Research article

GEEA

Geochemistry: Exploration, Environment, Analysis

https://doi.org/10.1144/geochem2022-055 | Vol. 23 | 2023 | geochem2022-055

Distribution and attribute analysis of soil selenium in Hebei Province, China

Liting Zhang[1,2,3], Weiming Xie[3], Wenlong Yu[3], Li Zhang[1], Shicong Zhang[3], Xi Wang[3], Xiaoyang Xie[3] and Hangxin Cheng[1*]

[1] Key Laboratory of Geochemical Cycling of Carbon and Mercury in the Earth's Critical Zone, Institute of Geophysical & Geochemical Exploration, Chinese Academy of Geological Sciences, Langfang 065000, China

[2] China University of Geosciences, Beijing 100083, China

[3] Hebei Institute of Geological Survey, Shijiazhuang 050081, China

HC, 0000-0001-8707-0578

* Correspondence: hangxin@vip.sina.com

Abstract: From 2004 to 2018, *c.* 73% of the land area in Hebei Province was surveyed by a 1:250 000 land quality geochemical survey of the National Geochemistry Survey of Land Quality. Based on the topsoil (0–20 cm) datasets (50 724 analysis samples) and attribute analysis of the key area with a semivariance function model, geographic detector model and random forest model, the content and spatial distribution of soil selenium (Se) in Hebei Province and the influencing factors of the distribution of Se in the Shijiazhuang–Xingtai–Handan area were obtained. The soil Se content in Hebei Province is low, ranging from 0.02 to 3.23 mg kg^{-1}, with an average of 0.20 mg kg^{-1}. The spatial distribution of soil Se was heterogeneous, and the marginal Se (0.125–0.175 mg kg^{-1}) area and the sufficient Se (0.175–0.40 mg kg^{-1}) area were widely distributed, but the acreage of the Se-rich (0.40–3.0 mg kg^{-1}) area was very limited, accounting for only 2.43% of the total survey area. However, the Shijiazhuang–Xingtai–Handan area was a significant Se enrichment area. The spatial distribution of Se in this area was restricted by natural and socioeconomic factors. Soil organic matter was the most influential factor and showed good coordination with Se. Coal burning superimposed on natural factors promotes the spatial differentiation of soil Se to a certain extent.

Keywords: soil selenium distribution; influence factors; semivariance function; GeoDetecor; random forest; Hebei Province

Received 27 October 2022; **revised** 11 January 2023; **accepted** 18 January 2023

Selenium (Se) was first discovered by Swedish chemist Jöns Jakob Berzelius in 1817. Its crustal abundance is 0.05 mg kg^{-1}, ranking 67th among the most abundant elements in the Earth's crust (Natasha *et al.* 2018). Selenium is an essential micronutrient for the human body and animals (Supriatin *et al.* 2015; Matos *et al.* 2017; Liu *et al.* 2018; Wang *et al.* 2018). There is a very narrow gap between Se deficiency and toxicity for living organisms, and both inadequate and excessive intake can lead to health problems (Li *et al.* 2012; Supriatin *et al.* 2015; Wang and Xu 2017). For example, Se deficiency can cause Keshan disease and Kashin–Beck disease (Tan and Huang 1991; Tan *et al.* 2002), while excess Se can cause poisoning, which can lead to nervous system disorders, nail and hair loss, paralysis and even death (Yang *et al.* 1983; Li *et al.* 2012; Song *et al.* 2020).

Soil Se content is the key factor affecting the Se content in plants, animals and the human body (Natasha *et al.* 2018; López-Bellido *et al.* 2019; Ngigi *et al.* 2020). However, the spatial distribution of Se in soil on the Earth's surface is heterogeneous. Some areas have high Se levels; for example, the Se content of the soil in the Egyptian North Nile Delta is 38.7 mg kg^{-1} (Shaheen *et al.* 2014). Some regions have very low Se levels; for example, the average Se content of agricultural soil in Bangladesh is only 0.044 mg kg^{-1} (Spallholz *et al.* 2008). The spatial distribution of Se in Chinese soil is also uneven (Tan *et al.* 2002; H. L. Liu *et al.* 2021), with 72% of the area lacking Se (Qin 2014; Yu *et al.* 2014), in particular a low-Se belt stretching from Heilongjiang Province in the NE to Yunnan Province in the SW (Wang and Gao 2001). However, there are also some typical Se-rich areas, such as Enshi Prefecture in Hubei Province (Zhu *et al.* 2008; Yuan *et al.* 2012), Ziyang County in Shaanxi Province (Luo *et al.* 2004; Tian *et al.* 2016), Kaiyang County in Guizhou Province (Ren and Yang 2014) and Taoyuan County in Hunan Province (Ni *et al.* 2016).

Numerous factors affect the distribution of Se in soil, which can be divided into natural and socioeconomic factors. Natural factors include soil parent rock or soil parent material, soil physicochemical properties, and topography. Socioeconomic factors include land use type, industrial and mining enterprise emissions and agricultural chemical consumption. Soil parent rock is the main source of Se in soil, and the composition of the parent rock determines its presence and (Malisa 2001; Gabos *et al.* 2014; Y. L. Liu *et al.* 2021). In general, soils formed from sedimentary rocks have high Se content, and those formed from magmatic rocks have low Se content (Wang and Gao 2001; Bujdoš *et al.* 2005). Soil physical and chemical properties also affect soil Se distribution. For example, soil organic matter (SOM) is one of the main physical and chemical properties that can affect the Se content in soil through selenite(+4) and selenate(+6) adsorption (Tan *et al.* 2002). Land use type is an important socioeconomic factor that strongly influences soil properties, particularly SOM. Thus, land use type has an important impact on the Se content and distribution in soil (Smith 2008; Tuttle *et al.* 2014; Xiao *et al.* 2020). Anthropogenic emissions are an important source of Se in the environment, including the consumption of petroleum and coal and the processing and extraction of metal elements (Wang and Gao 2001; Fernández-Martínez and Charlet 2009); in particular, the emission of coal combustion is the key to promoting the enrichment of Se in soil (Nriagu and Pacyna 1988; Wen and Carignan 2007).

Soil is an important basis for agricultural production. As a major agricultural province in China, the content, spatial distribution and influencing factors of soil Se have been reported in many studies in Hebei Province. However, most of the previous studies focused on a certain region, and there was never any research on the whole region. Benefiting from the 1:250 000 land quality geochemical survey of the National Geochemistry Survey of Land Quality project (NGSLQ), *c.* 73% of the land area of Hebei Province was surveyed from 2004 to 2018 (Fig. 1). Based on the topsoil (0–20 cm) test data of the survey results and taking all the regions where the geochemical survey was completed as the study area, this study aims to (1) evaluate the soil Se content, (2) explore the spatial distribution of soil Se in the study area and (3) determine the main factors affecting the Se distribution by establishing an attribute analysis model.

Materials and methods

Study area

Hebei Province covers an area of 188 800 km^2, with geographic coordinates of 113°27′ to 119°50′ (E) and 36°05′ to 42°40′ (N), located in the northern North China Plain (Fig. 1), surrounding the capital Beijing. It borders Bohai Bay to the east, the Taihang Mountains to the west, the Yanshan Mountains to the north and Henan Province to the south. Hebei Province has a temperate continental monsoon climate, with four distinct seasons in most areas. The annual sunshine duration is 2303.1 hours, the average annual precipitation is 484.5 mm, and the monthly average temperature is below 3°C. The terrain is high in the NW and low in the SE (Fig. 2a), with complex and diverse landforms, including the Bashang Plateau, Yanshan Mountains and Taihang Mountains, and Hebei Plain. Cenozoic sediments are widely distributed in plain areas, and igneous, sedimentary and metamorphic rocks are mainly distributed in mountainous and plateau areas (Fig. 2b). It is rich in mineral resources, mainly precious metals, nonferrous metals, iron mines and coal mines. Among them, precious metals mainly refer to gold mines; nonferrous metals mainly include copper, lead, zinc and molybdenum. All mineral resources are mainly distributed in the west, north and NE of Hebei Province (Fig. 2c). The soil types are diverse, mainly cinnamon soil, moisture soil, brown soil, chestnut soil, castano-cinnamon soils and skeleton soil (Fig. 2d). The class of cinnamon soil is semileached soil, with an average SOM content of 3.086%. The class of moisture soil is semihydromorphic soil, with an average SOM content of 2.360%. The class of brown soil is leached soil, with an average SOM content of 6.430%. The class of chestnut and castano-cinnamon soils is pedocal, with an average SOM content of 3.738%. The class of skeleton soil is primitive soil, and with an average SOM content of 1.596% (Chen *et al.* 2013). There are eight types of land use, of which the distribution area of cultivated land is the largest, accounting for *c.* 40% of the land area of Hebei Province (Fig. 2e).

Agriculture is an important pillar industry in Hebei Province, and the use of agricultural chemicals is very high. According to national statistics, the use of pesticides in Hebei reaches 57 300 tons every year, ranking 11th among the 31 provinces in China (except Taiwan, Hong Kong and Macao), and fertilizer use reaches 2 972 700 tons annually, ranking fourth among 31 provinces. With the development of the social economy, the province’s coal consumption continues to increase. In 1980, the total coal consumption was 26.5243 million tons. In the early 21st century, consumption was 101.813 8 million tons, and in 2019, coal consumption rose to 266 7423 million tons. Hebei Province is still dominated by coal power generation (Fig. 2f), which is an important method of coal consumption. Social development also drives population mobility, and an increasing number of people move into cities, which increases the gap between population density and road density in cities, towns and rural areas. The number of enterprises in Hebei Province also show an increasing trend, with the total number of enterprises in 2021 reaching 35 616, an increase of 11.6% over 2020. Among all types of enterprises, those closely related to the environment are mainly petroleum processing and petroleum product manufacturing, mining and smelting, manufacturing of chemical raw materials and chemical products, and pharmaceutical manufacturing.

