

# Contamination characteristics and source apportionment of trace metals in soils around Miyun Reservoir

Haiyang Chen<sup>1</sup> · Yanguo Teng<sup>1</sup> · Ruihui Chen<sup>1</sup> · Jiao Li<sup>1</sup> · Jinsheng Wang<sup>1</sup>

Received: 4 November 2015 / Accepted: 12 April 2016 / Published online: 23 April 2016  
© Springer-Verlag Berlin Heidelberg 2016

**Abstract** Due to their toxicity and bioaccumulation, trace metals in soils can result in a wide range of toxic effects on animals, plants, microbes, and even humans. Recognizing the contamination characteristics of soil metals and especially apportioning their potential sources are the necessary preconditions for pollution prevention and control. Over the past decades, several receptor models have been developed for source apportionment. Among them, positive matrix factorization (PMF) has gained popularity and was recommended by the US Environmental Protection Agency as a general modeling tool. In this study, an extended chemometrics model, multivariate curve resolution-alternating least squares based on maximum likelihood principal component analysis (MCR-ALS/MLPCA), was proposed for source apportionment of soil metals and applied to identify the potential sources of trace metals in soils around Miyun Reservoir. Similar to PMF, the MCR-ALS/MLPCA model can incorporate measurement error information and non-negativity constraints in its calculation procedures. Model validation with synthetic dataset suggested that the MCR-ALS/MLPCA could extract acceptable recovered source profiles even considering relatively larger error levels. When applying to identify the sources of trace metals

in soils around Miyun Reservoir, the MCR-ALS/MLPCA model obtained the highly similar profiles with PMF. On the other hand, the assessment results of contamination status showed that the soils around reservoir were polluted by trace metals in slightly moderate degree but potentially posed acceptable risks to the public. Mining activities, fertilizers and agrochemicals, and atmospheric deposition were identified as the potential anthropogenic sources with contributions of 24.8, 14.6, and 13.3 %, respectively. In order to protect the drinking water source of Beijing, special attention should be paid to the metal inputs to soils from mining and agricultural activities.

**Keywords** Soil contamination · Trace metals · Source apportionment · Multivariate curve resolution-alternating least squares · Positive matrix factorization

## Introduction

Soils are generally regarded as being continuous recipients of trace metals. However, excessive metal inputs into soils may lead to the changes of the soil physicochemical properties, deterioration of the soil biology and function, and other environmental problems (Zhao et al. 2014). Especially, due to their toxicity, bioaccumulation, and resistance to biochemical degradation, trace metals can reside in soils for long periods and sequentially have the potential to damage microbiota, flora, and fauna once they have been transformed from solid form into ionic moieties or through biomethylation to organometallic moieties (Chen et al. 2015). More importantly, trace metals in soils can even threaten human health via food chain or by ingestion, inhalation, or dermal absorption. For example, previous studies showed that environmental exposure to soils contaminated by cadmium (Cd) and lead (Pb) would result in reduction of human life expectancy by 9~10 years

---

Responsible editor: Zhihong Xu

**Electronic supplementary material** The online version of this article (doi:10.1007/s11356-016-6694-1) contains supplementary material, which is available to authorized users.

✉ Haiyang Chen  
superbg@163.com

<sup>1</sup> Engineering Research Center of Groundwater Pollution Control and Remediation, Ministry of Education, College of Water Sciences, Beijing Normal University, No 19, Xijiekouwai Street, Beijing 100875, China

(Lăcătușu et al. 1996). Chervona et al. (2012) also reported that exposure to high levels of chromium (Cr) would cause respiratory toxicity, skin sensitization, and increased risks of some cancers. Over the past decades, soil contamination by trace metals has been caused more and more attention in the world (Nriagu 1990; Chen et al. 1999; Teng et al. 2014).

For soil pollution prevention and control, the correct discrimination of natural and anthropogenic sources, and the quantitative determination of contributions for metals in soils become more important even if better measurements are to be established. Over the past decades, several receptor models have been developed for source apportionment studies, such as the chemical mass balance (CMB) model, primary component analysis/multiple linear regressions (PCA/MLR), factor analysis (FA), Unmix, and positive matrix factorization (PMF) (Gordon 1988; Norris and Duvall 2014). Among them, PMF has gained popularity and has been recommended by the US Environmental Protection Agency (USEPA) as a general modeling tool for source apportionment studies (Paatero 1997; Norris and Duvall 2014). Compared with conventional multivariate models (i.e., PCA, FA), PMF was developed to cope with uncertainties and error propagation problems and to achieve more statistically sound maximum likelihood solution in the analysis of noisy environment data (Paatero and Tapper 1994; Paatero 1997, 1999). Due to its attractive features, PMF has been widely applied to identify pollution sources and apportion contributions in various environmental media including sediments, soils, and air particulates (Brinkman et al. 2006; Chen et al. 2013; Lang et al. 2015).

In recent years, a chemometrics model, multivariate curve resolution-alternating least squares method (MCR-ALS) has been proposed for qualitative and quantitative mixture analysis of multivariate datasets (Tauler 2007, Tauler et al 2009). With the MCR-ALS, the factor loading and score matrices can be obtained by minimizing the sum of squared residuals via the alternating least-squares algorithm with non-negativity constraints. However, the MCR-ALS algorithm obtain optimal solution under the assumption that the experimental data have an identical, independent, and normally distributed error structure, which cannot be made in general and is not satisfied by ambient measurement data (Wentzell et al. 1997; Tauler et al. 2009; Dadashi et al. 2012). To take into account the frequent existence of heteroscedastic and correlated error structures inherent in the environmental data in most cases, multivariate curve resolution weighted alternating least squares (MCR-WALS), and multivariate curve resolution-alternating least squares based on maximum likelihood principal component analysis (MCR-ALS/MLPCA), have been proposed for the analysis of receptor data (Wentzell et al. 1997; Tauler et al. 2009; Stanimirova et al. 2011; Dadashi et al. 2012, 2013). Since they avoid the propagation of errors to the parameters of the optimization, the two extensions of MCR-ALS that allow for the incorporation of the measurement uncertainty

information can provide more reliable results than unweighted ordinary MCR-ALS, especially in the case of the presence of high amounts of noise (Wentzell et al. 2006; Dadashi et al. 2012). Compared with MCR-WALS, the MCR-ALS/MLPCA is only used as a preliminary data pretreatment before MCR-ALS analysis and naturally does not require changing the traditional MCR-ALS algorithm, which also brings it the possible advantage than MCR-WALS for its easier application (Dadashi et al. 2013).

Constructed in 1960s, Miyun Reservoir is now the main raw water source for Beijing's domestic water supply. More than ten million people inhabiting the city of Beijing depend on Miyun Reservoir for drinking water. Protecting the surrounding ecosystem of Miyun Reservoir is of great urgency for ensuring water security for this vitally important and rapidly growing city (Wang et al. 2008). However, with drastically increased human activities and fast urban expansion over the past 30 years, Miyun Reservoir has been influenced, particularly during the 1980s, by agricultural and industrial development in the northern portions of the Beijing region (Luo et al. 2010; Su et al. 2014). Several studies investigated the concentration distribution of trace metals in soils of the Miyun Reservoir watershed and found that runoff from non-point sources, direct dumping of wastes, mineral exploitation, and pollutants carried by rivers had resulted in elevated concentrations of trace metals in soils of that region (Gao and Liao 2007; Luo et al. 2010; Huang et al. 2013). It is well known that the trace metals in soils can reach surface and ground water through runoff and leaching and directly or indirectly affect human health through the water supply and aquatic and terrestrial food chains (Schipper et al. 2008). Some trace metals such as Cd, Cr, Pb, mercury (Hg), nickel (Ni), and arsenic (As) have been detected in the water and sediment of Miyun Reservoir (Liu et al. 2005; Qiao et al. 2013).

