speed, etc., during each precipitation event. Nevertheless, these conventional methods are always constrained by the complex field situations, leading to a lack of observation sites. Recently, Hu et al. (2020) proposed that the use of geographic information system (GIS), remote sensing products and other geographical information can help to extract comprehensive explanatory variables over large-scale region. However, this study only used land use maps as auxiliary variables, without considering other human variables.

Species Distribution Models (SDMs) can describe the relationships between species and corresponding explanatory variables, and predict their responses to the changing explanatory variables (Kearney and Porter, 2009) (Fig. 2). Another important advantage of SDM is that it can superimpose and calculate many explanatory variables at the same time. In this study, we systematically collected comprehensive corresponding variables from the available GIS and remote sensing products over the study area. The prediction results of SDM in this study not only considered the environmental background information (topography, water flow direction, etc.), but also took into account the influence of human activities (land use, GDP, population density, public transport, etc.). More importantly, this study also used the secondary carriers as explanatory variables to explain and predict the HMs in river water. Comparing to the output maps without considering the secondary HM concentrations, we found it may improve the prediction quality (Fig. 13, Fig. 14). For example, we didn't have soil samples in Jiuquan and Jiayuguan areas, but the SDM model predicted that there were HM hotspots in this place (Fig. 8). The spatial variations of these HM compositions were consistent with the previous studies (Guan et al., 2018), which proved the feasibility and universal applicability of our model. Incorporating more data layers of influencing factors into SDM can potentially increase the accuracy of the prediction of HM distributions. Therefore, more factors that affect HM transport should be considered in the future study when such data become available, including extreme precipitation events, geochemical backgrounds, etc.

5.2 Solution to the field sampling limitations

Field sampling provides core scientific data for studying the interactions among various carriers, with an important prerequisite of geographical continuity for establishing connections between different carriers (Fig. 1). To ensure this, the main solution of previous research was the 'proximity' strategy during the sampling process (Table 1). The typical research include the proximity between the soil and river water (Hassellov and von der Kammer, 2008; Quinton and Catt, 2007), soil and crops (Hu et al., 2020), atmospheric deposition and soil (Liang et al., 2017b), atmospheric deposition and vegetables (Sharma et al., 2008), flood plain and riverbed sediment (Hochella Jr et al., 2005). However, the major challenge in current proximity methods is how to ensure the spatial match among layers of different carriers, which otherwise would greatly reduce the accuracy of subsequent analysis, especially over the large-scale regions. Complex terrains, harsh weather, undeveloped public transport system, and other constraints at the field make it challenging or even impossible to collect multi-dimensional carriers at the same point (Li et al., 2020). For the entire catchment, especially with large scales, different carriers' collecting sites cannot distribute at the same sampling density or rate under actual circumstances. A dearth of consistent data for different carriers is often credited as a major limitation for HMs interaction research (Liu et al., 2019).

SDM has advantage of using the presence-only data, which are the field observations where the presences of the species were observed. SDM thus doesn't require the normal distribution and spatial continuation of the field data (Fig. 4), and is friendly to scientific research involves in large-scale area, where continuous sampling is normally impossible. The sampling in this study was based on SDM requirement and collected different carriers in typical areas across the entire catchment as much as possible. The output maps confirmed that the SDM worked well to map the HM distribution, especially in areas that lack of field samples.

5.3 Ecologicalization of HM contaminants by SDM

The spatial distribution of HMs is an important indicator to estimate the environmental pollution (Li et al., 2017b), which was mostly achieved by the Kriging interpolation as one of the most prevalent spatial interpolation methods in previous studies (Li et al., 2017b; Liang et al., 2017a). Strong evidence has shown that increasing the sampling density is the main approach to improve the accuracy of predictions (Lv, 2019), but the conventional interpolation methods ignore the influence of many determinant variables when investigating the HM transport and accumulation processes (Li et al., 2017b). Lv (2019) explored to apply the geostatistical techniques to estimate the HM spatial distribution based on the factors

resulting from the receptor models. Nickel et al. (2014) proposed to combine multivariate generalized linear models with Kriging interpolation to create HM maps at a high level of spatial resolution. However, both of these two studies had to increase the sampling density to ensure the accuracy of the HM prediction (Table 1), which is often challenging over large catchment. In addition, lots of preparatory work for these two studies was required, for example, the explanatory variables must be calculated and selected through the acceptor models beforehand.