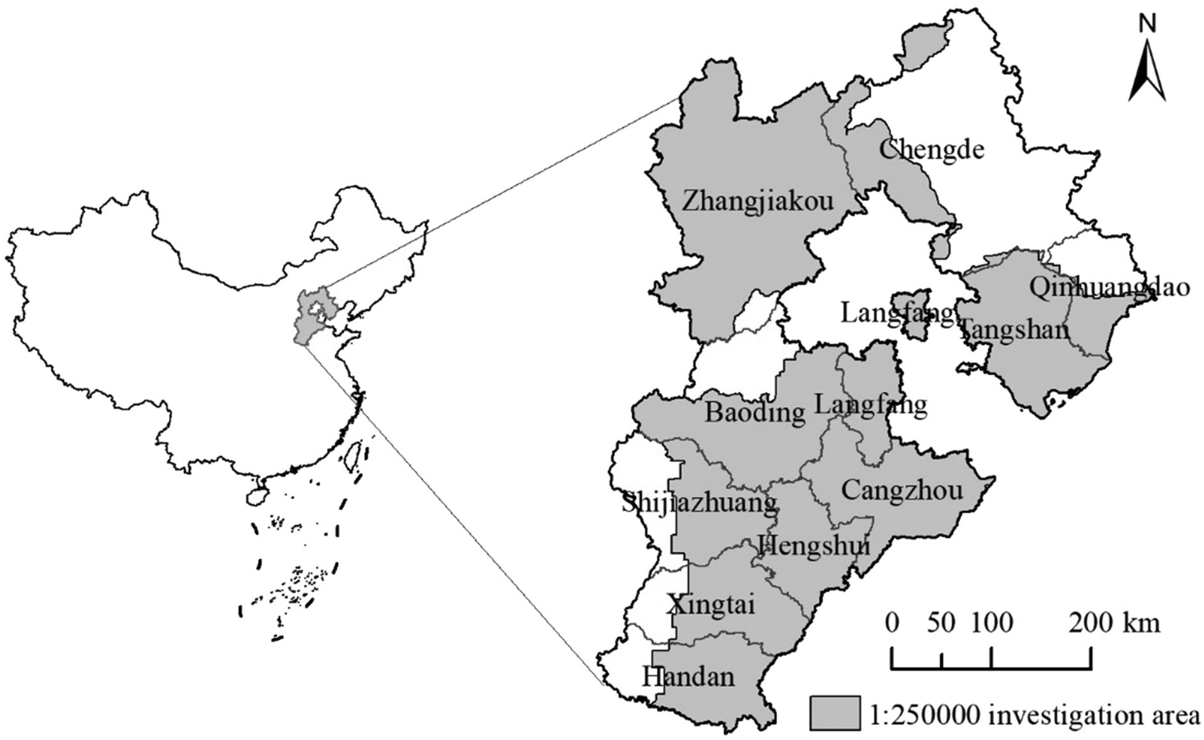

Fig. 1. Regional location and investigation area in Hebei Province.

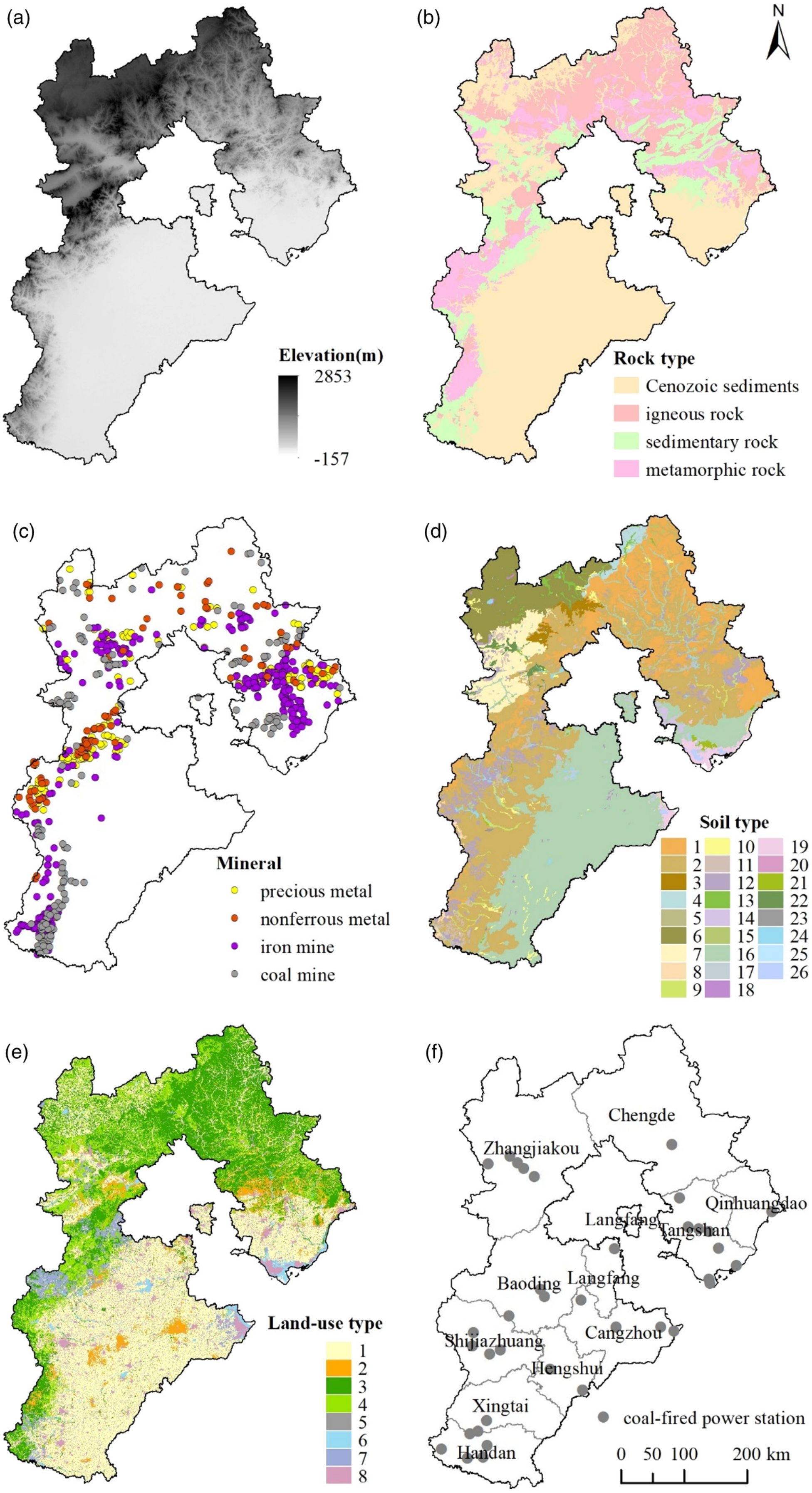

Fig. 2. Elevation (**a**), rock type (**b**), mineral (**c**), soil type (**d**), land-use type (**e**), and coal-fired power station (**f**) in Hebei Province. Notation: in Figure 2d, soils: 1–brown, 2–cinnamon, 3–grey cinnamon, 4–grey forest, 5–chernozem, 6–chestnut, 7–castanocinnamon, 8–loessal, 9–alluvial, 10–aeolian sandy, 11–chisley, 12–skeleton, 13–meadow, 14–Shajiang black, 15–mountain meadow, 16–moisture, 17–boggy, 18–saline, 19–salinized, 20–solonetz, 21–paddy, 22–irrigation-warping; 23–urban district, 24–lakes and reservoirs, 25–sandbanks and islands in rivers, 26–coastal salt farm or fish farm. In Figure 2e, 1–cultivated land, 2–garden plot, 3–forest land, 4–grassland, 5–transportation land, 6–water area and water conservancy facilities land, 7–other land, 8–towns, villages, industrial and mining land.

Sample collection and preparation

Sample collection and preparation were performed according to national standards outlined in 'Specification of multi-purpose regional geochemical survey (1:2 50 000) (DZ/T 0258-2014)' and 'Specification of land quality geochemical assessment (DZ/T 0295-2016)'. The 1:250 000 geochemical survey of land quality of the NGSLQ project adopted grid distribution points and collected topsoil samples (0–20 cm). The sampling densities were one sample/1 km^2, and then four samples within 4 km^2 were formed into a composite sample, each sample weighing ≥1000 g. The sampling medium was a mix of mature soil types in the work area, avoiding obvious point pollution areas, fertilizer masses and newly accumulated soils and ridges, and the sampling points were 100 m from the main road and railway. After drying naturally, the soil samples were screened with a 20-mesh nylon sieve, and 200 g was put into a special sample bag and sent to the laboratory. Based on the above principles, a total of 2 02 813 sample datasets and 50 724 analysis sample datasets of topsoil were obtained.

Chemical analysis and quality control

Sample analysis and quality control of 1:250 000 geochemical survey of land quality of the NGSLQ project according to national standards 'DZ/T 0258-2014' and 'DZ/T 0295-2016', have been described in detail in prior studies (Li *et al.* 2014; Zhang *et al.* 2020).

In the current study, soil samples were ground to 200 mesh before chemical analysis. Selenium was analysed by hydride generation-atomic fluorescence spectrometry after digestion by HCl, H_2SO_4, HNO_3 and $HClO_4$. Organic carbon (C_{org}) was analysed by the volumetric method (VOL) after the sample was oxidized by $K_2Cr_2O_7$. The pH was analysed by an ion selective electrode after extraction with distilled water, and the sample-to-water ratio was 1:2.5. The detection limits of Se, C_{org} and pH were 0.01 mg kg^{-1}, 0.1% and 0.10, respectively. The reportable rate of all samples for all analysed indictors was 100%.

The accuracy and precision of the analysis method were controlled by inserting the national first-grade certified reference material (GBW) (GBW07401 (GSS-1)-GBW07408 (GSS-8), GBW07423 (GSS-9)-GBW07430 (GSS-16), GBW07446 (GSS-17)-GBW07457 (GSS-28)) as a code sample during the analysis process. Specifically, 12 GBWs were inserted per 500 samples to control accuracy and four GBWs were inserted per 50 samples to control precision. The pass rates of accuracy and precision were 100%, in accordance with the requirements of national standards ≥98%. At least 5% of samples were randomly selected for the repeatability test, and the qualified rates of Se, C_{org} and pH were 98.35%, 98.39% and 97.27%, respectively, in accordance with the requirements of national standards ≥90%. All samples were analysed in accordance with quality control requirements.

Statistical analysis, mapping and modelling

Statistical analysis

The descriptive statistical parameters of soil Se content were calculated in Excel, including maximum, minimum, arithmetical mean, standard deviation, coefficient of variation, median value and five other percentiles (P2.5, P25, P75, P85 and P97.5).

Spatial distribution map creation

In ArcGIS 10.2, we gridded the discrete data using the ordinary kriging interpolation method and then created a spatial distribution map of soil Se with 0.125, 0.175, 0.400 and 3.00 mg kg^{-1} as interval values (Tan and Huang 1991; Tan *et al.* 2002).

Attribute analysis model

The semivariance function model

As an important part of geostatistics, the semivariance function can describe the structure and randomness of regional variables. Common fitting models are the linear, spherical, exponential and Gaussian models, and the main parameters include nugget, sill, partial sill and range. The value of nugget/sill is an important indicator reflecting the spatial variation in regional variables. The larger the value, the stronger the spatial variability, that is, the weaker the spatial correlation (Yang *et al.* 2009; Xie *et al.* 2013; Mamat *et al.* 2014; Song *et al.* 2020).

Outliers and data distributions affect the reliability of semivariance functions (Oliver and Webster 2014). In this paper, the outliers were repeatedly removed according to the Pauta criterion (mean plus or minus 3 standard deviations) until they were no different (Hou *et al.* 2020). The original soil Se content data did not conform to a normal distribution, so the normal score was used for normal transformation in SPSS 18.0, and skewness and kurtosis quantitative judgment were used to complete the normal test. The semivariance function is calculated and fitted in GS+ 9.0.