To formulate effective management plans and policies to protect the drinking water reservoir, it is the basic preconditions to identify the contamination characteristics of its surrounding soils and especially apportion the potential anthropogenic sources. Information on the significance and extent of soil contamination from different sources is so important that appropriate actions can be effectively targeted to reduce metal inputs to soils (Marmur et al. 2005; Hu and Cheng 2013). Although several researches mentioned above investigated the concentration distribution of trace metals in soils of Miyun Reservoir watershed, few reports which focused on source apportionment of soil metals can be found.

In the present study, the extended chemometrics model, multivariate curve resolution-alternating least squares based on maximum likelihood principal component analysis (MCR-ALS/MLPCA), taking the advantages of the MCR-ALS and MLPCA, was proposed to identify the potential sources of trace metals in soils around Miyun Reservoir. Similar to PMF, the MCR-ALS/MLPCA algorithm can incorporate measurement uncertainty information inherent in environment data to account

for the structure and level of the noise. In addition, the solutions obtained from MCR-ALS/MLPCA obey non-negativity constraints of the source profiles, which make their interpretation physically meaningful. As a comparison, PMF was also employed to the same dataset of soil metals collected around Miyun Reservoir. The objective of this study is (1) to demonstrate the value of the MCR-ALS/MLPCA model for source apportionment of soil metals and (2) to apportion the potential sources of trace metals in soils around Miyun Reservoir. The results will provide policy and decision makers with a practical tool to identify the pollution sources of soil metals, enhancing their abilities to design effective regulatory measures and control strategies to protect soil ecosystem.

## Materials and methods

### Study area

Miyun Reservoir, situated in the mountainous area of Miyun County, Beijing (north latitude 40° 23', east longitude 116° 50') (Fig. 1), is the largest reservoir in north China. Now, it is the main drinking water source for Beijing, with a surface area of 188 km<sup>2</sup> and watershed area of 15,788 km<sup>2</sup>. Historically, some metal mines existed in the north areas of the Miyun Reservoir. The Baihe and Bai Maguan Rivers are its main sources of water. The area has a typical monsoon-influenced climate, characterized by hot, humid summers due to the East Asian monsoon and generally cold, windy, dry winters due to

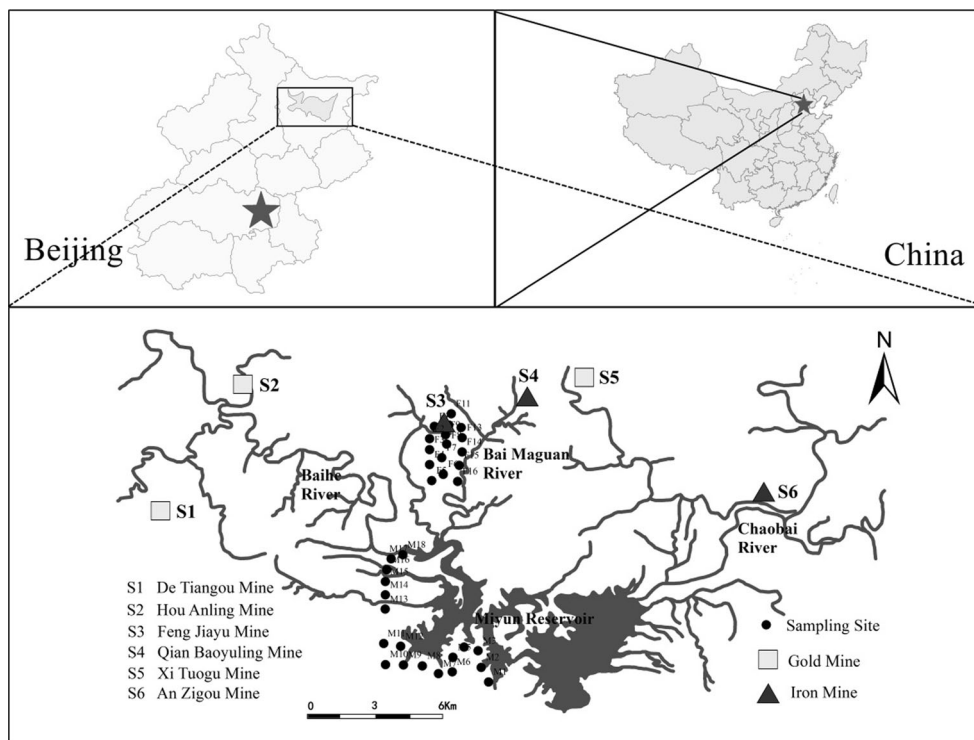
the vast Siberian anticyclone. Its annual temperature is about 11.5 °C, and the annual precipitation is about 600 mm. The main land uses in the area are forestry and farming which represent 50 and 15 % of the total land area, respectively. The soil types around Miyun Reservoir are mainly argosols, cambosols, and aridosols, namely, luvisols, alisols, cambisols, calcisols, and gypsisols, respectively (Luo et al. 2010).

### Sample collection and analysis

Surrounding Miyun Reservoir, a total of 34 surface soil samples were collected in May 2015. Figure 1 shows the location of the sampling sites. To evaluate the concentrations of trace metals in soils and their potential adverse effects to the reservoir, sample locations were most densely collected around the upper region of reservoir where the most human-impacted areas were located. Throughout the survey, a global positioning system was used to determine sampling positions. Approximately top 20 cm of surface soils were collected. After sampling, the soil samples were stored in sealed kraft packages to avoid contamination and transported to the laboratory immediately for further analysis. Preservation and transportation of the soil samples were performed based on the Technical Specification for Soil Environmental Monitoring of China (SEPAC 2004).

In the laboratory, soil samples were dried, and passed through a 100-mesh sieve, and then powdered and stored in acid washed and deionized water rinsed glass bottles prior to analysis. For content determinations, about 0.1 g subsamples

**Fig. 1** Study area and sampling sites of trace metals in soils around Miyun Reservoir



was subjected to a digestion solution with concentrated nitric acid and concentrated hydrochloric acid. After digestion using an electric digestion instrument, the sample solutions were filtered and adjusted to a suitable volume with double deionized water. The contents for trace metals were determined with inductively coupled plasma atomic emission spectrometry and inductively coupled plasma mass spectrometer. A necessary analytical quality control method was designed and followed during the analysis through careful standardization, procedural blank measurements, and spiked and duplicate samples. A total of 11 trace elements (Cd, Cr, As, Hg, Pb, Cu, Zn, Ni, Mn, Co, and V) were measured for each soil sample. Basic descriptive statistics were derived to provide a summary of the concentrations of these trace elements and presented in Table 1.

**Contamination and exposure risk assessment**

To assess the general contamination characteristics of trace metals in soils around Miyun Reservoir, the geoaccumulation index ( $I_{geo}$ ) was used with the equation:  $I_{geo} = \log_2(C_n/1.5B_n)$ , where  $C_n$  is the measured concentration of the element in soil (mg/kg),  $B_n$  is the geochemical background value of the corresponding element (mg/kg), the coefficient 1.5 is used to detect very small anthropogenic influences (Muller 1969). Figure 2 shows the boxplots of the  $I_{geo}$  for soil metals. Additionally, to identify the exposure risks posed by soil metals, the dose–response model recommended by USEPA (1989, 2001) was employed for health risk assessment. Three groups of population (adult males, adult females, and children) and three pathways (soil and dust ingestion, air inhalation, and dermal contact) were considered. Considering 1000 iterations, Monte-Carlo simulation which was done by programming an evaluation algorithm with MATLAB R2009b software was used to deal with the uncertainty caused by probabilistic parameters (Lonati et al. 2007; Chen et al. 2015). The probabilistic parameters were preferentially collected from the studies conducted in China. Those data unfilled by localized studies were derived from the USEPA guidelines and international agencies (see Table S1). Overview of health risk levels for the defined three population groups due to environmental exposure to soil metals were listed in Table 2.

