Spatial Distribution and Associated Risk Assessment of Heavy Metal Pollution in Farmland Soil Surrounding the Ganhe Industrial Park in Qinghai Province, China

Fang Yin 1, Wenrui Meng 2, Lei Liu 2,⁎, Kai Feng 3 and Cuijing Yin 4

1 Shaanxi Key Laboratory of Land Consolidation, School of Land Engineering, Chang’an University, Xi’an 710054, China; yinf@chd.edu.cn
2 School of Earth Science and Resources, Chang’an University, Xi’an 710054, China; 2022127011@chd.edu.cn
3 Qingdao Hisense School, Qingdao 266000, China; fengkai11@hisense.com
4 Xi’an Meihang Remote Sensing Information Co., Ltd., Xi’an 710199, China; 2019127014@chd.edu.cn
⁎ Correspondence: liul@chd.edu.cn; Tel.: +86-29-82339936

Abstract: The farmland around the industrial areas in the Upper Yellow River is crucial for agricultural production but is vulnerable to contamination from the surrounding industries. This research focused on analyzing the spatial distribution and environmental risks of heavy metal pollution in the farmland around the Ganhe Industrial Park in the Qinghai–Tibet Plateau. A total of 138 surface soil samples were collected, and the concentration of seven heavy metals (Cd, As, Pb, Cr, Cu, Ni, and Zn) was analyzed using the random forest (RF) model. Pollution indicators, including the pollution index and Nemero index, were used to evaluate the pollution levels of soil heavy metals. The human health and ecological risks were estimated using the hazard index (HI) and the potential ecological risk index (RI). Cd and Zn were identified as the primary soil pollutants in the study area, with Cd being more concentrated than other heavy metals. Heavy metal contamination was most severe in the central–eastern region of the study area, with a ring-shaped distribution, which correlated with the presence of zinc smelting and chemical plants. Furthermore, the study revealed that soil heavy metal contamination posed a health threat to the local population, with children being particularly vulnerable to non-carcinogenic risks when the HI was 1.21 and to potential carcinogenic risks when the CR was $2.27 \times 10^{-5}$. Additionally, heavy metal pollution caused a moderate to high ecological risk in 56.4% of the samples. The results highlighted the severe impact of soil heavy metal pollution on the delicate ecosystem of the Upper Yellow River and Qinghai–Tibet Plateau. The government should take action to improve soil environment management and prevent heavy metal pollution to protect the health of the local population and the ecological environment.

Keywords: spatial characteristics; random forest model; risk evaluation; pollution index; soil heavy metals; Qinghai–Tibet plateau

1. Introduction

Soil heavy metal pollution is considered to be one of the most serious environmental problems that can lead to health risks for humans and animals [1–6]. Heavy metals could enter the soil through various pathways, including bedrock weathering, precipitation, dust settlement, and waste discharge [7–9]. Wastes discharged by anthropogenic activities, such as mining, metal processing and smelting, chemical production, factory emissions, and sewage irrigation, are major sources of heavy metal pollution in soil [10–13]. Prior research has demonstrated that the elevated concentration of heavy metals in soil poses a threat to crops, animals, and human health through direct exposure (such as ingestion, dermal absorption, and inhalation) or via food chains [14–21].

Heavy metals in soils are difficult to degrade and decontaminate [8]. The presence of heavy metals in soils is a serious concern for soil environment management and heavy
metal pollution prevention. Several statistics-based methods have been developed to evaluate the soil heavy metal pollution degree, including pollution index (Pi), geo-accumulation index (Igeo), and Nemero index (P), which use classification criteria to assess pollution levels [8,22–24]. These methods could consider the influence of natural geological effects and anthropogenic activities on heavy metals in soil, with the metal background value as a parameter [25,26]. However, the relationship between the concentrations and various predictor variables often reveals non-linear features, making it challenging to predict the spatial distribution and influence factors of soil heavy metals. The random forest model (RF) has been used to overcome this challenge and predict the spatial distribution and influence factors of soil heavy metals [27]. In addition to pollution evaluation, risk assessment is also essential for heavy metal contamination. The potential ecological risk assessment (RI) is commonly used to assess the risk posed to the ecological environment and human health [8,28–30], while the Health Risk Assessment (HRA) is used to quantitatively assess the carcinogenic and non-carcinogenic risks of human exposure to certain contaminants [3,28].