The GeoDetector model

A geographical detector (GeoDetector) was proposed by Wang Jinfeng in 2005. It is a new statistical method to detect the spatial heterogeneity of variables and reveal the driving factors behind it. It mainly consists of four detectors, namely, factor detector, interaction detector, risk detector and ecological detector (J. F. Wang *et al.* 2010, 2016; J. F. Wang and Xu 2017).

Fifteen factors were selected for the modelling of GeoDetector, which can be divided into natural and socioeconomic factors. Natural factors include rock type, soil type, pH, SOM, elevation (H), slope and topographic wetness index (TWI), while socioeconomic factors include land-use type, road density, population density, annual coal consumption for power generation, annual fertilizer use, annual pesticide use, number of enterprises and number of mineral resources. The digital elevation model (DEM: resolution is 30 m; DEM data were sourced from the China Geospatial Data Cloud) was selected as the data source, and H, slope and TWI were extracted using ArcGIS 10.2. Road network data were derived from the Open Street Map, and population density, coal consumption for power generation, fertilizer use and pesticide use were derived from national statistics. The types of enterprises mainly include petroleum processing and petroleum product manufacturing, mining and smelting, manufacturing of chemical raw materials and chemical products, and pharmaceutical manufacturing. Mineral resources mainly include precious metals, nonferrous metals, iron mines and coal mines. SOM was obtained by multiplying the C_{org} content by 1.724.

ArcGIS 10.2 was used for data processing, including data discretization and resampling, and then data information was read and calculated in Excel GeoDetector.

The random forest model

The random forest (RF), proposed by Breiman (2001), is a machine-learning algorithm based on multiple decision trees that can evaluate the importance of different variables and reveal the relative importance of different environmental factors in the prediction and decision structure (Grimm *et al.* 2008; Jia *et al.* 2020). The mean decrease Gini is able to calculate the effect of each variable on the heterogeneity of the observed values at each node of the classification tree, which is an indicator used to indicate the importance of variables; the higher the value, the higher the importance.

The RF model selects the same variables as the GeoDetector model. Fifteen variables were resampled in ArcGIS 10.2 and then

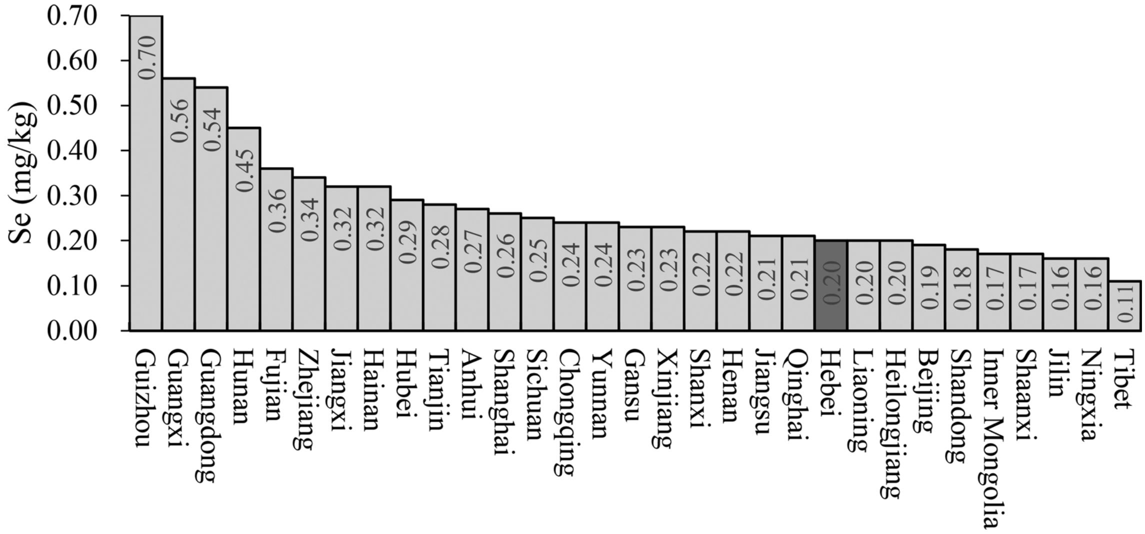

Fig. 3. Topsoil Se content in different provinces of China.

simulated and calculated using the randomForest, caret and pROC packages in R 4.1.1.

Results and discussion

Soil Se content

The Se content of topsoil in Hebei Province ranged from 0.02 to 3.23 mg kg^{-1}, with an average value of 0.20 mg kg^{-1}, which was lower than the background value of 0.29 mg/kg mg kg^{-1} in China (Wei *et al.* 1991). The soil Se content in different provinces of China (except Taiwan, Hong Kong and Macao) ranges from 0.11 to 0.70 mg kg^{-1} (Hou *et al.* 2020), and the Se content in Hebei Province is relatively low, ranking 22nd in China (Fig. 3).

The topsoil Se content of 11 cities in Hebei Province from high to low is as follows: Shijiazhuang>Handan>Tangshan>Baoding>Xingtai>Qinhuangdao ≥0.20 mg kg^{-1}>Hengshui>Cangzhou>Langfang>Chengde>Zhangjiakou (Table 1). The minimum value of Se content is in Tangshan, and the maximum value is in Baoding. The coefficient of variation (CV) reflects the degree of variation of the element content, with higher the valuesreflecting higher degrees of variation (Nezhad *et al.* 2015). The CV of the soil Se content in Tangshan and Handan was above 0.5, indicating a high degree of variation, while the CV of the other cities was between 0.2 and 0.5, indicating a moderate degree of variation and a relatively uniform Se content in the soil.

Soil Se spatial distribution

The soil Se content classification is generally as follows: Se-deficient (<0.125 mg kg^{-1}); Se-marginal (0.125–0.175 mg kg^{-1}); Se-sufficient (0.175–0.40 mg kg^{-1}), Se-rich (0.40–3.0 mg kg^{-1}), and Se-excessive (>3.0 mg kg^{-1}) (Tan and Huang 1991; Tan *et al.* 2002). The soil Se content in Hebei Province is mainly marginal and sufficient, the Se-deficient areas are mainly distributed in Zhangjiakou and western Chengde, the Se-rich areas are mainly located in Tangshan, Baoding, Shijiazhuang, Xingtai and Handan, and there is no area of Se surplus (Fig. 4).

The acreage of Se-rich areas in Hebei Province is very limited, *c.* 3244.23 km^2, accounting for only 2.43% of the total survey area, but it is very concentrated in spatial distribution. The largest and most concentrated Se-rich area is located in Shijiazhuang–Xingtai–Handan, showing a discontinuous banded distribution, and the corresponding Se content is 0.52 mg kg^{-1}, which is consistent with previous research results. For example, the soil Se content of the abnormal Se area in Gaocheng, Shijiazhuang is 0.47 mg kg^{-1} (Zhao *et al.* 2021) and 0.598 mg kg^{-1} in Handan (Zhang and Sun 2021).

Table 1. *Statistical parameters of soil Se content in 11 cities of Hebei Province*

City	Num.	Min.	P2.5	P25	P50	P75	P85	P97.5	Max.	AM	STD	CV
Zhangjiakou	25 123	0.04	0.09	0.13	0.15	0.17	0.19	0.32	3.02	0.16	0.07	0.45
Xingtai	2360	0.06	0.11	0.16	0.20	0.25	0.29	0.44	0.83	0.22	0.09	0.39
Tangshan	3193	0.02	0.09	0.16	0.21	0.27	0.30	0.52	3.00	0.23	0.14	0.61
Shijiazhuang	2090	0.03	0.12	0.19	0.23	0.31	0.37	0.54	1.27	0.26	0.11	0.44
Qinhuangdao	801	0.03	0.06	0.14	0.19	0.24	0.28	0.41	0.98	0.20	0.09	0.45
Langfang	1573	0.05	0.10	0.15	0.18	0.20	0.21	0.27	0.45	0.18	0.04	0.25
Hengshui	2191	0.10	0.13	0.17	0.19	0.21	0.22	0.27	0.81	0.19	0.04	0.21
Handan	2152	0.05	0.12	0.17	0.20	0.29	0.35	0.62	2.37	0.25	0.15	0.61
Chengde	3698	0.04	0.07	0.12	0.15	0.19	0.23	0.35	0.79	0.17	0.07	0.45
Cangzhou	3508	0.07	0.12	0.16	0.18	0.20	0.21	0.26	0.60	0.18	0.04	0.21
Baoding	4035	0.04	0.12	0.18	0.21	0.25	0.28	0.43	3.23	0.22	0.10	0.44
Hebei province	50 724	0.02	0.09	0.14	0.17	0.21	0.24	0.40	3.23	0.20	0.09	0.45

All concentrations are in mg kg^{-1}.
Num. = number of samples; *P* = percentile (P50 = median); Min. = minimum; Max. = maximum; AM = arithmetical mean; STD = standard deviation; CV = coefficient of variation.

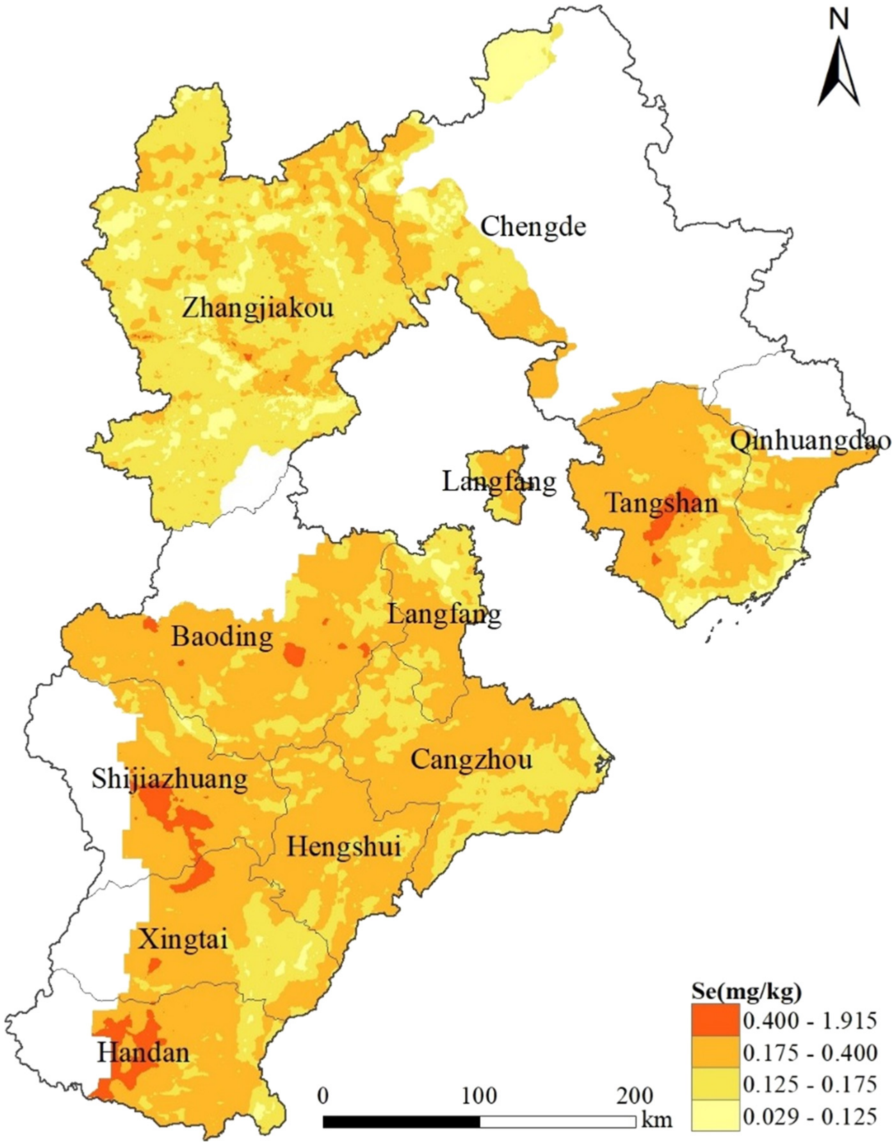

Fig. 4. Topsoil Se spatial distribution of Hebei Province.