**Source apportionment tool**

In the present study, the MCR-ALS/MLPCA model was proposed for source apportionment of trace metals in soils. The general principle of MCR-ALS/MLPCA method is to solve a bilinear problem which can be expressed mathematically in the following way (Tauler 2007):

$$\begin{cases} D = S \times R + E \\ S_{ik} \geq 0, R_{kj} \geq 0, k = 1, 2, \dots, p; i = 1, 2, \dots, m; j = 1, 2, \dots, n \end{cases} \quad (1)$$

**Table 1** Summary statistics of trace metals in soils around Miyun Reservoir (mg/kg)

	Cd	Cr	As	Hg	Pb	Cu	Zn	Ni	Mn	Co	V	Se
Range	0.11~0.28	36.4~123.5	1.1~8.5	0.03~0.06	14.0~31.9	18.6~38.9	53.3~84.7	18.4~37.8	384~812	8.8~15.2	65.7~104.6	0.01~0.3
Concentration <sup>a</sup>	0.26±0.12	65.2±252.4	7.7±4.9	0.05±0.05	22.9±41.7	31.7±27.3	82.4±22.1	32.6±19.6	725±328	13.6±6.0	100.0±29.9	0.315±0.115
C.V. <sup>b</sup>	0.4	2.0	0.6	0.9	1.3	0.7	0.3	0.5	0.4	0.4	0.3	0.4
Kurtosis	8.9	30.2	6.0	27.6	25.5	18.9	3.7	4.3	8.9	4.5	4.4	2.8
Skewness	2.0	5.3	1.5	4.9	4.7	3.7	1.0	1.6	2.1	1.3	1.1	-0.3
ABC <sup>c</sup>	0.067	66.1	8.7	0.031	23.6	22.9	105.7	30.7	682	17	89.7	0.187
Grade II <sup>d</sup>	0.6	300	25	0.5	300	100	250	50	-	-	-	-

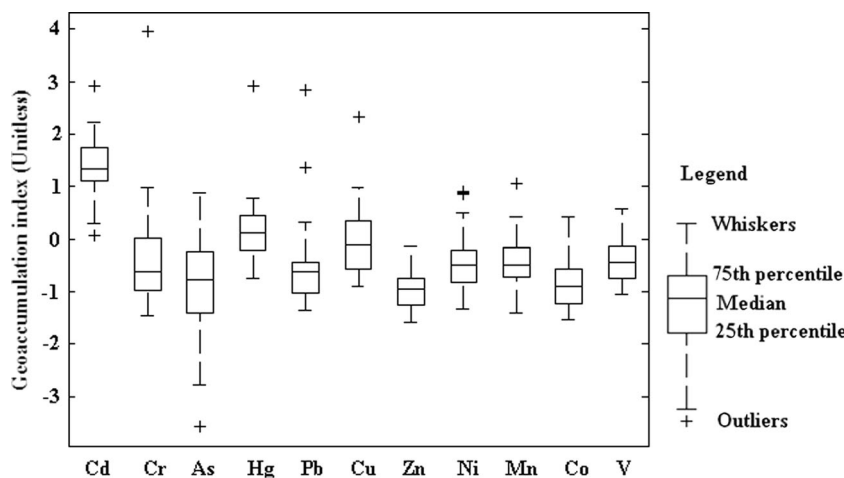
<sup>a</sup> Average ± standard deviation

<sup>b</sup> Coefficient of variation

<sup>c</sup> Average background concentration

<sup>d</sup> Chinese soil guidelines

**Fig. 2** Boxplots of the geoaccumulation index for trace metals in soils around Miyun Reservoir



where  $m$ ,  $n$ , and  $p$  are the number of samples, compounds, and sources, respectively; the matrix  $E$  is the residual matrix.

Owning the advantages of MLPCA (Wentzell et al. 1997), the MCR-ALS/MLPCA model can include uncertainty information in its curve resolution procedure with the respective maximum likelihood projections implemented in the steps of the alternating least-squares algorithm. It minimizes the weighted sum of residuals ( $Q_w$ ) which can be expressed mathematically in the following way (Dadashi et al. 2012):

$$\begin{cases} Q_w = \sum_{i=1}^m \sum_{j=1}^n (d_{ij} - \tilde{d}_{ij})^2 / \sigma_{ij}^2 \\ \{S, R\} = \arg \cdot \min_{S,R} Q_w, \quad s_{ik} \geq 0, \quad r_{ik} \geq 0, \quad k = 1, 2, \dots, p \end{cases} \quad (2)$$

Where  $d_{ij}$  is the measurement of the  $j$ th chemical component for the  $i$ th sample and  $\tilde{d}_{ij}$  is the prediction with the model given a complexity;  $\sigma_{ij}$  is the uncertainty associated with element  $d_{ij}$  of the data matrix  $D$ .

The basic algorithms for MCR-ALS/MLPCA are described in detail as follows. Initially, the number of significant factor and an initial estimation for matrix  $R$  or  $S$  must be provided. Due to its symmetry of the MCR-ALS algorithm, either of the factor loadings or factor scores can be chosen as the starting point. Herein, the factor loading matrix  $R$  is considered to be initialized since it will be smaller and follow a more systematic variation (Wentzell et al. 2006). Several methods, such as the random assignment with positive values (RAPV), random selection from the original data  $D$  (RSOD),