The combination of these methods can provide a more comprehensive evaluation of the potential ecological and health risk of toxic metals in soil. However, most previous studies were carried out mainly in floodplains, ports, mining areas, or urban–rural ecotones, with few studies conducted in industrial park areas with high altitudes [1–3,11,12]. High-altitude areas have different characteristics for the migration and transformation of heavy metals, including low temperature, underdeveloped soil, and sensitive ecosystem, making it necessary to study heavy metal contamination in these areas as well.

The Upper Yellow River is located in the north–eastern Qinghai–Tibet Plateau, which is an area that makes up 16.2% of the total area of the Yellow River Basin [24]. The area has an average altitude of approximately 3500 m and is known as the “third pole” of the Earth due to its sensitivity to climate change in Asia and the Northern Hemisphere [31]. The most widely distributed non-zonal vegetation types in the region are alpine meadow and alpine steppe [31]. As the largest industrial park in the Upper Yellow River, the high-pollution factories in the Ganhe Industrial Park may pose a threat to the residents of the whole county, the farmland, and even the ecosystem of the Upper Yellow River and Qinghai–Tibet Plateau, due to the release of contaminants to the water, air, and soil (Figure 1). Therefore, this study aims to quantitatively evaluate the pollution status and potential risks of heavy metal concentrations in Ganhe Industrial Park by applying various methods, such as the random forest model, pollution index, Nemero index, potential ecological risk assessment, and health risk assessment [32,33]. The study not only investigated the relative environmental, economic, and natural conditions in the study area but also estimated human health and ecological risks by using the hazard index (HI) and potential ecological risk index (RI) to provide theoretical support for contamination management for the industrial park and even for the Upper Yellow River. The findings of the study could confirm the extent of heavy metal pollution in the study area, as well as the potential risks associated with this pollution, to maintain a sustainable and healthy agricultural system in the Upper Yellow River, which could also be used in similar regions in China and other developing countries.
2. Materials and Methods

2.1. Study Area

The study area is located in the north-eastern Qinghai Province, which is a typical temperate continental climate region characterized by hot summers, cold winters, and low annual precipitation. The area covers approximately 200 km$^2$, with coordinates ranging from 36°29′49″ N–36°36′47″ N to 101°25′44″ E–101°35′2″ E (Figure 1). The elevation gradually declines from south to north, ranging from 3600 m to 2100 m, and the terrain belongs to the Qinghai–Tibet Plateau.

The land use types are dominated by farmland, grassland, and construction land, and the farmland in this study’s area is mainly used for growing grain crops (Figure 1). The soil types in the area include Chernozem, Kastanozem, Calcisols, and Cryosols, according to the classification system of the Food and Agriculture Organization (FAO) [34]. The residential and factory estates extend along two parallel drainages in the north-south direction, interspaced by farmlands and pastures. The factories include various metals (aluminum,
copper, lead, zinc, iron), smelting plants, chemical plants, inorganic salt manufacturing plants, etc. The soil in the study area is heavily polluted with heavy metals, mainly as a result of industrial contaminants released into the air, water, and soil, as well as excessive use of pesticides and fertilizers.

2.2. Sampling and Analysis

A total of 138 soil samples were collected between 29 September 2020 and 3 October 2020. Semi-random distribution method was used for selecting the sampling points, taking into account the accessibility and spatial coverage of the study area (Figure 1). For each sampling point, a 20 m × 20 m square was built in the center, and five samples were collected from the four corners and the center of each square. The collected soil samples were mixed to obtain a composite sample for each sampling point. The soil samples were collected from a depth of 0–20 cm from the natural surface and were approximately 2 kg in weight. Rocks and debris were removed from the samples. The soil samples were air-dried, crushed, and sieved through 200 mesh polyethylene for analysis. The heavy metal concentrations of Cd, As, Pb, Cr, Cu, Ni, and Zn were analyzed using inductively coupled plasma mass spectrometry (ICP-MS, Thermo Fisher X7, Bremen, Germany) instruments following digestion with an HCl-H2O-HF-HNO3 mixture [35]. The analysis was conducted in a laboratory with an environment temperature of 24 °C and a humidity of 40%. The ICP-MS detection limits for all elements are calculated as three times the standard deviation of the calibration blank measurements (1:1 v/v HNO3: MQ-water); instrumental stability and tuning were checked using a solution of 10 µg/L of Li, Be, Bi, Ce, Co, In, Pb and U in HNO3 (2% v/v), and 156CeO+/140Ce+<0.3%. Internal standardization was performed using 103Rh for all elements. For ICP-MS determination, certified reference materials (GSS-3a and Gss-5a), blank, and samples were measured for quality control. The GSS-3a and Gss-5a were used for the validation of the measurement. The measured results were compared with the certified reference values, and the results were within the 95% confidence level. Moreover, repeated analysis of samples from different sampling areas was adopted, and the results of duplicate samples were consistent within the margin of error.