Attribute analysis

According to the distribution of soil Se in Hebei Province, Tangshan, Baoding, Shijiazhuang, Xingtai and Handan all have Se-rich areas, but Shijiazhuang–Xingtai–Handan has the largest and most concentrated Se-rich area in Hebei Province, and this area is also an important agricultural production base, with a total cultivated land area of 1 939 483 hectares, much higher than Tangshan's cultivated land area of 557 942 hectares. In addition, Shijiazhuang is the capital of Hebei Province and has an important economic and social status. Overall, this paper focuses on Shijiazhuang–Xingtai–Handan to explore the main factors affecting the spatial distribution of soil Se.

Semivariance function model result

A total of 6274 analysis sample datasets were obtained after excluding outliers. The results of the normality test are shown in Table 2. The skewness and kurtosis values are −0.005 and −0.057, respectively, *c.* 0. At the test level of $\alpha = 0.05$, the Z scores of skewness and kurtosis are −0.169 and −0.928, respectively, satisfying the range of variables limited by the hypothesis (±1.96), and the data are considered to follow a normal distribution.

The semivariance function model has many important parameters. Nuggets are caused by measurement error or spatial variation. The sill is a stable constant of the semivariance function when the sample spacing increases to a certain extent. The nugget/sill ratio is the base effect, which reflects the spatial variation degree of regional variables and measures the influence degree of regional and nonregional factors on variables. Regional factors refer to natural factors, while nonregional factors refer to socioeconomic or human factors (Xie *et al.* 2013). When the ratio is less than 25%, the variables have strong spatial correlation and are mainly affected by natural factors. When the ratio is between 25% and 75%, the variables are influenced by natural and socioeconomic factors. When the ratio is greater than 75%, the variables have strong spatial variability, and socioeconomic factors play a major role (Cambardella *et al.* 1994; Martín *et al.* 2006). The range represents the distance at which the variable is spatially dependent. The coefficient of determination (R^2) and residual sum of square (RSS) reflect the fitting accuracy of the theoretical model; the larger R^2 and

Table 2. *Results of the soil Se content data normal test*

	Value	STD	Z-score
Skewness	−0.005	0.031	−0.169
Kurtosis	−0.057	0.062	−0.928

STD = standard deviation; Z-score = Value/STD.

Table 3. *Semivariance function model parameters*

Element	Model	Nugget (C_0)	Sill ($C_0 + C$)	$C_0/(C_0 + C)$	Range (A_0)	R^2	RSS
Se	Spherical	0.315	1.133	27.8%	98.2 km	0.998	1.71E-03

the smaller RSS are, the higher the accuracy (Mamat *et al.* 2014; Song *et al.* 2020).

According to the results (Table 3 and Fig. 5), the semivariance function conforms to the spherical model. R^2 and RSS are 0.998 and 1.71×10^{-3}, respectively, indicating that the model has high accuracy. The nugget/sill ratio is 27.8%, in the range of 25%–75%, indicating that soil Se is affected by natural and socioeconomic factors and has a moderate spatial correlation with a range of 98.2 km.

GeoDetector model results

The data required by the GeoDetector model include dependent and independent variables. The independent variable data should be a type quantity, and when it is a numerical quantity, it needs to be discretized. Discretization can be based on professional knowledge or classification methods such as equal interval, quantile and natural discontinuous point classification (Wang and Xu 2017). In this paper, the dependent variable is the soil Se content, and the independent variables are the 15 factors mentioned above, among which rock type, soil type and land-use type are type quantities, and the others are numerical quantities. pH, H, number of enterprises and mineral resources are manually classified according to data characteristics. Slope data are classified according to Y. L. Liu *et al.* (2021). The SOM, TWI, road density, population density, coal consumption for power generation, fertilizer and pesticide use data are discretized using the natural discontinuity point classification method. The classification results of the 15 factors are shown in Table 4 and Figure 6.

A factor detector and interaction detector are used to quantify the effects of different factors on the soil Se spatial distribution. The factor detector is used to detect the spatial differentiation of dependent variables and measure the explanatory ability of different factors to the spatial differentiation of dependent variables with the value of q. The value of q is between 0 and 1, and the larger the value, the stronger the explanatory ability. The interaction detector is used to evaluate the changes in the explanatory ability of dependent variables when different factors work together, including nonlinear attenuation q (X1∩X2)<Min (q (X1), q (X2)), single-factor nonlinear attenuation Min (q (X1), q (X2))<q (X1∩X2)<Max (q (X1), q (X2)), double-factor enhancement q (X1∩X2)>Max (q (X1), q (X2)), independent q (X1∩X2) = q (X1) + q (X2), and nonlinear enhancement q (X1∩X2)>q (X1) + q(X2) (Wang and Xu 2017).

The factor detector results (Table 5) show that all factors passed the significance test ($P < 0.05$) except TWI. Different factors have different explanatory abilities for the spatial distribution of soil Se, with q values ranging from 0.013 to 0.718, and the sequence from large to small is SOM>coal>population density>pesticide>soil type>fertilizer>road density>H > pH>rock type>land-use type> enterprise>mineral slope. SOM is the most critical factor, with a q value of 0.718, accounting for 33.18% of the explanatory power (Fig. 7), which is much higher than the second factor, coal burning. Coal and population density, both socioeconomic, are relatively important influencing factors, accounting for 14.92% and 11.08%, respectively. The influence of other factors is low, accounting for less than 10%.

The results of the interaction detector (Table 6) show that there are two kinds of changes in explanatory ability when two factors (X1, X2) act together: double-factor enhancement, q (X1∩X2) >Max (q (X1), q (X2)), and nonlinear enhancement, q (X1∩X2)>q (X1) + q (X2). q (SOM∩X_n) is significantly higher than the other factor interaction results, and q (SOM∩pesticide) is the largest, indicating that the interaction of SOM and pesticides had the greatest influence on the spatial distribution of soil Se. In addition, q (coal∩population density), q (coal∩pesticide) and q (population density∩pesticide) are also significant, indicating that the superposition of these factors has an important effect on soil Se distribution.

The results of GeoDetector modelling show that the distribution of soil Se is affected by both natural factors and socioeconomic factors, which is consistent with the results of semivariance function analysis. SOM, coal burning, population density and pesticides are the four most important factors.

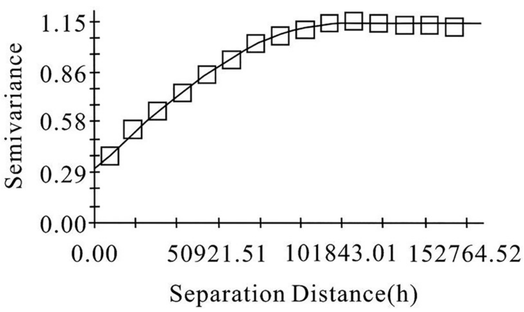

Fig. 5. Semivariance function model-fitting results.

RF model results

The RF model was established with soil Se content as the dependent variable and the aforementioned 15 factors as independent variables. The soil Se content was divided into four levels with 0.2, 0.3 and 0.4 mg kg^{-1} as the interruption values. The independent variable data were divided into a training set and a test set at a ratio of 7:3.

Accuracy (0-1), Kappa (0-1) and AUC (0-1) are indicators to evaluate the accuracy of the model, and the higher the value, the better the model effect. The number of variables selected at each node (mtry) and the number of decision trees (ntree) affect the accuracy of the model. By default, mtry is the quadratic root of the number of variables, ntree is 500, and optimization of these two parameters can obtain a better model effect. Table 7 shows that the model error is minimal when mtry is 6. The optimal ntree is searched under the condition of mtry = 6. As shown in Figure 8, the model error tends to be basically stable when ntree is 500. Through the above attempts, mtry and ntree are set at 6 and 500, respectively. By comparison, it can be seen that the model has higher accuracy under the optimal parameters (Table 8).

The mean decrease in Gini represents the importance of variables. The results show (Figs 9 and 10) that the soil Se spatial distribution is influenced by both natural factors and socioeconomic factors. SOM and pH are the most important natural factors, especially SOM, which has the strongest influence. Its mean decrease in Gini is as high as 1062.86, and the importance proportion is 34.98%, which is much higher than the second rank pH. Coal burning is the most important socioeconomic factor, accounting for nearly 10%. The influence of other variables is weak, and the importance proportion is mostly below 7%.