In this study, we innovatively adopt the SDM by ecologicalizing the HMs in catchment as different species. When studying the spatial distributions of different HMs in the same carrier, "different HMs in the same carrier" were regarded as "species of the same trophic level" to predict their spatial distributions (Fig. 5). Similarly, the HM transport among different carriers were converted into the food chain relationships in the SDM (Fig. 5), and each carrier was assumed to have its own habitats. A few research entities had been ecologicalized as different species in previous studies, such as human emotions (Li et al., 2021), visitors and hikers (Coppes and Braunisch, 2013). Based on our HM output maps, the prediction accuracy has been significantly improved, thus we believe SDMs have big potentials for mapping the HM distributions. Due to the scarcity of field data in this study, some typical big-data driven SDM models such as Maxent cannot be used. Instead, we built our own SDM framework to address the issue. Future work can be conducted to explore the use of different advanced SDMs to further improve the heavy metal mapping research.

6 Conclusion

The spatial inconsistence and mismatch of the field samplings for different HM carriers profoundly hinder our ability to map HMs over large areas. To address this issue, this study developed a novel methodological framework for mapping HM contamination over the large-scale catchment with a species distribution model. Results found that the variables of distance to road, GDP, and nightlight played a relatively important role in accumulation of soil HMs. The output maps of HMs from soil, sediment, and wet deposition could respectively reflect the influence of industry contaminants, hydraulic sorting, and precipitation washout process, which proved the rationality of SDM in predicting the spatial distributions of individual carriers. The environment and human variables, together with the HM transport from the secondary carriers were selected as explanatory variables to predict the HMs in river water, which helped to better predict spatial dynamics and transport mechanism of HMs. Although future work is needed to overcome existing biases and challenges, the methodological

framework proposed in this study with ecological perspective of SDM provides a new valuable future rigorous exploration in hydrological chemistry and pollution research.

Acknowledgements

We acknowledge the editor and reviewers for their valuable advice on improving this study. We thank professor Jianquan Liu, Dr. Apurva Kakade from Lanzhou University for their encouragement and suggestions on this study, and also thank Dr. Jian Zhang for sharing the sediment dataset. This work was supported by the Natural science Foundation of China (Project No:41801262). We want to thank the resources and environment data center of the Chinese Academy of Sciences for data support.

Declaration of interests

Dear Editor,

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.

Thanks very much for your consideration!

Zunyi Xie
Professor at Henan University

References

Ali, H., Khan, E. and Ilahi, I. 2019. Environmental chemistry and ecotoxicology of hazardous heavy metals: environmental persistence, toxicity, and bioaccumulation. Journal of chemistry 2019.

Badry, A., Palma, L., Beja, P., Ciesielski, T.M., Dias, A., Lierhagen, S., Jenssen, B.M., Sturaro, N., Eulaers, I. and Jaspers, V.L. 2019. Using an apex predator for large-scale monitoring of trace element contamination: Associations with environmental, anthropogenic and dietary proxies. Science of the total environment 676, 746-755.

Burnham, K.P., Anderson, D.R. and Huyvaert, K.P. 2011. AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. Behavioral ecology and sociobiology 65(1), 23-35.

Cable, E. and Deng, Y. 2018. Trace elements in atmospheric wet precipitation in the detroit metropolitan area: levels and possible sources. Chemosphere 210, 1091-1098.

Cheng, G., Li, X., Zhao, W., Xu, Z., Feng, Q., Xiao, S. and Xiao, H. 2014. Integrated study of the water–ecosystem–economy in the Heihe River Basin. National science review 1(3), 413-428.

Clemente, P., Calvache, M., Antunes, P., Santos, R., Cerdeira, J.O. and Martins, M.J. 2019. Combining social media photographs and species distribution models to map cultural ecosystem services: The case of a Natural Park in Portugal. Ecological indicators 96, 59-68.

Coppes, J. and Braunisch, V. 2013. Managing visitors in nature areas: where do they leave the trails? A spatial model. Wildlife biology 19(1), 1-11.

Elith, J. and Franklin, J. (2013) Encyclopedia of Biodiversity: Second Edition, pp. 692-705, Elsevier Inc.

Fang, H., Huang, L., Wang, J., He, G. and Reible, D. 2016. Environmental assessment of heavy metal transport and transformation in the Hangzhou Bay, China. Journal of hazardous materials 302, 447-457.

Gajbhiye, T., Pandey, S.K., Kim, K.-H., Szulejko, J.E. and Prasad, S. 2016. Airborne foliar transfer of PM bound heavy metals in Cassia siamea: a less common route of heavy metal accumulation. Science of the Total Environment 573, 123-130.

Gassama, N., Curie, F., Vanhooydonck, P., Bourrain, X. and Widory, D. 2021. Determining the regional geochemical background for dissolved trace metals and metalloids in stream waters: Protocol, results and limitations—The upper Loire River basin (France). Water 13(13), 1845.

Ge, Y., Li, X., Huang, C. and Nan, Z. 2013. A Decision Support System for irrigation water allocation along the middle reaches of the Heihe River Basin, Northwest China. Environmental Modelling & Software 47, 182-192.