2.3. Risk Assessment Methods

2.3.1. Random Forest Model

The RF model was used to establish the non-linear relationship between the independent variables and the dependent variable [27]. The independent variables include factory distance, slope, road distance, annual rainfall, average annual temperature, and elevation of the samples, while the dependent variable is heavy metals in the soil.

The model-building process involved several steps [27]. Firstly, K bootstrap samples were selected using the self-expanding sampling method to train the regression tree. Then, m subsets were randomly extracted from M feature values, and the best subset was selected when the tree was split (m ≤ M). Each decision tree was built without pruning and allowed to grow to the maximum extent. Finally, the prediction results were obtained through voting.

In the modeling process, the research area was divided into 900 (30 × 30) rectangular grids with a grid size of 500 m, ensuring that the number of training grids was greater than 10% of the total grid. Sample points located in the same grid were eliminated if they had the same content values, resulting in only 136 sample points being used. The optimal parameters and accuracy of the training sets corresponding to verification sets are shown in Table 1, which was generated during the model evaluation process.

The sample dataset of 136 points was split into 70% for modeling and 30% for verification. A Python program was developed using ArcGIS Pro 2.5 to optimize the model parameters in order of the number of trees, the maximum tree depth, the number of random sampling variables, the minimum leaf size, and the accuracy of the training, and verification sets was measured (Table 1).
Table 1. Parameters and accuracy of random forest model.

<table>
<thead>
<tr>
<th>Heavy Metal</th>
<th>Tree Number</th>
<th>Minimum of Leaf Size</th>
<th>Maximum of Tree Depth</th>
<th>Number of Random Sampling</th>
<th>$R^2$ Modeling Set</th>
<th>$R^2$ Validation Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cd</td>
<td>61</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>0.856</td>
<td>0.634</td>
</tr>
<tr>
<td>As</td>
<td>195</td>
<td>7</td>
<td>2</td>
<td>4</td>
<td>0.571</td>
<td>0.399</td>
</tr>
<tr>
<td>Pb</td>
<td>160</td>
<td>6</td>
<td>1</td>
<td>3</td>
<td>0.692</td>
<td>0.518</td>
</tr>
<tr>
<td>Cr</td>
<td>100</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>0.923</td>
<td>0.699</td>
</tr>
<tr>
<td>Cu</td>
<td>151</td>
<td>10</td>
<td>1</td>
<td>5</td>
<td>0.545</td>
<td>0.570</td>
</tr>
<tr>
<td>Ni</td>
<td>194</td>
<td>4</td>
<td>7</td>
<td>3</td>
<td>0.919</td>
<td>0.796</td>
</tr>
<tr>
<td>Zn</td>
<td>160</td>
<td>6</td>
<td>3</td>
<td>3</td>
<td>0.883</td>
<td>0.616</td>
</tr>
</tbody>
</table>

2.3.2. Pollution Index

The pollution index is a single-factor assessment method used to evaluate heavy metal pollution and the associated pollution risk [10,36]. It can be calculated by dividing the actual concentration of a specific heavy metal by the corresponding background values of the same metal in the soil, as indicated in the provided Equation (1)

$$P_i = \frac{C_i}{S_i}$$

where $P_i$ represents the pollution index of the heavy metal $i$; $C_i$ represents the chemical analyzed value of soil heavy metal $i$; and $S_i$ represents the standard value of soil heavy metal quality. The standard value of soil heavy metal quality is determined by using the screening values defined in the Soil Environmental Quality Standard (GB 15618-2018) issued by the Chinese Ministry of Environmental Protection in 2018 [37].