Table 4. *Classification results of the 15 factors*

Factors	1	2	3	4	5	6	7
Rock type	pluvial–alluvial sediment	alluvium	Lacustrine sediment	aeolian sediment	acidite	basite	intermediate rock
Soil type	alluvial soil	paddy soil	boggy soil	moisture soil	saline soil	chisley soil	Shajiang black soil
pH	5.9–6.5	6.5–7	7–7.5	7.5–8	8–8.7	/	/
SOM (%)	0.10–1.10	1.10–1.40	1.40–1.66	1.66–1.92	1.92–2.14	2.14–2.44	2.44–2.85
H (m)	−23–49	49–88	88–152	152–243	243–418	418–1018	/
Slope (°)	0–5	5–8	8–15	15–25	25–35	35–58.86	/
TWI	2.93–6.24	6.24–7.43	7.43–8.90	8.90–10.37	10.37–11.74	11.74–13.12	13.12–14.77
Land-use type	cultivated land	garden plot	forest land	grassland	transportation land	water area and water conservancy facilities land	other land
Road density (km km^{-2})	0.00–0.94	0.94–1.53	1.53–2.09	2.09–2.64	2.64–3.27	3.27–4.05	4.05–5.00
Population density (people/km^2)	67–208	208–407	407–581	581–654	654–731	731–829	829–942
Coal (10 000 t/a)	10.46–12.56	12.56–20.98	20.98–25.19	25.19–31.50	31.50–46.23	46.23–58.86	58.86–67.28
Fertilizer (kg hm^{-2}·a)	159.01–163.89	163.89–271.31	271.31–354.32	354.32–451.97	451.97–530.09	530.09–622.86	622.86–700.99
Pesticide (kg hm^{-2}·a)	2.21–3.01	3.01–4.80	4.80–6.80	6.80–7.79	7.79–8.59	8.59–11.98	11.98–15.97
Enterprise/4 km^2	0	0–2	2–6	6–12	12–21	21–47	/
Mineral/4 km^2	0	1	2	3	/	/	/
Factors	8	9	10	11	12	13	
Rock type	vein rock	carbonatite	clastic rock	clay rock	regionol metamorphic rock	contact metamorphic rock	
Soil type	skeleton soil	cinnamon soil	aeolian sandy soil	/	/	/	
pH	/	/	/	/	/	/	
SOM (%)	2.85–3.48	3.48–4.70	4.70–9.57	/	/	/	
H (m)	/	/	/	/	/	/	
Slope (°)	/	/	/	/	/	/	
TWI	14.77–16.79	16.79–19.64	19.64–26.34	/	/	/	
Land-use type	towns, villiages, industrial and mining land	/	/	/	/	/	
Road density (km km^{-2})	5.00–6.14	6.14–7.59	7.59–10.03	/	/	/	
Population density (people/km^2)	942–1111	1111–3274	3274–8206	/	/	/	
Coal (10 000 t/a)	67.28–94.64	94.64–126.21	126.21–547.12	/	/	/	
Fertilizer (kg hm^{-2}·a)	700.99–764.46	764.46–871.88	871.88–1404.09	/	/	/	
Pesticide (kg hm^{-2}·a)	15.97–19.36	19.36–22.95	22.95–53.07	/	/	/	
Enterprise/4 km^2	/	/	/	/	/	/	
Mineral/4 km^2	/	/	/	/	/	/	

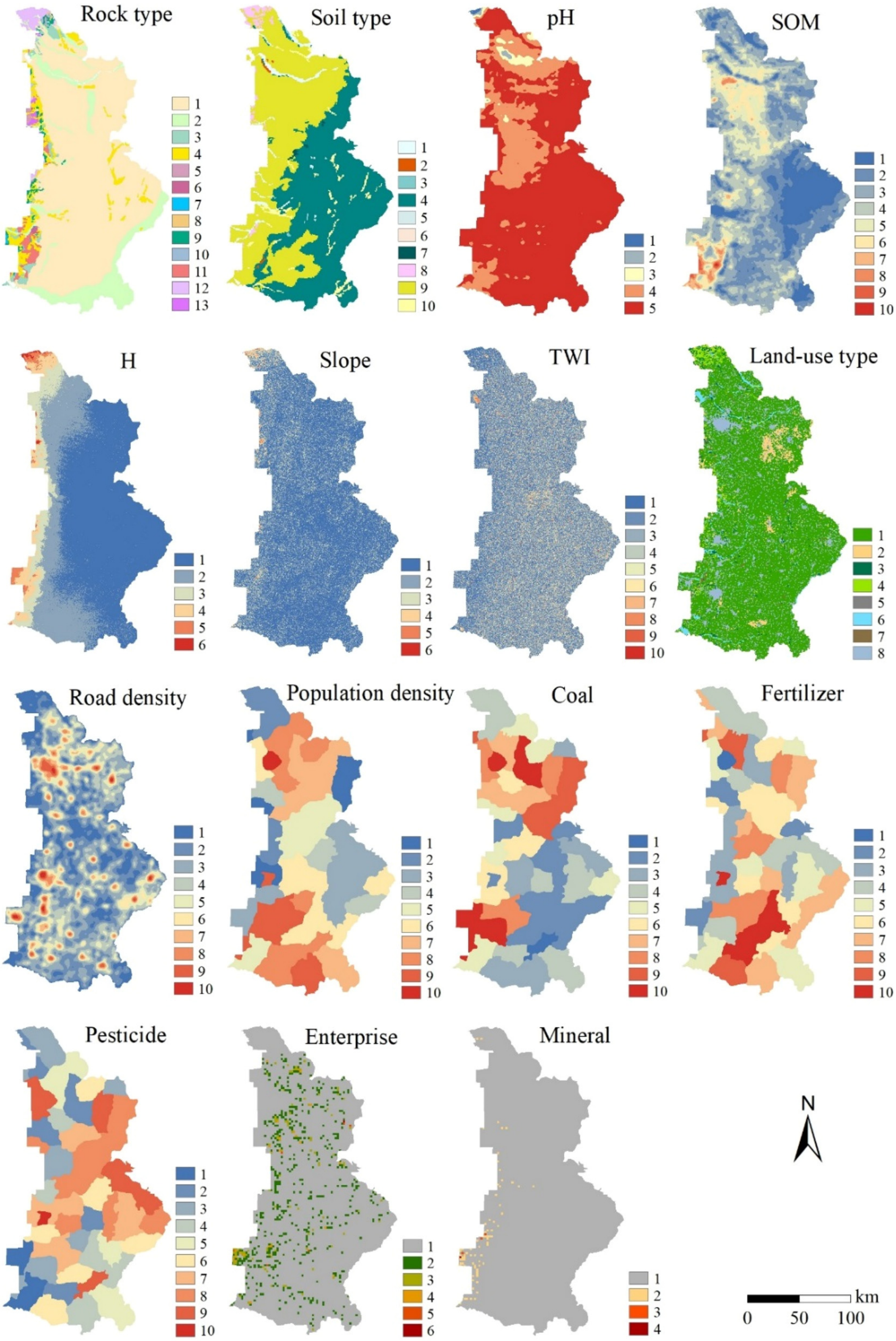

Fig. 6. Classification results of 15 factors. Refer to Table 4 for different factor category information.

Model results comparison

The results of the semivariance function, GeoDetector and RF are very consistent, which reveals that the spatial distribution of soil Se in the Shijiazhuang–Xingtai–Handan area is affected by natural and socioeconomic factors. q and the mean decrease in Gini are used to represent the influence of variables in GeoDetector and RF, respectively. By comparison, SOM and coal burning are both very important factors in the two models, especially SOM, ranked first, and whose influence is much higher than that of the other variables. pH is a variable with large differences, ranking ninth in GeoDetector and second in RF, which may be caused by the different principles and methods of the two models. There are also differences in the ranking of other variables, but the results of the two models show that the influence of these variables is very low,

Table 5. *Factor detector results*

Factors	*q*	*p*	Factors	*q*	*p*
Rock type	0.059	0	Road density	0.120	0
Soil type	0.151	0	Population density	0.240	0
pH	0.075	0	Coal	0.323	0
SOM	0.718	0	Fertilizer	0.123	0
H	0.085	0	Pesticide	0.203	0
Slope	0.008	0.007	Enterprise	0.018	0
TWI	0.006	0.122	Mineral	0.013	0
Land-use type	0.029	0			

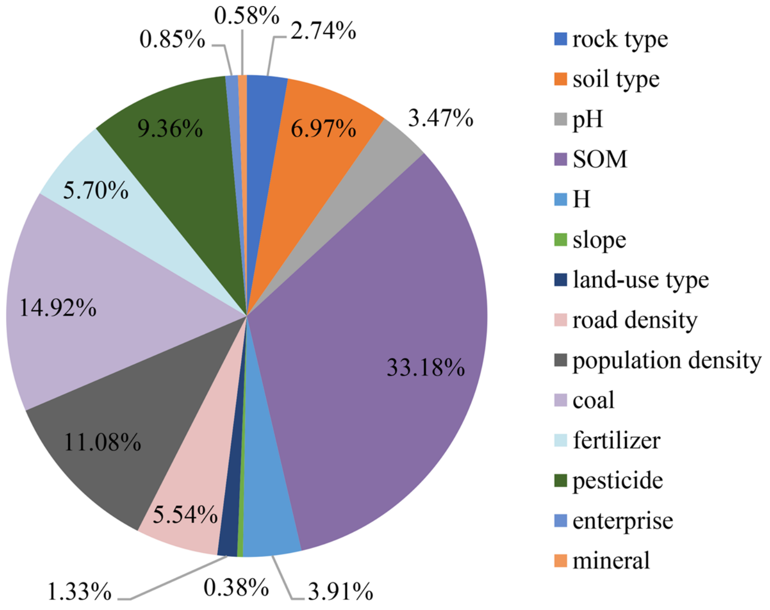

Fig. 7. Different factor *q* proportions.

Table 6. *Interaction detector results*

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15
X1	0.059														
X2	0.216	0.151													
X3	0.144	0.218	0.075												
X4	0.748	0.733	0.734	0.718											
X5	0.191	0.188	0.184	0.743	0.085										
X6	0.072	0.160	0.084	0.720	0.094	0.008									
X7	0.073	0.163	0.081	0.725	0.095	0.016	0.006								
X8	0.100	0.180	0.105	0.727	0.128	0.041	0.042	0.029							
X9	0.191	0.270	0.203	0.743	0.248	0.136	0.138	0.150	0.120						
X10	0.372	0.436	0.308	0.778	0.436	0.260	0.258	0.270	0.372	0.240					
X11	0.385	0.404	0.400	0.773	0.395	0.329	0.330	0.356	0.423	0.655	0.323				
X12	0.213	0.305	0.254	0.765	0.313	0.138	0.139	0.159	0.260	0.560	0.595	0.123			
X13	0.265	0.374	0.281	0.796	0.344	0.220	0.216	0.251	0.382	0.633	0.614	0.549	0.203		
X14	0.079	0.166	0.096	0.723	0.106	0.029	0.028	0.049	0.134	0.272	0.330	0.146	0.222	0.018	
X15	0.069	0.158	0.091	0.722	0.091	0.021	0.019	0.043	0.135	0.261	0.326	0.137	0.205	0.028	0.013

X1 = rock type; X2 = soil type; X3 = pH; X4 = SOM; X5 = H; X6 = slope; X7 = TWI; X8 = land-use type; X9 = road density; X10 = population density; X11 = coal; X12 = fertilizer; X13 = pesticide; X14 = enterprise; X15 = mineral.

Table 7. *Model error rates with different mtry*

mtry	1	2	3	4	5	6	7	8
error rate	0.3346	0.2354	0.2219	0.2133	0.2145	0.2127	0.2128	0.2135
mtry	9	10	11	12	13	14	15	
error rate	0.2198	0.2232	0.2187	0.2192	0.2150	0.2238	0.2249	

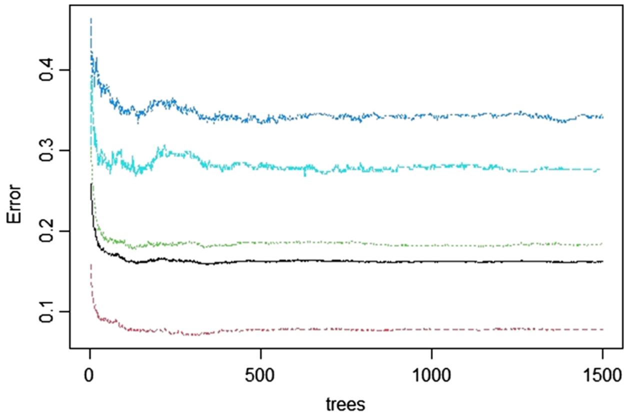

Fig. 8. Relationship between ntree and model error.