**Table 2** Boxplots of the geoaccumulation index for trace metals in soils around Miyun Reservoir

	Adult males				Adult females				Children			
	Mean	S.D.	Percentiles		Mean	S.D.	Percentiles		Mean	S.D.	Percentiles	
			50th	90th			50th	90th			50th	90th
Non-carcinogenic Hazard Index												
Cd	1.20E-02	4.90E-03	1.10E-02	1.80E-02	1.00E-02	4.50E-03	9.50E-03	1.60E-02	1.80E-02	7.70E-03	1.60E-02	2.70E-02
Cr	2.50E-01	5.00E-01	1.30E-01	3.20E-01	2.20E-01	4.60E-01	1.20E-01	2.90E-01	3.90E-01	8.00E-01	2.10E-01	5.10E-01
As	9.10E-03	5.30E-03	8.20E-03	1.30E-02	9.60E-03	5.60E-03	8.70E-03	1.40E-02	1.90E-02	1.10E-02	1.80E-02	2.80E-02
Hg	4.50E-04	3.90E-04	3.70E-04	5.80E-04	4.20E-04	3.60E-04	3.40E-04	5.40E-04	7.40E-04	6.30E-04	6.10E-04	9.50E-04
Pb	2.00E-03	2.70E-03	1.50E-03	2.80E-03	2.30E-03	3.00E-03	1.60E-03	3.10E-03	4.90E-03	6.40E-03	3.50E-03	6.70E-03
Cu	1.10E-03	7.40E-04	8.60E-04	1.60E-03	1.00E-03	7.00E-04	8.10E-04	1.50E-03	1.80E-03	1.30E-03	1.50E-03	2.80E-03
Zn	1.10E-04	2.90E-05	1.10E-04	1.50E-04	1.20E-04	3.00E-05	1.10E-04	1.60E-04	2.30E-04	6.00E-05	2.20E-04	3.10E-04
Ni	7.40E-03	3.90E-03	6.40E-03	1.60E-02	6.80E-03	3.50E-03	5.90E-03	1.50E-02	1.20E-02	6.20E-03	1.00E-02	2.60E-02
Carcinogenic risks												
Cd	1.00E-10	4.40E-11	9.50E-11	1.60E-10	1.20E-10	5.20E-11	1.10E-10	1.90E-10	1.10E-10	4.80E-11	1.00E-10	1.70E-10
Cr	7.20E-06	1.50E-05	3.80E-06	9.50E-06	8.40E-06	1.70E-05	4.40E-06	1.10E-05	8.20E-06	1.70E-05	4.30E-06	1.10E-05
As	4.40E-06	2.60E-06	4.00E-06	6.40E-06	4.20E-06	2.50E-06	3.90E-06	6.10E-06	3.00E-06	1.70E-06	2.70E-06	4.30E-06
Pb	3.00E-08	4.00E-08	2.20E-08	4.20E-08	3.50E-08	4.60E-08	2.50E-08	4.80E-08	3.40E-08	4.50E-08	2.50E-08	4.70E-08
Ni	1.90E-09	9.70E-10	1.60E-09	4.10E-09	2.20E-09	1.10E-09	1.90E-09	4.90E-09	2.00E-09	1.10E-09	1.70E-09	4.40E-09



and simple to use interactive self-modeling analysis (SIMPLISMA), can be used to initially defined the elements  $S$  (Windig and Guilment 1991; Stanimirova et al. 2011).

Second, the estimation of  $D^*$  in the subspace defined by its principal components, which can be obtained using MLPCA, is then given by the maximum likelihood projection of  $D$ , as described by Eq. (3) (Wentzell et al. 2006). In contrast to conventional principal component analysis, MLPCA incorporates uncertainty information into the decomposition process and can deal with heteroscedastic errors and correlated error structures (Wentzell et al. 1997).

$$d_i^* = d_i \cdot \Sigma_i^{-1} V_p \left( V_p^T \Sigma_i^{-1} V_p \right)^{-1} V_p^T \tag{3}$$

Where,  $d_i$  represents the  $i$ th row vector of  $D$ ;  $\Sigma_i$  is the error covariance matrix for  $i$ th row;  $V_p$  is the derived loadings matrix for the  $p$  component either by PCA (once at the beginning of the optimization) or MLPCA (at each ALS iteration of the optimization) (Dadashi et al. 2012);  $D^*$  is the projection of the original dataset onto the loadings subspace.

After data projection, the factor score matrix  $S$  can be obtained given loading matrix  $R$  using the following equations described by Eq. (4) to solve the non-negative least-squares problem. In order to limit the problem of the rotational ambiguity of the solutions, non-negativity constraints were imposed to matrix  $S$ . The simple method is to replace negative elements with zero or an extremely small value (i.e.,  $10^{-12}$ ). Other approaches, such as the non-negativity least-squares (NNLS), fast non-negativity least-squares algorithm (FNNLS) can also be used (Van and Keenan 2004).

$$\begin{cases} \{S\} = \arg \cdot \min_S \|D^* - SR\|, & S_{kj} \geq 0, \quad k = 1, 2, \dots, p \\ S = D^* R^T (RR^T)^{-1} \end{cases} \tag{4}$$

Following this step, the estimation obtained for  $S$  is then used to re-estimate the profiles matrix  $R$  under the non-negativity constraint in the following way:

$$\begin{cases} \{R\} = \arg \cdot \min_R \|D^* - SR\|, & S_{ik} \geq 0, \quad k = 1, 2, \dots, p \\ R = (S^T S)^{-1} S^T D^* \end{cases} \tag{5}$$

The iterative procedure described by Eqs. (3)–(5) is repeated until convergence, which can be tested by checking for the insignificant changes in  $S$  and/or  $R$ , or the difference in the  $Q_w$  values obtained in two consecutive iterations.

As a comparison, PMF was also applied to the same dataset of soil metals. Similar to MCR-ALS/MLPCA, the PMF model also use a weighted least-squares fit with the known error estimates of the elements to minimize the  $Q_w$  (Paatero and Tapper 1994). Over the past years, the bilinear PMF2 and multilinear engine (ME-2) algorithms have been developed to solve the minimization problem described by Eq. (2)

(Paatero and Tapper 1994; Paatero 1999). Here, the latest version of EPA PMF program, EPA-PMF5.0, which is based on the ME-2 engine, was comparably employed for source apportionment of trace metals in soils around Miyun Reservoir. Further details of applications of EPA PMF5.0 can be found elsewhere (Norris and Duvall 2014).

### Synthetic dataset development

Before applying to the dataset of soil metals, the capability of the MCR-ALS/MLPCA model was investigated using simulated dataset which was generated in the case of noisy interference, considering six hypothetical environmental sources ( $p=6$ ). The simulated dataset contained 15 variables ( $J=15$ ) and 50 samples ( $I=50$ ). The initial source profiles ( $R_0$  and  $S_0$ ) were generated from a log normal distribution of random numbers (mean, 0.01 and standard deviation, 1). Previous studies showed that the shapes of the initially generated profiles in this way were very similar to those obtained in environmental data in which scores for only some of the samples and loadings for some of the variables were much larger than for the others (Dadashi et al. 2012). Sequentially, the error-free data matrix  $D_0$  was generated by multiplying the initial distribution profile matrix  $S_0$  ( $50 \times 6$ ) by the composition profile matrix  $R_0$  ( $6 \times 15$ ). Considering the most general case of error structure for environment data, error matrix with constant and proportional uncorrelated errors were generated and added to the error-free datasets  $D_0$  to generate simulated dataset  $D$ . The elements of error matrix of standard deviations,  $\sigma_{ij}$ , were obtained as the square root of the sum of the squares of the constant part,  $a$ , and the proportional part as (Stanimirova et al. 2011):

$$\sigma_{ij} = \sqrt{a^2 + b^2 d_{ij}^2} \tag{6}$$

where  $b$  is the proportional coefficient. In this simulated case, different noise levels were considered and compared with a dataset of parameters. The constant part was taken to be 1 and 5 % of the maximum value of the error-free data, while the proportional part was taken as 5, 10, 15, 20, 25 and 30 % of the elements in dataset  $D_0$ . Thus, a total of 12 datasets with different noise levels were calculated. The final error matrix in each case was found by multiplying element by element the matrix of normally distributed numbers by the matrix of standard deviations, and was then added to the error-free datasets giving the 12 dataset configurations.

The validation analysis was performed applying non-negativity constraint with FNNLS algorithm (Van and Keenan 2004). The initial estimate of loading matrix  $S$  was obtained from the purest variables using the SIMPLISMA method in all cases (Windig and Guilment 1991). All calculations, according to the implemented routines, were performed

with MATLAB R2009b software on a personal computer (Intel(R) Core (TM) 2 Duo CPU P8700, 2.53 GHz with 4 GB RAM) using the Microsoft Windows XP (service pack 3) operating system.

## Results and discussion

### Validation results of MCR-ALS/MLPCA using synthetic datasets

The MCR-ALS/MLPCA model was validated using the simulated dataset. To check and evaluate the performance of the apportionment tool, the lack of fit (lof) and explained variance ( $R^2$ ) were calculated (Tauler et al. 2009). Additionally, evaluation of the quality in the recovery of source profiles is possible because the true profiles are known beforehand. Therefore, the comparable pairwise correlation coefficients ( $r^2$ ) were also calculated between the theoretical profiles and the model results with different levels of measurement errors (Dadashi et al. 2013).