A higher value of $P_i$ indicates more severe contamination of the soil. When $P_i$ is less than or equal to 1, the soil is considered uncontaminated. A $P_i$ value between 1 and 2 indicates that the soil is low to moderately contaminated, while a $P_i$ value between 2 and 3 indicates that the soil is moderate to heavily contaminated. Finally, if $P_i$ is greater than 3, the soil is considered heavy to extremely contaminated [8].

2.3.3. Nemero Index

The Nemero index is a useful tool for assessing the environmental risk of pollutants, specifically heavy metals, in soil. The index takes into account the different toxic effects of each heavy metal and calculates a comprehensive measure of the environmental risk based on their concentrations. The formula for calculating the Nemero index is as follows:

$$P = \sqrt{\left(\frac{1}{n} \sum \frac{C_i}{S_i}\right)^2 + \left(\frac{C_i}{S_i}\right)_{\text{MAX}}^2}$$

where $P$ is the Nemero index; $\left(\frac{C_i}{S_i}\right)_{\text{MAX}}^2$ is the maximum value of the pollution index of the heavy metal; $\left(\frac{1}{n} \sum \frac{C_i}{S_i}\right)^2$ is the average value of the individual pollution index. The quality of the soil environment is classified into 5 levels based on the Nemero index, according to Table 2 [38].

Table 2. Grading criteria for the Nemero index [38].

<table>
<thead>
<tr>
<th>Grade</th>
<th>$P$</th>
<th>Pollution Levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>$P \leq 0.7$</td>
<td>Clean (safe level)</td>
</tr>
<tr>
<td>II</td>
<td>$0.7 &lt; P \leq 1$</td>
<td>Warning limit (precaution level)</td>
</tr>
<tr>
<td>III</td>
<td>$1 &lt; P \leq 2$</td>
<td>Slight pollution (slightly polluted level)</td>
</tr>
<tr>
<td>IV</td>
<td>$2 &lt; P \leq 3$</td>
<td>Moderate pollution (moderately polluted level)</td>
</tr>
<tr>
<td>V</td>
<td>$P &gt; 3$</td>
<td>Heavy pollution (seriously polluted level)</td>
</tr>
</tbody>
</table>
2.3.4. Potential Health Risk Assessment

Health Risk Appraisal (HRA) is a quantitative method used to assess the human risk of exposure to environmental pollutants, including heavy metals [39,40]. Ingestion, inhalation, and dermal absorption are the three primary pathways through which heavy metals can enter the human body [41,42]. According to the model provided by the US Environmental Protection Agency (USEPA), the chronic daily intake (CDI) of the seven heavy metals through multiple exposure pathways was calculated using the following equations:

\[
\text{CDI}_{\text{ingest}} = \frac{C_{\text{soil}} \times \text{IngR} \times \text{EF} \times \text{ED} \times \text{CF}}{\text{BW} \times \text{AT}} \quad (3)
\]

\[
\text{CDI}_{\text{inhal}} = \frac{C_{\text{soil}} \times \text{InhR} \times \text{EF} \times \text{ED}}{\text{PEF} \times \text{BW} \times \text{AT}} \quad (4)
\]

\[
\text{CDI}_{\text{dermal}} = \frac{C_{\text{soil}} \times \text{SA} \times \text{AF} \times \text{ABS} \times \text{EF} \times \text{ED} \times \text{CF}}{\text{BW} \times \text{AT}} \quad (5)
\]

where \( C_{\text{soil}} \) represents the average of the measured concentration of each soil heavy metal for all the samples in the study area (mg \( \cdot \) kg\(^{-1} \)). CDI\(_{\text{ingest}}\), CDI\(_{\text{inhal}}\), and CDI\(_{\text{dermal}}\) represent the potential intake value of each heavy metal through ingest, inhale, and dermal absorption. The other parameters in Equation (3)–(5) are shown in Table 3.

Table 3. Parameters used in CDI estimation for non-carcinogenic risk and carcinogenic risk (adapted from US Environmental Protection Agency [43]).