Table 8. *Comparison of model accuracy with different parameters*

Indicators	Default parameters	Optimal parameters
Accuracy	83.40%	84.00%
Kappa	0.7435	0.7535
AUC	0.9527	0.9531

and the conclusions are consistent on this point, so there is no further discussion.

Influence factors of Se spatial distribution

Although the results of attribute analysis show that rock type is not the key factor affecting the distribution of soil Se, there is no doubt that rock or parent material is the most important material base of soil (Fordyce *et al.* 2000, 2010; Spadoni *et al.* 2007; Li *et al.* 2008). Quaternary pluvial–alluvial sediments are most widely distributed in the Shijiazhuang–Xingtai–Handan area, and the Taihang Mountains in the west are the main provenance area. After weathering, denudation and migration, the weathering products of mountain rocks accumulate in the piedmont plain. Therefore, the Quaternary sediments of Shijiazhuang, Xingtai and Handan are a mixture of various rock types. The main rock types in the Taihang Mountains include acidite (mainly granite), limestone, dolomite, quartzite, sandstone, shale, gneiss and granulite, with different Se contents (Table 9). Sedimentary rock has the highest Se content, followed by metamorphic and magmatic rock. These rocks are the original source of Se in the soil.

As shown in Table 9, the element content will change after the rock forms into soil. On the one hand, the soil inherits the parent rock, such as shale, rock and soil, which have relatively high Se content. On the other hand, soil element secondary enrichment and the Se content of all types of soil are significantly higher than those of the parent rock, indicating that the soil Se content is more affected by factors other than the parent rock (Li 2010). Based on previous studies, the physical and chemical properties of soil have an important influence on the content of elements, especially in the long process of soil formation. The influence of parent rock or parent material is gradually weakened, but the effect of physico-chemical properties will continue to increase (Wang *et al.* 2013).

SOM is generally considered to be the factor most closely related to soil Se levels (Yamada *et al.* 2009; Roca-Perez *et al.* 2010). The mean SOM content in the Shijiazhuang–Xingtai–Handan area is 1.54%, which is strongly correlated with Se, and the complex correlation coefficient (R^2) is 0.59 (Fig. 11). With the increase in SOM content, the soil Se content increases (Table 10). Se in soil mainly exists in the form of selenate (SeO_4^{2-}) and selenite (SeO_3^{2-}), and SOM is an important part of the adsorption of selenate (+6) and selenite (+4) (Weng *et al.* 2011). Selenite is strongly adsorbed and has low mobility, so is more sorbed than selenate in soil (A.A. *et al.* 2020). The presence of higher levels of SOM helps to create anoxic zones that facilitate reduction and thus immobilization of Se (Li *et al.* 2017; Wang *et al.* 2018).

Socioeconomic factors also have an important impact on soil element content, and coal burning is the most important nonnatural factor affecting soil Se content in the Shijiazhuang–Xingtai–Handan area. The Fengfeng Handan, Xingtai, Jingxing and Yuanshi mining areas in Shijiazhuang are important coal production bases in Hebei Province (Yang 2006). The Se content in coal is high; for example, in the Xingtai coal mine it is 3.99 mg kg^{-1} (Yu *et al.* 2010), which is higher than the national average of 3.91 mg kg^{-1} (L. Wang *et al.* 2010). Figures 2f and 4 show that the distribution of coal-fired power stations in Shijiazhuang–Xingtai–Handan has a good spatial correspondence with Se-rich areas, which should be due to the high volatility of Se. More than 75% of Se in coal is released into the atmosphere during combustion, and then into the soil through dry and wet deposition (L. Wang *et al.* 2010; Zhao *et al.* 2021). In previous studies,, the Se flux into the soil through atmospheric deposition was 5.32–6.01 g $hm^{-2}\cdot a$ in the Shijiazhuang–Xingtai–Handan area, accounting for *c.* 55% of

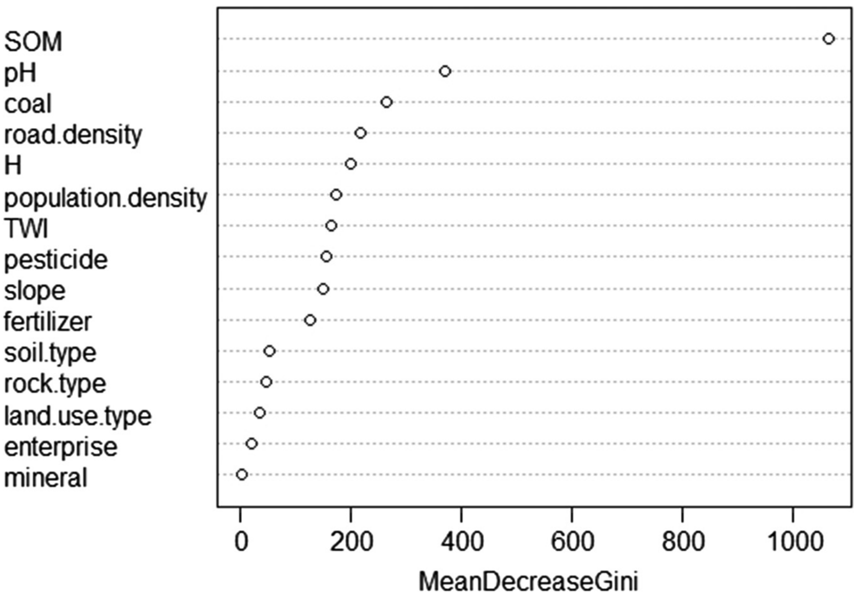

Fig. 9. Mean decrease in Gini of different factors.

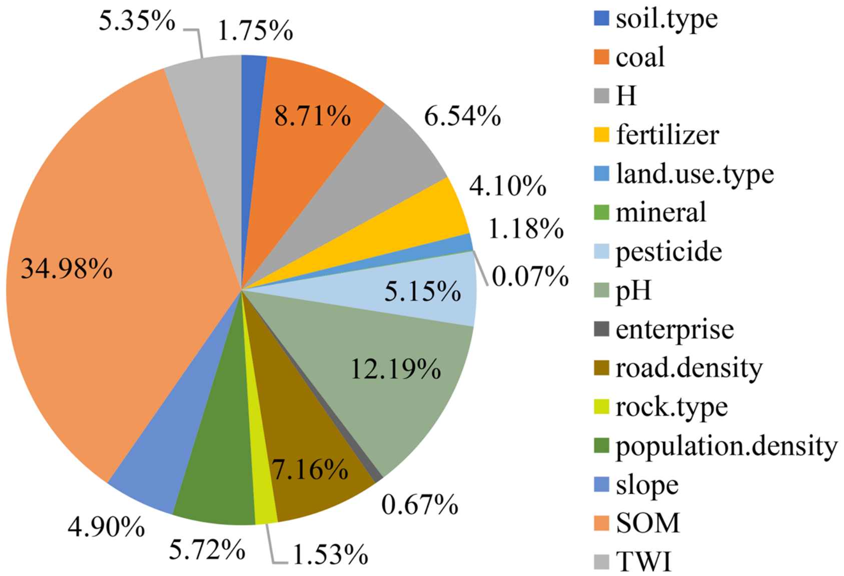

Fig. 10. The proportion of different factors mean decrease in Gini.

Table 9. *Se content in different rocks and soils*

Rock type	Se_{rock} (mg kg^{-1})	Se_{soil} (mg kg^{-1})	Se_{soil}/Se_{rock}
magmatic rock			
acidite	0.033	0.15	4.55
sedimentary rock			
limestone	0.046	0.57	12.40
dolomite	0.057	0.30	5.26
quartzite	0.016	0.17	10.60
sandstone	0.059	0.17	2.88
shale	0.222	0.34	1.53
metamorphic rocks			
gneiss	0.074	0.18	2.43
granulite	0.068	0.17	2.50

Se_{rock} and Se_{soil} from Gong (1997).

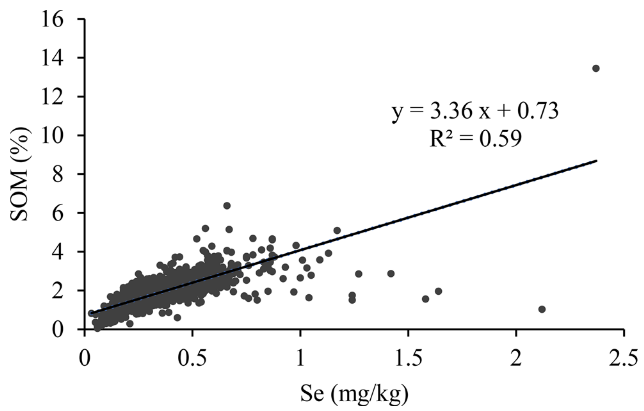

Fig. 11. Correlation between soil Se and SOM.

Table 10. *SOM and soil Se content*

SOM (%)	0.10–1.10	1.10–1.40	1.40–1.66	1.66–1.92	1.92–2.14
Se (mg kg^{-1})	0.15	0.19	0.22	0.27	0.34
SOM (%)	2.14–2.44	2.44–2.85	2.85–3.48	3.48–4.70	4.70–9.57
Se (mg kg^{-1})	0.40	0.49	0.58	0.68	1.03

various anthropogenic input sources (Li 2010; Zhao 2020), indicating that coal burning emissions are an important source of soil Se.

In general, the rocks in the Taihang Mountains are the material basis of the soil in the Shijiazhuang–Xingtai–Handan area. In the process of soil formation, the influence of physicochemical properties on the soil Se content is gradually enhanced, and SOM is the most important factor. Socioeconomic factors promote the spatial differentiation of soil Se, and coal burning is the most important nonnatural influencing factor.

Conclusion

(1) The topsoil Se content in Hebei Province ranged from 0.02 to 3.23 mg kg^{-1}, with an average of 0.20 mg kg^{-1}, which is lower than the Chinese background value of 0.29 mg kg^{-1}. The Se contents of the 11 cities are different. Shijiazhuang has the highest Se content at 0.26 mg kg^{-1}, while Zhangjiakou has the lowest at 0.16 mg kg^{-1}.

(2) The spatial distribution of soil Se in Hebei Province is heterogeneous; the marginal Se area and sufficient Se area are the most extensive, but the Se-rich area is very limited, accounting for only 2.43% of the total survey area, whereas the distribution of Se-rich areas is relatively concentrated. The Shijiazhuang–Xingtai–Handan area is the largest and most concentrated Se-rich area with a content that reaches 0.52 mg kg^{-1}.