It can be seen from Table 3 that the MCR-ALS/MLPCA model reproduced each source profile well. For the lower error levels with constant noise of 1 %, the comparison between theoretical and MCR-ALS/MLPCA results shows a difference in the calculated lack of fit values from 5.8 to 23.3 %. On the other hand, for the higher levels of constant noise (5 %), the relevant values for lof and  $R^2$  were 18.9~25.3 and 98.20~96.75 %, respectively. The pairwise correlation coefficients calculated between the corresponding profiles obtained from MCR-ALS/MLPCA and the hypothetical true profiles also supported the conclusion that the fitted profiles were resolved for synthetic dataset with constant and proportional measurement errors ( $r^2 > 0.8$ ). Generally, if the results of the correlation coefficient are close to 1, this indicates that there is

a strong linear relationship between the profiles (Dadashi et al. 2012). On contrast, if it is zero, this means that there is no linear relationship between them. Previous studies suggested that correlation coefficients more than 0.80 would represent acceptable precision for an individual model analysis (Terrado et al. 2009).

However, as was expected, the profiles were recovered with higher lack of fit when data had higher levels of measurement errors (Table 3). It was also not surprising that a higher noise would imply greater difficulty in recovering the true profiles with lower pairwise correlation coefficients. Overall, according to the validation results shown in Table 3, it could be found that the MCR-ALS/MLPCA model might obtain acceptable recovered profiles even when the constant noise was 5 % of the maximum value and the proportional part was up to 25 % of the elements in the error-free data matrix.

### General characteristics of trace metals in soils

Table 1 lists the concentration range, standard deviation, skew, and coefficient of variation of trace metals in soils around Miyun Reservoir. To facilitate the evaluation and comparison, the average background concentrations (ABCs) of soils in Beijing and the soil quality standards of China are also presented in Table 1. As shown in Table 1, the average concentrations of all analyzed metals except for As, Zn, and Co are higher than their corresponding ABCs (CNEMC 1990). Especially, the pollution indices by comparing median concentration of Cd, Hg, and Cu with respect to relevant background value were 3.8, 1.6, and 1.4, respectively, indicating moderate contamination. In the soil quality standards of China (Chinese Environmental Protection Administration (CEPA) 1995), the grade II level can be used as the threshold values for protecting human health. In this study, approximately 15 % samples for Ni exceeded their corresponding grade II

**Table 3** Validation results for synthetic datasets using the MCR-ALS/MLPCA model with different levels of measurement errors

Simulation run	Noise levels		lof%	$R^2$ %	Pairwise correlation coefficients ( $r$ )					
	a (%)	b (%)			s1	s2	s3	s4	s5	s6
1	1	5	5.8	99.83	0.937	0.997	0.999	0.989	0.977	0.989
2	1	10	8.9	99.60	0.920	0.989	0.998	0.983	0.989	0.982
3	1	15	12.6	99.20	0.901	0.977	0.995	0.956	0.991	0.952
4	1	20	16.2	98.68	0.907	0.960	0.994	0.939	0.986	0.921
5	1	25	19.8	98.02	0.861	0.993	0.928	0.951	0.967	0.896
6	1	30	23.3	97.25	0.868	0.984	0.922	0.947	0.970	0.879
7	5	5	18.9	98.20	0.917	0.948	0.912	0.939	0.932	0.900
8	5	10	19.6	98.06	0.853	0.923	0.891	0.901	0.969	0.904
9	5	15	21.1	97.75	0.812	0.909	0.834	0.937	0.863	0.919
10	5	20	22.6	97.41	0.889	0.858	0.826	0.957	0.813	0.859
11	5	25	24.0	97.08	0.803	0.847	0.841	0.858	0.851	0.839
12	5	30	25.3	96.75	0.804	0.815	0.858	0.804	0.742	0.843

values. The results showed that the soils in the study area had been influenced by trace metals in varied degree, which was consistent with the previous studies where mining activities and agricultural application had caused heavy metal contamination in the soils around Miyun Reservoir (Luo et al. 2009; Huang et al. 2013). Meanwhile, it can also be seen from Table 1 that the elements Cr, Hg, Pb, Cu, and Ni had relatively wide concentration ranges and high coefficients of variation (>50 %), which suggested that these metal inputs to the soils would be attributable to anthropogenic sources.

To further identify the impact of anthropogenic activities, the  $I_{geo}$  for each element in every sample was considered. Generally, the soils can be classified as unpolluted ( $I_{geo} \leq 0$ ), unpolluted to moderately polluted ( $0 < I_{geo} \leq 1$ ), moderately polluted ( $1 < I_{geo} \leq 2$ ), moderately to heavily polluted ( $2 < I_{geo} \leq 3$ ), heavily polluted ( $3 < I_{geo} \leq 4$ ), heavily to extremely polluted ( $4 < I_{geo} \leq 5$ ), or extremely polluted ( $I_{geo} > 5$ ) (Muller 1969). As shown in Fig. 2, the  $I_{geo}$  values of most elements varied the most, mainly ranging from unpolluted level to moderately polluted level, indicating human activities. In particular, it should be noticed that the  $I_{geo}$  values for Cd in 85 % samples lay above the moderately contaminated level. High Cd concentration could be associated with industrial activities and anthropogenic wastes, including industrial discharges, sewage sludge, and municipal solid waste (Huang et al. 2013; Zhao et al. 2014).

### Exposure risks of soil metals

As shown in Table 2, overall, the exposure risks posed by trace metals in soils to human health were acceptable. According to USEPA (1989), carcinogenic risks surpassing  $1 \times 10^{-4}$  are viewed as unacceptable. Herein, the average carcinogenic risk values of Cr, As, Cd, Pb, and Ni for all three populations were less than  $1 \times 10^{-5}$ , and the average total carcinogenic risk index which was estimated by summing the individual carcinogenic risk index for Cd, Cr, As, Pb, and Ni were below  $1 \times 10^{-4}$ . Similarly, the soil metals potentially posed low non-carcinogenic risks to the public. For the total exposure hazard index which was calculated by summing the individual non-carcinogenic hazard index for all metals applied to the three populations, the proportion of soils whose hazard index lay between 0.1 and 0.5 was higher than 95 %, suggesting the exposed individual is unlikely to experience obvious adverse health effects (USEPA 2001).

Comparatively speaking, people were most exposed to Cr and Cd because of their high concentrations in soils or low reference dose values. For instance, the average hazard indices of Cr and Cd through all three exposure pathways for children respectively accounted for 87.3, and 4.0 % of the total exposure hazard index (TEHI). Similarly, for adult males and females, the total average hazard indices of Cr and Cd through three pathways accounted for 92.9 and 91.9 % of the TEHI

value, respectively. From Table 2, it should be noticed that the carcinogenic risks for children was less than that for adults. However, due to the behavioral and physiological characteristics of hand-to-mouth activities for soils, higher respiration rates per unit body weight, and increased gastrointestinal absorption of some substances, children have a higher susceptibility of non-carcinogenic exposure risks to soil metals per unit body weight than adults (Chen et al. 2015).