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Units</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Adults</td>
<td>Children</td>
</tr>
<tr>
<td>ABS</td>
<td>(1.00 \times 10^{-3} ) (0.03 for As)</td>
<td>(1.00 \times 10^{-3} ) (0.03 for As)</td>
</tr>
<tr>
<td>AF</td>
<td>(0.07)</td>
<td>(0.20)</td>
</tr>
<tr>
<td>AT(_{\text{carcinogenic}})</td>
<td>days</td>
<td>25,550</td>
</tr>
<tr>
<td>AT(_{\text{non-carcinogenic}})</td>
<td>days</td>
<td>9490</td>
</tr>
<tr>
<td>BW</td>
<td>(70.0)</td>
<td>(15.0)</td>
</tr>
<tr>
<td>CF</td>
<td>(1.00 \times 10^{-6})</td>
<td>(1.00 \times 10^{-6})</td>
</tr>
<tr>
<td>ED</td>
<td>(26.0)</td>
<td>(6.00)</td>
</tr>
<tr>
<td>EF</td>
<td>(350)</td>
<td>(350)</td>
</tr>
<tr>
<td>InhR</td>
<td>(20.0)</td>
<td>(7.60)</td>
</tr>
<tr>
<td>IngR</td>
<td>(100)</td>
<td>(200)</td>
</tr>
<tr>
<td>PEF</td>
<td>(1.36 \times 10^9)</td>
<td>(1.36 \times 10^9)</td>
</tr>
<tr>
<td>SA</td>
<td>(6032)</td>
<td>(2373)</td>
</tr>
</tbody>
</table>

The hazard quotient (HQ) and hazard index (HI) are commonly used metrics in human health risk assessments for evaluating the potential chronic effects of exposure to multiple heavy metals. The HQ is calculated by dividing the estimated exposure level of a particular heavy metal over a specified time period by a reference dose (RfD) for a similar exposure period. The RfD is an estimate of the amount of a substance that a person can be exposed to on a daily basis over a lifetime without appreciable health risks [43]. HQ and HI are calculated using Equations (6) and (7), respectively:

\[
\text{HQ} = \frac{\text{CDI}}{\text{RfD}} \quad (6)
\]

\[
\text{HI} = \sum \text{HQ} = \text{HQ}_{\text{ingest}} + \text{HQ}_{\text{inhal}} + \text{HQ}_{\text{dermal}} \quad (7)
\]

where RfD is the reference dose of heavy metals (mg \( \cdot \) kg\(^{-1} \)) and is different for each heavy metal, as shown in Table 3 [43]. HI > 1 indicates that the environment has a serious potential for non-carcinogenic risk; HI < 1 suggests that there is no significant potential for non-carcinogenic risks in the environment [44].
The Carcinogenic Risk (CR) assessment methodology is utilized to estimate the risk of carcinogenesis for humans who are exposed to the external environment via three distinct pathways [45]. The Lifetime Carcinogenic Risk (LCR) is calculated as the aggregate of the three CRs [45]. Equations (8) and (9) are used to compute both CR and LCR:

\[
CR = CDI \times CSF
\]

\[
LCR = CDI_{\text{ingest}} \times CSF + CDI_{\text{inhal}} \times CSF + CDI_{\text{dermal}} \times CSF
\]

where CSF refers to the carcinogenic slope factor of heavy metals and is different for each heavy metal (Table 4). According to the USEPA, for heavy metals, the minimum CR value is \(1.00 \times 10^{-6}\), and CR values between \(1.00 \times 10^{-6}\) and \(1.00 \times 10^{-4}\) suggest no severe carcinogenic risk to human health [6].

Table 4. The RfD values and CSF values of the seven heavy metals in the study area (data referred to [10,20,29,41,42,46]).