(3) The soil Se of the Shijiazhuang–Xingtai–Handan area originally comes from the rocks in the Taihang Mountain area, and its spatial distribution is influenced by natural and socioeconomic factors. SOM is the most important natural factor, which has a strong adsorption capacity for soil Se and has the strongest influence among all factors. Coal burning is the most important socioeconomic factor and the main anthropogenic source of soil Se. It acts together with other factors to promote the spatial differentiation of soil Se.

Scientific editing by Scott Wood

Acknowledgments We would like to thank AJE (https://www.aje.cn/) for its linguistic assistance during the preparation of this manuscript. We would also like to express our sincere thanks to the editors and reviewers for their critical and constructive comments and suggestions.

Author contributions

LZ: data curation (lead), formal analysis (lead), investigation (equal), methodology (lead), writing – original draft (lead), writing – review & editing (lead); **WX**: funding acquisition (lead), investigation (lead), project administration (lead), resources (lead); **WY**: data curation (equal), investigation (equal), resources (equal); **LZ**: writing – original draft (equal), writing – review & editing (equal), **SZ**: formal analysis (supporting), visualization (lead); **XW**: investigation (supporting); **XX**: investigation (supporting); **HC**: conceptualization (lead), methodology (equal), writing – review & editing (equal)

Funding

This work was funded by the China Geological Survey (200414200007, 12120113000800, GZTR20080102, DD20160325-01, DD20221770 and DD20221770-04) and Institute of Geophysical and Geochemical Exploration, Chinese Academy of Geological Sciences (AS2022J09).

Competing interests

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.

Data availability

Data sharing is not applicable to this article as no datasets were generated or analysed during the current study.

References

A.A., E.-S., Seeda, M.A., Yassen, A., Khater, A. and Zaghloul, S.M. 2020. Selenium behavior in the soil, water, plants and its implication for human health. A review. *Current Science International*, **9**, 173–197, https://doi.org/10.36632/csi/2020.9.2.17

Breiman, L. 2001. Random forest. *Machine Learning*, **45**, 5–32, https://doi.org/10.1023/A:1010933404324

Bujdoš, M., Mul'ová, A., Kubová, J. and Medved', J. 2005. Selenium fractionation and speciation in rocks, soils, waters and plants in polluted surface mine environment. *Environmental Geology*, **47**, 353–360, https://doi.org/10.1007/s00254-004-1157-2

Cambardella, C.A., Moorman, T.B., Novak, J.M., Parkin, T.B., Karlen, D.L., Turco, R.F. and Konopka, A.E. 1994. Field-scale variability of soil properties in central iowa soils. *Soil Science Society of America Journal*, **58**, 1501–1511, https://doi.org/10.2136/sssaj1994.03615995005800050033x

Chen, Y.M., Feng, S.Q., Yang, T.X., Zhang, W.Q. and Wang, S.B. 2013. Statistical characteristics of organic matter content from different soil classes in China. *Journal of Fudan University(Natural Science,)*, **52**, 220–224, https://doi.org/10.15943/j.cnki.fdxb-jns.2013.02.012 [in Chinese with English abstract].

Fernández-Martínez, A. and Charlet, L. 2009. Selenium environmental cycling and bioavailability: a structural chemist point of view. *Reviews in Environmental Science & Bio/technology*, **8**, 81–110, https://doi.org/10.1007/s11157-009-9145-3

Fordyce, F.M., Zhang, G.D., Green, K. and Liu, X.P. 2000. Soil, grain and water chemistry in relation to human selenium-responsive diseases in Enshi District, China. *Applied Geochemistry*, **15**, 117–132, https://doi.org/10.1016/S0883-2927(99)00035-9

Fordyce, F.M., Brereton, N., Hughes, J., Luo, W. and Lewis, J. 2010. An initial study to assess the use of geological parent materials to predict the Se concentration in overlying soils and in five staple foodstuffs produced on them in Scotland. *Science of the Total Environment*, **408**, 5295–5305, https://doi.org/10.1016/j.scitotenv.2010.08.007

Gabos, M.B., Alleoni, L.R.F. and Abreu, C.A. 2014. Background levels of selenium in some selected Brazilian tropical soils. *Journal of Geochemical Exploration*, **145**, 35–39, https://doi.org/10.1016/j.gexplo.2014.05.007

Gong, J.Z. 1997. *Report of Regional Petrogeochemical Survey in Hebei Province*. Hebei Geophysical Survey Institute, https://doi.org/10.35080/n01.c.93352 [in Chinese].

Grimm, R., Behrens, T., Märker, M. and Elsenbeer, H. 2008. Soil organic carbon concentrations and stocks on Barro Colorado Island – digital soil mapping using Random Forests analysis. *Geoderma*, **146**, 102–113, https://doi.org/10.1016/j.geoderma.2008.05.008

Hou, Q.Y., Yang, Z.F., Yu, T., Xia, X.Q., Cheng, H.X. and Zhou, G.H. 2020. *Geochemical Parameters of China*. Geological Publishing House, Beijing [in Chinese].

Jia, X.L., Fu, T.T., Hu, B.F., Shi, Z., Zhou, L.Q. and Zhu, Y.W. 2020. Identification of the potential risk areas for soil heavy metal pollution based on the source-sink theory. *Journal of Hazardous Materials*, **393**, 122424, https://doi.org/10.1016/j.jhazmat.2020.122424

Li, Z.N. 2010. *The Study on Source Tracking of Se Anomaly and Ecological Appraisal in Plain Terrain of Hebei*. Master thesis, Shijiazhuang University of Economics [in Chinese with English abstract].

Li, Y.H., Wang, W.Y., Luo, K.L. and Li, H.R. 2008. Environmental behaviors of selenium in soil of typical selenosis area, China. *Journal of Environmental Sciences*, **20**, 859–864, https://doi.org/10.1016/S1001-0742(08)62138-5

Li, S.H., Xiao, T.F. and Zheng, B.S. 2012. Medical geology of arsenic, selenium and thallium in China. *Science of the Total Environment*, **421–422**, 31–40, https://doi.org/10.1016/j.scitotenv.2011.02.040

Li, M., Xi, X. *et al.* 2014. National multi-purpose regional geochemical survey in China. *Journal of Geochemical Exploration*, **139**, 21–30, https://doi.org/10.1016/j.gexplo.2013.06.002

Li, Z., Liang, D., Peng, Q., Cui, Z., Huang, J. and Lin, Z. 2017. Interaction between selenium and soil organic matter and its impact on soil selenium bioavailability: a review. *Geoderma*, **295**, 69–79, https://doi.org/10.1016/j.geoderma.2017.02.019

Liu, Y.L., Yuan, Y.Y. and Luo, K.L. 2018. Regional distribution of longevity population and elements in drinking water in Jiangjin District, Chongqing City, China. *Biological Trace Element Research*, **184**, 287–299, https://doi.org/10.1007/s12011-017-1159-z

Liu, H.L., Wang, X.Q. *et al.*. 2021. Concentration and distribution of selenium in soils of mainland China, and implications for human health. *Journal of Geochemical Exploration*, **220**, 106654, https://doi.org/10.1016/j.gexplo.2020.106654

Liu, Y.L., Tian, X.L., Liu, R., Liu, S.L. and Zuza, A. V. 2021. Key driving factors of selenium-enriched soil in the low-Se geological belt: a case study in Red Beds of Sichuan Basin, China. *Catena*, **196**, 104926, https://doi.org/10.1016/j.catena.2020.104926

López-Bellido, F.J., Sanchez, V., Rivas, I., López-Bellido, R.J. and López-Bellido, L. 2019. Wheat grain selenium content as affected by year and tillage system in a rainfed Mediterranean Vertisol. *Field Crops Research*, **233**, 41–48, https://doi.org/10.1016/j.fcr.2019.01.006

Luo, K.L., Xu, L.R., Tan, J.A., Wang, D.H. and Xiang, L.H. 2004. Selenium source in the selenosis area of the Daba region, South Qinling Mountain, China. *Environmental Geology*, **45**, 426–432, https://doi.org/10.1007/s00254-003-0893-z

Malisa, E.P. 2001. The behaviour of selenium in geological processes. *Environmental Geochemistry and Health*, **23**, 137–158, https://doi.org/10.1023/A:1010908615486

Mamat, Z., Yimit, H., Ji, R.Z.A. and Eziz, M. 2014. Source identification and hazardous risk delineation of heavy metal contamination in Yanqi basin, northwest China. *Science of the Total Environment*, **493**, 1098–1111, https://doi.org/10.1016/j.scitotenv.2014.03.087

Martín, R.J.A., Arias, M.L. and Corbí, G.J.M. 2006. Heavy metals contents in agricultural topsoils in the Ebro basin (Spain). Application of the multivariate geoestatistical methods to study spatial variations. *Environmental Pollution*, **144**, 1001–1012, https://doi.org/10.1016/j.envpol.2006.01.045

Matos, R.P., Lima, V.M.P., Windmöller, C.C. and Nascentes, C.C. 2017. Correlation between the natural levels of selenium and soil physicochemical characteristics from the Jequitinhonha Valley (MG), Brazil. *Journal of Geochemical Exploration*, **172**, 195–202, https://doi.org/10.1016/j.gexplo.2016.11.001

Natasha, M.S., Niazi, N.K., Khalid, S., Murtaza, B., Bibi, I. and Rashid, M.I. 2018. A critical review of selenium biogeochemical behavior in soil-plant system with an inference to human health. *Environmental Pollution*, **234**, 915–934, https://doi.org/10.1016/j.envpol.2017.12.019

Nezhad, M.T.K., Tabatabaii, S.M. and Gholami, A. 2015. Geochemical assessment of steel smelter-impacted urban soils, Ahvaz, Iran. *Journal of Geochemical Exploration*, **152**, 91–109, https://doi.org/10.1016/j.gexplo.2015.02.005

Ngigi, P.B., Du, L.G., Masinde, P.W. and Lachat, C. 2020. Selenium deficiency risk in central Kenya highlands: an assessment from the soil to the body. *Environmental Geochemistry and Health*, **42**, 2233–2250, https://doi.org/10.1007/s10653-019-00494-1

Ni, R., Luo, K., Tian, X., Yan, S., Zhong, J. and Liu, M. 2016. Distribution and geological sources of selenium in environmental materials in Taoyuan County, Hunan Province, China. *Environmental Geochemistry and Health*, **38**, 927–938, https://doi.org/10.1007/s10653-015-9772-2

Nriagu, J.O. and Pacyna, J.M. 1988. Quantitative assessment of worldwide contamination of air, water and soils by trace metals. *Nature*, **333**, 134–139, https://doi.org/10.1038/333134a0

Oliver, M.A. and Webster, R. 2014. A tutorial guide to geostatistics: computing and modelling variograms and kriging. *Catena*, **113**, 56–69, https://doi.org/10.1016/j.catena.2013.09.006

Qin, J.F. 2014. Historical review on the research of selenium in China. *Guangdong Trace Elements Science*, **21**, 36–51, https://doi.org/10.16755/j.cnki.issn.1006-446x.2014.12.009 [in Chinese with English abstract].