### Source apportionment of trace metals in soils

In the present study, the validated MCR-ALS/MLPCA model was employed to identify the potential sources of soil metals and apportion their contributions. Information about the measurement uncertainty was incorporated during the analyses of MCR-ALS/MLPCA algorithm. To determinate the significant number of factors, the cumulative percent variance, and the weighted sum of squares of differences between calculated and measured data,  $Q_w$ , were considered. In a common case, the  $Q_w$  value should approximately equal the number of degrees of freedom,  $df = m \times n - p \times (n + m)$  for a good fit (Bzdusek et al. 2006). Herein, with PCA, two factors would explain more than 85 % of the variances. However, the weighted sum of square residuals for four factors was approximately equal the number of degrees of freedom. Meanwhile, for the MCR-ALS/MLPCA model, a relatively large reduction for the  $Q_w$  value was not observed with complexities larger than four. Therefore, four factors, which explained 91.5 % of the variances were further considered in this study.

### Comparison of source profiles generated by the MCR-ALS/MLPCA and PMF

To facilitate the evaluation, the EPA PMF program (v5.0) was comparatively applied to the same dataset of trace metals in soils around Miyun Reservoir. According to the operation guides of EPA PMF software, before applying apportionment model to the dataset, individual sample was inspected and those with unusually high or low concentrations were identified as possible outliers (Marmur et al. 2005). Concentration values below the detection limit were substituted by half of the detection limit as recommended in the literature. Additionally, it should be noticed that the selection of the matrix initialization can influence the final solution of PMF and MCR-ALS/MLPCA, in part because their solutions may represent only one of a range of feasible solutions (Wentzell et al. 2006). Therefore, in order to guarantee the optimality of the solution, the global minima of PMF solution was tested 20 times using different seeds for pseudo-random initial values, and the MCR-ALS/MLPCA algorithm has also been run 20 times from different initial estimates of the composition profiles which were obtained from the purest variables using the SIMPLISMA (Windig and Guilment 1991) method at each



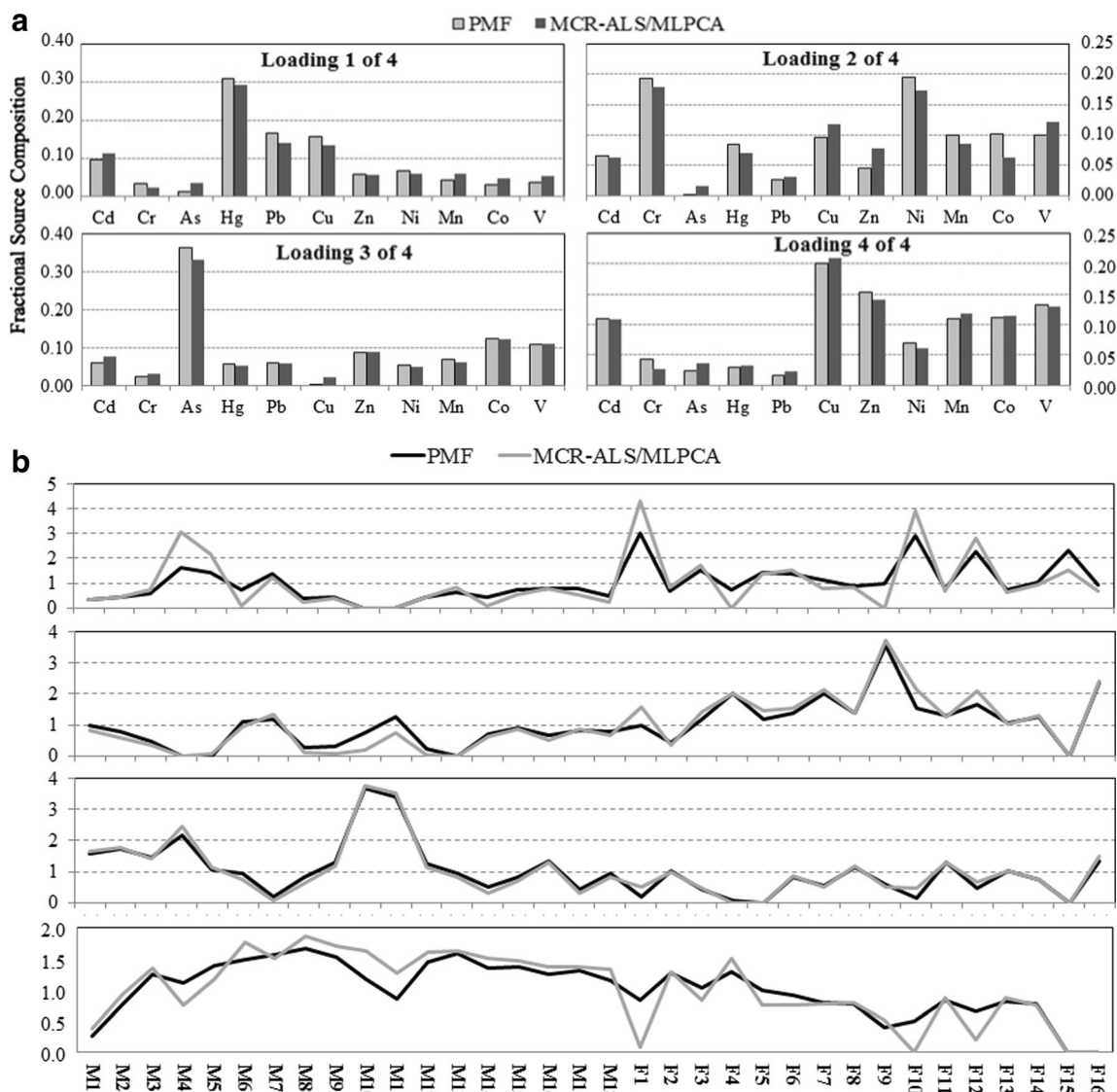
rerun. The final solution of PMF and MCR-ALS/MLPCA was the one with the lowest value of the respective objective function.

Source profiles for the trace metals in soils around Miyun Reservoir were generated using PMF and MCR-ALS/MLPCA with complexity four and presented in Fig. 3. From the qualitative viewpoint, it can be concluded that most of the data in the figure showed the relatively highly overlapping profiles, with little differences among them. To further evaluate their similarities, pairwise correlation coefficients between resolved profiles obtained from the two models were calculated and listed in Table 4. It can be seen that the pairwise correlation coefficients between source profiles obtained by PMF and MCR-ALS/MLPCA were higher than 0.90. Therefore, the probable conclusion was that the two methods, MCR-ALS/MLPCA and PMF which are based on completely different

algorithm, extracted the similar source profiles for the studied environmental data. The results also indirectly suggested the performance of MCR-ALS/MLPCA as a potential apportionment tool. For conciseness reasons, the MCR-ALS/MLPCA solution was further discussed and interpreted here.

*Source identification of trace metals in soils*

As shown in Fig. 3a, the first factor (loading 1 of 4) that accounted for 68.2 % of the total variance was predominantly loaded on Hg and Pb and slightly loaded on Cu and Cd. In Beijing area, high concentrations for Hg could be associated with coal combustion (Chen et al. 2015). The released amount of Hg from coal combustion cannot be ignored since coal consumption around Beijing-Tianjin-Hebei region has increased significantly over the past several decades (Jiang



**Fig. 3** Loading profiles (a) and contribution profiles (b) of trace metals in soils around Miyun Reservoir derived from the PMF and MCR-ALS/MLPCA models

**Table 4** Pairwise correlation coefficients between source profiles of trace metals in soils around Miyun Reservoir derived from the PMF and MCR-ALS/MLPCA models

PMF vs. MCR-ALS/MLPCA	Loading profiles				Contribution profiles			
	1	2	3	4	1	2	3	4
Factor 1	0.924	0.520	0.371	0.197	0.913	0.354	0.026	0.100
Factor 2	0.469	0.944	0.291	0.424	0.356	0.915	0.631	0.302
Factor 3	0.080	0.277	0.959	0.152	0.185	0.384	0.937	0.465
Factor 4	0.224	0.321	0.183	0.913	0.272	0.005	0.662	0.906

et al. 2006). On the other hand, as has been found in previous studies in other areas, Pb and Cd are likely to have mainly come from anthropogenic sources of vehicular traffic (Kadi 2009; Zhao et al. 2014). Although leaded gasoline has not been used in Beijing since 1997, soils could act as a reservoir for lead pollution over the years. Therefore, the first factor may be an anthropogenic component due to the industry discharge, coal combustion, and traffic emission through atmospheric deposition.