<table>
<thead>
<tr>
<th></th>
<th>Cd</th>
<th>As</th>
<th>Pb</th>
<th>Cr</th>
<th>Cu</th>
<th>Ni</th>
<th>Zn</th>
</tr>
</thead>
<tbody>
<tr>
<td>RfD\text{mg}</td>
<td>(1.00 \times 10^{-3})</td>
<td>(3.00 \times 10^{-4})</td>
<td>(3.50 \times 10^{-3})</td>
<td>(3.00 \times 10^{-3})</td>
<td>(4.00 \times 10^{-2})</td>
<td>(2.00 \times 10^{-2})</td>
<td>(3.00 \times 10^{-1})</td>
</tr>
<tr>
<td>RfD\text{derm}</td>
<td>(1.00 \times 10^{-5})</td>
<td>(1.23 \times 10^{-4})</td>
<td>(5.25 \times 10^{-4})</td>
<td>(6.00 \times 10^{-5})</td>
<td>(1.20 \times 10^{-2})</td>
<td>(5.40 \times 10^{-3})</td>
<td>(6.00 \times 10^{-2})</td>
</tr>
<tr>
<td>RfD\text{inhale}</td>
<td>(1.00 \times 10^{-5})</td>
<td>(3.00 \times 10^{-4})</td>
<td>(3.52 \times 10^{-3})</td>
<td>(2.86 \times 10^{-5})</td>
<td>(4.00 \times 10^{-2})</td>
<td>(9.00 \times 10^{-5})</td>
<td>(3.00 \times 10^{-1})</td>
</tr>
<tr>
<td>CSF\text{ingest}</td>
<td>1.50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CSF\text{derm}</td>
<td></td>
<td>3.66</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CSF\text{inhale}</td>
<td>6.30</td>
<td></td>
<td>42.0</td>
<td></td>
<td></td>
<td></td>
<td>8.40 (\times 10^{-1})</td>
</tr>
</tbody>
</table>

2.3.5. Potential Ecological Risks

The potential ecological risk index (RI) was a widely used method for assessing the potential ecological risk of soil heavy metal pollution. The RI considers both the concentration of heavy metals in the soil and the potential ecological impact of these metals [8,28,47]. This method takes into account various factors, such as the concentration of heavy metals, their toxicity, environmental quality standards, and ecological effects, in order to evaluate the potential impact of heavy metals on ecosystems [8]. The RI index can be calculated using the following equation:

\[
RI = \sum_{i=1}^{m} E_i^j = \sum_{i=1}^{m} T_i^j \times \left( \frac{C_i}{C_i^0} \right)
\]

where RI is the comprehensive potential ecological risk index; \(E_i^j\) is the individual potential ecological hazard index value of heavy metal \(i\); \(T_i^j\) is the toxicity factor value of heavy metal \(i\) [48]; \(C_i\) is the measured content of heavy metal \(i\); \(C_i^0\) is the reference value of heavy metal \(i\). The toxicity factor values of Pb, Cu, Cr, Cd, Zn, Ni, and As are 5, 5, 2, 30, 1, 5, and 10, respectively [47].

Based on the classification criteria of \(E_i^j\) and RI proposed by Hakanson (1980), \(E_i^j < 40\) means low potential ecological risk; \(40 \leq E_i^j < 80\) indicates moderate potential risk; \(80 \leq E_i^j < 160\) represents a considerable potential risk; \(160 \leq E_i^j < 320\) means high potential risk; \(E_i^j \geq 320\) indicates significantly very high risk. Four categories of RI values define, which are low risk (RI < 150), moderate risk (150 \(\leq\) RI < 300), considerable risk (300 \(\leq\) RI < 600), and high risk (RI \(\geq\) 600) [47].

3. Results

3.1. Characteristics of Heavy Metal Contamination

3.1.1. Statistics of the Soil Heavy Metals

The statistics of the soil heavy metals concentrations are listed in Table 5. Zn had the lowest mean value (147 mg·kg\(^{-1}\)), followed by Cr (89 mg·kg\(^{-1}\)), while Cd had the lowest mean value of 1.52 mg·kg\(^{-1}\). The obvious variations occurred in the measured
heavy metals concentrations, varying between 0.16 and 21.8 mg·kg\(^{-1}\) for Cd, 3.68 and 20.8 mg·kg\(^{-1}\) for As, 17.0 and 223 mg·kg\(^{-1}\) for Pb, 47.2 and 389 mg·kg\(^{-1}\) for Cr, 16.0 and 46.1 mg·kg\(^{-1}\) for Cu, 21.3 and 93.2 mg·kg\(^{-1}\) for Ni, 48.6, and 1.54 × 10\(^3\) mg·kg\(^{-1}\) for Zn. The following metals which exceeded level 1 (the risk screening values in the standard [37]) were 65 samples (47.1% of total) for Cd, 13 samples (9.40%) for Zn, 2 samples (1.40%) for Cr, 1 sample (0.70%) for Pb, and none for the others (As, Cu and Ni). For Cd, 13 samples even exceeded level 2 (the risk intervention values in the standard).