Goth, A., Michelsen, A. and Rousk, K. 2019. Railroad derived nitrogen and heavy metal pollution does not affect nitrogen fixation associated with mosses and lichens at a tundra site in Northern Sweden. Environmental Pollution 247, 857-865.

Guan, Q., Wang, F., Xu, C., Pan, N., Lin, J., Zhao, R., Yang, Y. and Luo, H. 2018. Source apportionment of heavy metals in agricultural soil based on PMF: A case study in Hexi Corridor, northwest China. Chemosphere 193, 189-197.

Guisan, A. and Zimmermann, N.E. 2000. Predictive habitat distribution models in ecology. Ecological modelling 135(2-3), 147-186.

Harrigan, R.J., Thomassen, H.A., Buermann, W. and Smith, T.B. 2014. A continental risk assessment of West Nile virus under climate change. Global change biology 20(8), 2417-2425.

Hassellov, M. and von der Kammer, F. 2008. Iron oxides as geochemical nanovectors for metal transport in soil-river systems. Elements 4(6), 401-406.

Hijmans, R.J. and Elith, J. 2013. Species distribution modeling with R. R Cran Project.

Hochella Jr, M.F., Moore, J.N., Putnis, C.V., Putnis, A., Kasama, T. and Eberl, D.D. 2005. Direct observation of heavy metal-mineral association from the Clark Fork River Superfund Complex: Implications for metal transport and bioavailability. Geochimica et cosmochimica acta 69(7), 1651-1663.

Hong, N., Zhu, P., Liu, A., Zhao, X. and Guan, Y. 2018. Using an innovative flag element ratio approach to tracking potential sources of heavy metals on urban road surfaces. Environmental Pollution 243, 410-417.

Hu, B., Xue, J., Zhou, Y., Shao, S., Fu, Z., Li, Y., Chen, S., Qi, L. and Shi, Z. 2020. Modelling bioaccumulation of heavy metals in soil-crop ecosystems and identifying its controlling factors using machine learning. Environmental Pollution 262, 114308.

Kearney, M. and Porter, W. 2009. Mechanistic niche modelling: combining physiological and spatial data to predict species' ranges. Ecology letters 12(4), 334-350.

Legendre, P. and Anderson, M.J. 1999. Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments. Ecological monographs 69(1), 1-24.

Levin, N., Kyba, C.C., Zhang, Q., de Miguel, A.S., Román, M.O., Li, X., Portnov, B.A., Molthan, A.L., Jechow, A. and Miller, S.D. 2020. Remote sensing of night lights: A review and an outlook for the future. Remote Sensing of Environment 237, 111443.

Li, J., Li, Z., Brandis, K.J., Bu, J., Sun, Z., Yu, Q. and Ramp, D. 2020. Tracing geochemical pollutants in stream water and soil from mining activity in an alpine catchment. Chemosphere 242, 125167.

Li, J., Li, Z. and Feng, Q.J.E.E.S. 2017a. Impact of anthropogenic and natural processes on the chemical compositions of precipitation at a rapidly urbanized city in Northwest China. 76(10), 1-14.

Li, X., Yang, H., Zhang, C., Zeng, G., Liu, Y., Xu, W., Wu, Y. and Lan, S. 2017b. Spatial distribution and transport characteristics of heavy metals around an antimony mine area in central China. Chemosphere 170, 17-24.

Li, Y., Fei, T., Huang, Y., Li, J., Li, X., Zhang, F., Kang, Y. and Wu, G. 2021. Emotional habitat: Mapping the global geographic distribution of human emotion with physical environmental factors using a species distribution model. International Journal of Geographical Information Science 35(2), 227-249.

Liang, J., Feng, C., Zeng, G., Gao, X., Zhong, M., Li, X., Li, X., He, X. and Fang, Y. 2017a. Spatial distribution and source identification of heavy metals in surface soils in a typical coal mine city, Lianyuan, China. Environmental Pollution 225, 681-690.

Liang, J., Feng, C., Zeng, G., Zhong, M., Gao, X., Li, X., He, X., Li, X., Fang, Y. and Mo, D. 2017b. Atmospheric deposition of mercury and cadmium impacts on topsoil in a typical coal mine city, Lianyuan, China. Chemosphere 189, 198-205.

Liu, H.-L., Zhou, J., Li, M., Hu, Y.-m., Liu, X. and Zhou, J. 2019. Study of the bioavailability of heavy metals from atmospheric deposition on the soil-pakchoi (Brassica chinensis L.) system. Journal of hazardous materials 362, 9-16.

Lu, Z., Feng, Q., Xiao, S., Xie, J., Zou, S., Yang, Q. and Si, J. 2021. The impacts of the ecological water diversion project on the ecology-hydrology-economy nexus in the lower reaches in an inland river basin. Resources, Conservation and Recycling 164, 105154.