The second factor (loading 2 of 4) that was responsible for 17.7 % of the total variances was heavily weighted on Ni and Cr and slightly loaded on Cu and V, which was also likely due to anthropogenic influences. Previous studies showed that the significant human activities that affect concentrations of Cr and Ni in soils around Miyun Reservoir were primarily iron ore mining activities (Xue et al. 2000). According to Luo et al. (2010), mining activities in the northeast areas of the Miyun Reservoir have caused serious Cr contamination and the elevated concentrations of Ni in the soils. In Miyun Country, the iron ore deposit is about  $9 \times 10^9$  tons, accounting for 98 % of its total deposit in Beijing. The mining activities, such as iron, chromite, and gold mines, were historically existing in the north areas of the Miyun Reservoir (Liu et al. 2005) and presented in Fig. 1. Thus, it was reasonable to assign this factor to mining operations.

The third factor (loading 3 of 4) was predominantly loaded on As and Co. However, the mean and median concentrations of As and Co were not very high, indicating that the element was probably predominantly presented at the natural background concentrations controlled by the parent material (Table 1). Approximately 60.0 % of samples for As and 70.6 % of samples for Co were below their corresponding background concentrations. Meanwhile, about 94.1 % of samples for As and 91.2 % of samples for Co were identified as uncontaminated using geoaccumulation index (Fig. 2). It should be noticed that high concentrations for As could be associated with industrial activities and coal combustion (Luo et al. 2012). Thus, the third factor might be a mix source.

The fourth factor (loading 4 of 4) was only responsible for 3.6 % of the total variances and was heavily weighted on Cu and Zn and slightly loaded on Cd, Co, Mn, and V. Previous studies showed that the applications of fertilizers, pesticides, and fungicides for agricultural practices had increased the

concentrations of Zn and Cu in soils (Luo et al. 2009). According to statistics, a total input of  $\sim 1.4 \times 10^5$  tons of fertilizer was estimated to be applied as agrochemical products to agricultural land in Beijing annually. The applications of metal-containing fungicides in orchards which were historically distributed to the east of the reservoir would result in increased Cu content in soils. Additionally, the animal feeding is usually regarded as the other major source of Zn, Cu, and Cd in surface soils (Wu et al. 2010). According to Liu et al. (2005), there was quite a scale of livestock and poultry farming in the reservoir's upstream watershed and the farms were concentrated in the surrounding villages of the reservoir. Therefore, the fourth factor may be related to agricultural activities due to livestock farming, fertilizers, and agrochemical application.

#### Source contributions

Figure 3b shows the obtained contribution profiles. It can be seen that the soil metals in sites F1~F16 were mainly determined by mining operation and coal combustion. The apportionment results were compatible with that area's environmental data. Historically, there existed an iron mine named as Feng JiaYu Mine in the sampling area (Fig. 1). The main component of iron mine tailings was ferrous chromite. Although most of the mining occurred in a closed pit, the process involving the depositing of many tailings was the most likely cause of elevated contaminations for Cr and Ni in soils (Huang et al. 2013). The results also indicated that it was unworthy to ignore the influence of human activities for trace metals in the study area due to industrialization in the past decades. Based on the average of the individual percent contributions, the overall percent contribution from each source was calculated. Mining activities occupied the largest contribution (24.8 %), followed by agricultural activities (14.6 %) and atmospheric deposition (13.3 %). The metals released from different sources would transport to the reservoir and directly or indirectly impacted the water quality and ecologic system of that region to a considerable extent. To protect the drinking water reservoir of Beijing, agricultural and industrial activities in this area should be strictly limited and regulated.

## Conclusions

To identify the contamination characteristics of trace metals in soils around Miyun Reservoir of Beijing, in this study, an extended chemometrics model of MCR-ALS/MLPCA was proposed for source apportionment. Model validation with synthetic dataset showed that the MCR-ALS/MLPCA could obtain acceptable results even considering relatively larger error levels. When applying to identify measurement data, it extracted similar source profiles with PMF, albeit they were based on completely different algorithms. In addition, compared with PMF, MCR-ALS/MLPCA is a relatively simpler algorithm that only requires the optimization of fewer parameters. Therefore, MCR-ALS/MLPCA will be an optional method to be employed for source apportionment of trace metals in soils. Overall, the soils around Miyun Reservoir were contaminated by trace metals in moderate degree. Mining activities, atmospheric deposition, and fertilizers and agrochemicals were apportioned as the potential anthropogenic sources determining the contents of trace metals in soils with contributions of 24.8, 13.3, and 14.6 %, respectively. If the water quality of the Miyun Reservoir is to be protected, concentrations of Cd, Cr, Pb, Hg, Cu, and Ni in its surrounding soils should not be allowed to increase further.

**Acknowledgments** This study was financially supported by the National Natural Science Foundation of China (no. 41303069), the Fundamental Research Funds for the Central Universities (no. 105578GK), and Major Science and Technology Program for Water Pollution Control and Treatment (2014ZX07201-010).

## References

Brinkman G, Vance G, Hannigan MP, Milford JB (2006) Use of synthetic data to evaluate positive matrix factorization as a source apportionment tool for PM<sub>2.5</sub> exposure data. *Environ Sci Technol* 40(6):1892–1901

Bzdusek PA, Christensen ER, Lee CM, Pakdeesasuk U, Freedman DL (2006) PCB congeners and dechlorination in sediments of Lake Hartwell, South Carolina, determined from cores collected in 1987 and 1998. *Environ Sci Technol* 40:109–119

Chen HM, Zheng CR, Tu C, Zhu YG (1999) Heavy metal pollution in soils in China: status and countermeasures. *Ambio* 28:130–134

Chen HY, Teng YG, Wang JS (2013) Source apportionment of trace element pollution in surface sediments using positive matrix factorization combined support vector machines: application to the Jinjiang River, China. *Biol Trace Elem Res* 151(3):462–470

Chen HY, Teng YG, Lu S, Wang Y, Wang J (2015) Contamination features and health risk of soil heavy metals in China. *Sci Total Environ* 512–523:143–1561

Chervona Y, Arita A, Costa M (2012) Carcinogenic metals and the epigenome: understanding the effect of nickel, arsenic, and chromium. *Metallomics* 4(7):619–627

China National Environmental Monitoring Center (CNEMC) (1990) The soil background value in China [R]. China Environmental Science Press, Beijing

Chinese Environmental Protection Administration (CEPA) (1995) Environmental quality standard for soils (GB15618-1995) [R]. CEPA, Beijing

Dadashi M, Abdollahi H, Tauler R (2012) Maximum likelihood principal component analysis as initial projection step in multivariate curve resolution analysis of noisy data. *Chemometr Intell Lab Syst* 118:33–40