Ren, H.L. and Yang, R.D. 2014. Distribution and controlling factors of selenium in weathered soil in Kaiyang County, Southwest China. *Chinese Journal of Geochemistry*, **33**, 300–309, https://doi.org/10.1007/s11631-014-0691-1

Roca-Perez, L., Gil, C. *et al.* 2010. Selenium and heavy metals content in some Mediterranean soils. *Journal of Geochemical Exploration*, **107**, 110–116, https://doi.org/10.1016/j.gexplo.2010.08.004

Shaheen, S.M., Rinklebe, J., Frohne, T., White, J.R. and Delaune, R.D. 2014. Biogeochemical Factors Governing Cobalt, Nickel, Selenium, and Vanadium Dynamics in Periodically Flooded Egyptian North Nile Delta Rice Soils. *Soil Science Society of America Journal*, **78**, 1065–1078, https://doi.org/10.2136/sssaj2013.10.0441

Smith, P. 2008. Land use change and soil organic carbon dynamics. *Nutrient Cycling in Agroecosystems*, **81**, 169–178, https://doi.org/10.1007/s10705-007-9138-y

Song, T.J., Cui, G., Su, X.S., He, J., Tong, S.Z. and Liu, Y. 2020. The origin of soil selenium in a typical agricultural area in Hamatong River Basin, Sanjiang Plain, China. *Catena*, **185**, 104355, https://doi.org/10.1016/j.catena.2019.104355

Spadoni, M., Voltaggio, M., Carcea, M., Coni, E., Raggi, A. and Cubadda, F. 2007. Bioaccessible selenium in Italian agricultural soils: comparison of the biogeochemical approach with a regression model based on geochemical and pedoclimatic variables. *Science of the Total Environment*, **376**, 160–177, https://doi.org/10.1016/j.scitotenv.2007.01.066

Spallholz, J.E., Mallory Boylan, L., David Robertson, J., Smith, L., Rahman, M.M., Hook, J. and Rigdon, R. 2008. Selenium and arsenic content of agricultural soils from Bangladesh and Nepal. *Toxicological and Environmental Chemistry*, **90**, 203–210, https://doi.org/10.1080/02772240701419347

Supriatin, S., Weng, L.P. and Comans, R.N.J. 2015. Selenium speciation and extractability in Dutch agricultural soils. *Science of the Total Environment*, **532**, 368–382, https://doi.org/10.1016/j.scitotenv.2015.06.005

Tan, J.A. and Huang, Y.J. 1991. Selenium in geo-ecosystem and its relation to endemic diseases in China. *Water Air & Soil Pollution*, **57–58**, 59–68, https://doi.org/10.1007/BF00282869

Tan, J.A., Zhu, W.Y., Wang, W.Y., Li, R.B., Hou, S.F., Wang, D.C. and Yang, L.S. 2002. Selenium in soil and endemic diseases in China. *Science of the Total Environment*, **284**, 227–235, https://doi.org/10.1016/S0048-9697(01)00889-0

Tian, H., Ma, Z. *et al.* 2016. Geochemical characteristics of selenium and its correlation to other elements and minerals in selenium-enriched rocks in Ziyang County, Shaanxi Province, China. *Journal of Earth Science*, **27**, 763–776, https://doi.org/10.1007/s12583-016-0700-x

Tuttle, M.L.W., Fahy, J.W., Elliott, J.G., Grauch, R.I. and Stillings, L.L. 2014. Contaminants from cretaceous black shale: II. Effect of geology, weathering, climate, and land use on salinity and selenium cycling, Mancos Shale landscapes, southwestern United States. *Applied Geochemistry*, **46**, 72–84, https://doi.org/10.1016/j.apgeochem.2013.12.011

Wang, Z.J. and Gao, Y.X. 2001. Biogeochemical cycling of selenium in Chinese environments. *Applied Geochemistry*, **16**, 1345–1351, https://doi.org/10.1016/S0883-2927(01)00046-4

Wang, J.F. and Xu, C.D. 2017. Geodetector: Principle and prospective. *Acta Geographica Sinica*, **72**, 116–134, https://doi.org/10.11821/dlxb201701010 [in Chinese with English abstract].

Wang, J.F., Li, X.H., Christakos, G., Liao, Y.L., Zhang, T., Gu, X. and Zheng, X.Y. 2010. Geographical detectors-based health risk assessment and its application in the neural tube defects study of the Heshun Region, China. *International Journal of Geographical Information Science*, **24**, 107–127, https://doi.org/10.1080/13658810802443457

Wang, L., Ju, Y., Liu, G., Chou, C.L., Zheng, L. and Qi, C. 2010. Selenium in Chinese coals: distribution, occurrence, and health impact. *Environmental Earth Sciences*, **60**, 1641–1651, https://doi.org/10.1007/s12665-009-0298-8

Wang, J., Li, H.R., Li, Y.H., Yu, J.P., Yang, L.S., Feng, F.J. and Chen, Z. 2013. Speciation, distribution, and bioavailability of soil selenium in the tibetan plateau kashin-beck disease area – a case study in Songpan County, Sichuan Province, China. *Biological Trace Element Research*, **156**, 367–375, https://doi.org/10.1007/s12011-013-9822-5

Wang, J.F., Zhang, T.L. and Fu, B.J. 2016. A measure of spatial stratified heterogeneity. *Ecological Indicators*, **67**, 250–256, https://doi.org/10.1016/j.ecolind.2016.02.052

Wang, D., Dinh, Q.T. *et al.* 2018. Effect of selenium-enriched organic material amendment on selenium fraction transformation and bioavailability in soil. *Chemosphere*, **199**, 417–426, https://doi.org/10.1016/j.chemosphere.2018.02.007

Wei, F.S., Chen, J.S., Wu, Y.Y. and Zheng, C.J. 1991. Study on soil environmental background value in China. *Environmental Science*, **12**, 12–19 [in Chinese].

Wen, H.J. and Carignan, J. 2007. Reviews on atmospheric selenium: emissions, speciation and fate. *Atmospheric Environment*, **41**, 7151–7165, https://doi.org/10.1016/j.atmosenv.2007.07.035

Weng, L.P., Vega, F.A., Supriatin, S., Bussink, W. and Van Riemsdijk, W.H. 2011. Speciation of Se and DOC in soil solution and their relation to Se bioavailability. *Environmental Science and Technology*, **45**, 262–267, https://doi.org/10.1021/es1016119

Xiao, K.C., Tang, J.J., Chen, H., Li, D.J. and Liu, Y.X. 2020. Impact of land use/land cover change on the topsoil selenium concentration and its potential bioavailability in a karst area of southwest China. *Science of The Total Environment*, **708**, 135201, https://doi.org/10.1016/j.scitotenv.2019.135201

Xie, H.L., Wang, P. and Huang, H.S. 2013. Ecological risk assessment of land use change in the poyang lake eco-economic zone, China. *International Journal of Environmental Research and Public Health*, **10**, 328–346, https://doi.org/10.3390/ijerph10010328

Yamada, H., Kamada, A., Usuki, M. and Yanai, J. 2009. Total selenium content of agricultural soils in Japan. *Soil Science and Plant Nutrition*, **55**, 616–622, https://doi.org/10.1111/j.1747-0765.2009.00397.x

Yang, S.C. 2006. Coal resources exploration and exploitation status quo in Hebei province. *Coal Geology of China*, **18**, 5–10, https://doi.org/10.3969/j.issn.1674-1803.2006.04.002 [in Chinese with English abstract].

Yang, G.Q., Wang, S.Z., Zhou, R.H. and Sun, S.Z. 1983. Endemic selenium intoxication of humans in China. *The American Journal of Clinical Nutrition*, **37**, 872–881, https://doi.org/10.1093/ajcn/37.5.872

Yang, P.G., Mao, R.Z., Shao, H.B. and Gao, Y.F. 2009. The spatial variability of heavy metal distribution in the suburban farmland of Taihang Piedmont Plain, China. *Comptes Rendus - Biologies*, **332**, 558–566, https://doi.org/10.1016/j.crvi.2009.01.004

Yu, H.C., Duan, D.J., Xu, J. and Zhao, J. 2010. Occurrence and enrichment of trace elements Pb and Se of 5 # coal seam from Gequan mine, Xingtai coalfield. *Journal of Hebei University of Engineering (Natural Science Edition)*, **27**, 54–56, https://doi.org/10.3969/j.issn.1673-9469.2010.03.014 [in Chinese with English abstract].

Yu, T., Yang, Z. *et al.* 2014. The origin and geochemical cycle of soil selenium in a Se-rich area of China. *Journal of Geochemical Exploration*, **139**, 97–108, https://doi.org/10.1016/j.gexplo.2013.09.006

Yuan, L., Yin, X., Zhu, Y., Li, F., Huang, Y., Liu, Y. and Lin, Z. 2012. Selenium in Plants and Soils, and Selenosis in Enshi, China: Implications for Selenium Biofortification. *In:* Yin, X. and Yuan, L. (eds) *Phytoremediation and Biofortification: Two Sides of One Coin*. Springer, Netherlands, Dordrecht, 7–31, https://doi.org/10.1007/978-94-007-1439-7_2

Zhang, M.M. and Sun, Y.Z. 2021. Source of selenium in Handan geochemical anomaly belt: evidences from petrology and geochemistry of Upper Paleozoic in western Handan, China. *Journal of Geochemical Exploration*, **226**, 106770, https://doi.org/10.1016/j.gexplo.2021.106770

Zhang, L., McKinley, J.M., Cooper, M., Peng, M., Wang, Q., Song, Y. and Cheng, H. 2020. A regional soil and river sediment geochemical study in Baoshan area, Yunnan province, southwest China. *Journal of Geochemical Exploration*, **217**, 106557, https://doi.org/10.1016/j.gexplo.2020.106557

Zhao, Y. 2020. *Study on the Genesis and Utilization of Selenium Rich Soil in Handan and Xingtai Area*. Master thesis, Hebei GEO University [in Chinese with English abstract].

Zhao, Y., Luan, W., Lou Guo, H.Q., Cai, K., Ma, Z.S. and Dun, Y.R. 2021. Characteristics, causes and ecological environment health evaluation of Selenium – enriched soil in Gaocheng District, Shijiazhuang City, Hebei Province. *Geology in China*, **48**, 764–776, https://doi.org/10.12029/gc20210307 [in Chinese with English abstract].

Zhu, J.M., Wang, N., Li, S.H., Li, L., Su, H.C. and Liu, C.X. 2008. Distribution and transport of selenium in Yutangba, China: impact of human activities. *Science of the Total Environment*, **392**, 252–261, https://doi.org/10.1016/j.scitotenv.2007.12.019