Luo, M., Yu, H., Liu, Q., Lan, W., Ye, Q., Niu, Y. and Niu, Y. 2021. Effect of river-lake connectivity on heavy metal diffusion and source identification of heavy metals in the middle and lower reaches of the Yangtze River. Journal of Hazardous Materials 416, 125818.

Lv, J. 2019. Multivariate receptor models and robust geostatistics to estimate source apportionment of heavy metals in soils. Environmental pollution 244, 72-83.

Meite, F., Alvarez-Zaldívar, P., Crochet, A., Wiegert, C., Payraudeau, S. and Imfeld, G. 2018. Impact of rainfall patterns and frequency on the export of pesticides and heavy-metals from agricultural soils. Science of the Total Environment 616, 500-509.

Nickel, S., Hertel, A., Pesch, R., Schröder, W., Steinnes, E. and Uggerud, H.T. 2014. Modelling and mapping spatio-temporal trends of heavy metal accumulation in moss and natural surface soil monitored 1990–2010 throughout Norway by multivariate generalized linear models and geostatistics. Atmospheric Environment 99, 85-93.

Quinton, J.N. and Catt, J.A. 2007. Enrichment of heavy metals in sediment resulting from soil erosion on agricultural fields. Environmental science & technology 41(10), 3495-3500.

Sarkar, D.J., Sarkar, S.D., Das, B.K., Sahoo, B.K., Das, A., Nag, S.K., Manna, R.K., Behera, B.K. and Samanta, S. 2021. Occurrence, fate and removal of microplastics as heavy metal vector in natural wastewater treatment wetland system. Water Research 192, 116853.

Senduran, C., Gunes, K., Topaloglu, D., Dede, O.H., Masi, F. and Kucukosmanoglu, O.A. 2018. Mitigation and treatment of pollutants from railway and highway runoff by pocket wetland system; A case study. Chemosphere 204, 335-343.

Sharma, R.K., Agrawal, M. and Marshall, F.M. 2008. Heavy metal (Cu, Zn, Cd and Pb) contamination of vegetables in urban India: A case study in Varanasi. Environmental pollution 154(2), 254-263.

Stankwitz, C., Kaste, J.M. and Friedland, A.J. 2012. Threshold increases in soil lead and mercury from tropospheric deposition across an elevational gradient. Environmental science & technology 46(15), 8061-8068.

Stojic, N., Pucarevic, M. and Stojic, G. 2017. Railway transportation as a source of soil pollution. Transportation Research Part D: Transport and Environment 57, 124-129.

Taiwo, A.M., Harrison, R.M. and Shi, Z. 2014. A review of receptor modelling of industrially emitted particulate matter. Atmospheric environment 97, 109-120.

Trujillo-González, J.M., Torres-Mora, M.A., Keesstra, S., Brevik, E.C. and Jiménez-Ballesta, R. 2016. Heavy metal accumulation related to population density in road dust samples taken from urban sites under different land uses. Science of the Total Environment 553, 636-642.

Wijesiri, B., Liu, A., He, B., Yang, B., Zhao, X., Ayoko, G. and Goonetilleke, A. 2019. Behaviour of metals in an urban river and the pollution of estuarine environment. Water research 164, 114911.

Yan, K. and Ding, Y. 2020. The overview of the progress of Qilian Mountain National Park System Pilot Area. International Journal of Geoheritage and Parks 8(4), 210-214.

Yang, J., Su, K. and Ye, S.J.J.o.G.S. 2019. Stability and long-range correlation of air temperature in the Heihe River Basin. 29(9), 1462-1474.

Zang, F., Wang, H., Zhao, C., Nan, Z., Wang, S., Yang, J. and Li, N. 2021. Atmospheric wet deposition of trace elements to forest ecosystem of the Qilian Mountains, Northwest China. Catena 197, 104966.

Zhang, J., Geng, H., Pan, B., Hu, X., Chen, L., Wang, W., Chen, D. and Zhao, Q. 2020a. Climatic zonation complicated the lithology controls on the mineralogy and geochemistry of fluvial sediments in the Heihe River basin, NE Tibetan Plateau. Quaternary International 537, 33-47.

Zhang, Y., Lu, Y., Zhou, Q. and Wu, F. 2020b. Optimal water allocation scheme based on trade-offs between economic and ecological water demands in the Heihe River Basin of Northwest China. Science of The Total Environment 703, 134958.

Zhou, Y., Wang, L., Xiao, T., Chen, Y., Beiyuan, J., She, J., Zhou, Y., Yin, M., Liu, J. and Liu, Y. 2020. Legacy of multiple heavy metal (loid) s contamination and ecological risks in farmland soils from a historical artisanal zinc smelting area. Science of the Total Environment 720, 137541.