Dadashi M, Abdollahi H, Tauler R (2013) Application of maximum likelihood multivariate curve resolution to noisy data sets. *J Chem Theory Comput* 27:34–41

Gao ZD, Liao HJ (2007) Investigation and assessment on pollution of soil heavy metals in up-stream of Miyun Reservoir. *China Environ Protection Indus* 8:23–26

Gordon GE (1988) Receptor models. *Environ Sci Technol* 22:1132–1142

Hu Y, Cheng H (2013) Application of stochastic models in identification and apportionment of heavy metal pollution sources in the surface soils of a large-scale region. *Environ Sci Technol* 47:3752–3760

Huang XX, Zhu Y, Ji HB (2013) Distribution, speciation, and risk assessment of selected metals in the gold and iron mine soils of the catchment area of Miyun Reservoir, Beijing, China. *Environ Monit Assess* 185:8525–8545

Jiang GB, Shi JB, Feng XB (2006) Mercury pollution in China. *Environ Sci Technol* 40:3672–3678

Kadi MW (2009) Soil pollution hazardous to environment: a case study on the chemical composition and correlation to automobile traffic of the roadside soil of Jeddah city, Saudi Arabia. *J Hazard Mater* 168:1280–1283

Lăcătușu R, Răuță C, Cârstea S, Ghelase I (1996) Soil–plant–man relationships in heavy metal polluted areas in Romania. *Appl Geochem* 11:105–107

Lang YH, Li GL, Wang XM, Peng P, Bai J (2015) Combination of Unmix and positive matrix factorization model identifying contributions to carcinogenicity and mutagenicity for polycyclic aromatic hydrocarbons sources in Liaohe delta reed wetland soils, China. *Chemosphere* 120:431–437

Liu X, Xu Q, Ge X, Liu L, Wu D (2005) Speciation analysis of metals in the sediments of Miyun Reservoir. *Sci China Ser D-Earth Sci* 35(Suppl):288–295

Lonati G, Cernuschi S, Giugliano M, Grosso M (2007) Health risk analysis of PCDD/F emissions from MSW incineration: comparison of probabilistic and deterministic approaches. *Chemosphere* 67:S334–S343

Luo L, Ma YB, Zhang SZ, Wei DP, Zhu YG (2009) An inventory of trace element inputs to agricultural soils in China. *J Environ Manage* 90:2524–2530

Luo W, Lu YL, Zhang Y, Fu WY, Wang B, Jiao WT, Wang G, Tong XJ, Giesy JP (2010) Watershed-scale assessment of arsenic and metal contamination in the surface soils surrounding Miyun Reservoir, Beijing, China. *J Environ Manage* 91:2599–2607

Luo XS, Yu S, Zhu YG, Li XD (2012) Trace metal contamination in urban soils of China. *Sci Total Environ* 421–422:17–30

Marmur A, Unal A, Mulholland JA, Russell AG (2005) Optimization-based source apportionment of PM<sub>2.5</sub> incorporating gas-to-particle ratios. *Environ Sci Technol* 39:3245–3254

Muller G (1969) Index of geoaccumulation in sediments of the Rhine River. *Geophys J R Astron Soc* 2:108–118

Norris G, Duvall R (2014) EPA positive matrix factorization (PMF) 5.0 fundamentals & user guide [R]. US Environmental Protection Agency, Washington, DC

Nriagu JO (1990) A history of global metal pollution. *Science* 272:223–224

Paatero P (1997) Least squares formulation of robust nonnegative factor analysis. *Chemometr Intell Lab Syst* 37:23–35

Paatero P (1999) The multilinear engine-A table-driven, least squares program for solving multilinear problems, including the n-way parallel factor analysis model. *J Comput Graph Stat* 8:854–888

- Paatero P, Tapper U (1994) Positive matrix factorization: a nonnegative factor model with optimal utilization of error estimates of data values. *Environmetrics* 5:111–126
- Qiao MM, Ji HB, Zhu XF, Chen Y (2013) Fraction distribution and risk assessment of heavy metal in sediments of Miyun Reservoir. *J Agro-Environ Sci* 7:1–8
- Schipper PNM, Bonten LTC, Plette ACC, Moolenaar SW (2008) Measures to diminish leaching of heavy metals to surface waters from agricultural soils. *Desalination* 226:89–96
- Stanimirova I, Tauler R, Walczak B (2011) A comparison of positive matrix factorization and the weighted multivariate curve resolution method. Application to environmental data. *Environ Sci Technol* 45:10102–10110
- State Environmental Protection Administration of China (SEPA) (2004) The technical specification for soil environmental monitoring (HJ/T 166-2004) [R]. The Industrial Standard of The People's Republic of China Environmental Protection, Beijing
- Su M, Yu JW, Pan SL, An W, Yang M (2014) Spatial and temporal variations of two cyanobacteria in the mesotrophic Miyun Reservoir, China. *J Environ Sci* 26:289–298
- Tauler R (2007) Application of non-linear optimization methods to the estimation of multivariate curve resolution solutions and of their feasible band boundaries in the investigation of two chemical and environmental simulated data sets. *Anal Chim Acta* 595:289–298
- Tauler R, Viana M, Querol X, Alastuey A, Flight RM, Wentzell PD, Hopke PK (2009) Comparison of the results obtained by four receptor modelling methods in aerosol source apportionment studies. *Atmos Environ* 43:3989–3997
- Teng YG, Wu J, Lu SJ, Wang YY, Jiao XD, Song LT (2014) Soil and soil environmental quality monitoring in China: a review. *Environ Int* 69:177–199
- Terrado M, Barceló D, Tauler R (2009) Quality assessment of the multivariate curve resolution alternating least squares (MCR-ALS) method for the investigation of environmental pollution patterns in surface water. *Environ Sci Technol* 43:5321–5326
- USEPA (1989) Risk assessment guidance for Superfund. Human health evaluation manual, (part A) [R], vol 1. Office of emergency and remedial response, Washington, DC [EPA/540/1-89/002]
- USEPA (2001) Supplemental guidance for developing soil screening levels for superfund sites [R]. Office of Solid Waste and Emergency Response, Washington, DC [OSWER9355.4-24]
- Van BM, Keenan MR (2004) Fast algorithm for the solution of large-scale non-negativity-constrained least squares problems. *J Chemom* 18:441–450
- Wang ZJ, Hong JM, Du GS (2008) Use of satellite imagery to assess the trophic state of Miyun Reservoir, Beijing, China. *Environ Pollut* 155:13–19
- Wentzell PD, Andrews DT, Hamilton DC, Faber K, Kowalski BR (1997) Maximum likelihood principal component analysis. *J Chemom* 11(4):339–366
- Wentzell PD, Karakach TK, Roy S, Martinez MJ, Allen CP, Werner-Washburne M (2006) Multivariate curve resolution of time course microarray data. *BMC Bioinformatic* 7:343–353
- Windig W, Guilment J (1991) Interactive self-modeling mixture analysis. *Anal Chem* 63:1425–1432
- Wu S, Xia XH, Lin CY, Chen X, Zhou CH (2010) Levels of arsenic and heavy metals in the rural soils of Beijing and their changes over the last two decades (1985–2008). *J Hazard Mater* 179:860–868
- Xue HB, Sigg L, Gächter R (2000) Transport of Cu, Zn and Cd in a small agricultural catchment. *Water Res* 34:2558–2568
- Zhao L, Xu YF, Hou H, Shangguan YX, Li FS (2014) Source identification and health risk assessment of metals in urban soils around the Tanggu chemical industrial district, Tianjin, China. *Sci Total Environ* 468–469:654–662