Table 5. Statistics of soil heavy metal concentrations (mg·kg\(^{-1}\)) in the study area.

<table>
<thead>
<tr>
<th></th>
<th>Cd</th>
<th>As</th>
<th>Pb</th>
<th>Cr</th>
<th>Cu</th>
<th>Ni</th>
<th>Zn</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max (mg·kg(^{-1}))</td>
<td>21.8</td>
<td>20.8</td>
<td>223</td>
<td>389</td>
<td>46.1</td>
<td>93.2</td>
<td>1.54 × 10(^3)</td>
</tr>
<tr>
<td>Min (mg·kg(^{-1}))</td>
<td>0.16</td>
<td>3.68</td>
<td>17.0</td>
<td>47.2</td>
<td>16.0</td>
<td>21.3</td>
<td>48.6</td>
</tr>
<tr>
<td>Mean (mg·kg(^{-1}))</td>
<td>1.52</td>
<td>11.7</td>
<td>35.6</td>
<td>89.3</td>
<td>26.0</td>
<td>36.2</td>
<td>147</td>
</tr>
<tr>
<td>CV (%)</td>
<td>180</td>
<td>24.0</td>
<td>74.0</td>
<td>45.0</td>
<td>15.0</td>
<td>20.0</td>
<td>121</td>
</tr>
<tr>
<td>Screening value (mg·kg(^{-1}))</td>
<td>0.600</td>
<td>25.0</td>
<td>170</td>
<td>250</td>
<td>100</td>
<td>190</td>
<td>300</td>
</tr>
<tr>
<td>Control value (mg·kg(^{-1}))</td>
<td>4.00</td>
<td>100</td>
<td>1000</td>
<td>1300</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Low risk (n)</td>
<td>73</td>
<td>138</td>
<td>137</td>
<td>136</td>
<td>138</td>
<td>138</td>
<td>125</td>
</tr>
<tr>
<td>Medium risk (n)</td>
<td>52</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>13</td>
</tr>
<tr>
<td>High risk (n)</td>
<td>13</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

3.1.2. Spatial Distribution of RF Model

The results of the RF model pass the significance test with a level of 0.01. The accuracy of seven heavy metals was greater than 0.5, and the \(R^2\) of Cd, Cr, Ni, and Zn was above 0.8 (Table 1). From the perspective of the precision of the validation set, the precision for the other six heavy metals except As was greater than 0.5 (the \(R^2\) of four heavy metals, Cd, Cr, Ni, and Zn, is above 0.6). Finally, the results were visualized in ArcGIS Pro 2.5. The weights of six factors (factory distance, slope, road distance, annual rainfall, annual average temperature, and elevation) were calculated (Figures 2 and 3).

Cd, Pb, and Zn had similar spatial distribution (Figure 2), with high pollution risk in the middle–east of the study area and low risk in the west and north–east regions. The high-risk regions were mainly concentrated in the concentrated areas of factories and enterprises. Similarly, Cr and Ni had similar spatial distributions, with high pollution risk in the north of the central region and low risk in the south–east and north–west regions. Cu had a different spatial distribution, with high pollution risk mainly concentrated in the middle of the study area and medium-risk areas in the south–west. However, the results for As were poor and did not accurately reflect the spatial distribution of pollution risks due to the relatively low \(R^2\).

The relative importance of six independent factors (factory distance, slope, road distance, annual rainfall, annual average temperature, and elevation) was analyzed (Figure 3). Specifically, the factory distance factor had the highest weight in the modeling process for all seven heavy metals. Precipitation and annual average temperature were also important factors in predicting the pollution risks of Cr and Ni. The slope factor was important for predicting the pollution risks of Cd, Pb, and Cu. For As, the weight of the factory distance factor is not significantly higher than the weights of the other factors, and the weights of the other five factors are relatively equal. Considering the poor RF modeling accuracy of As (Figure 2), it can be inferred that none of the six factors is the main determinant of As soil contamination.
Figure 2. Prediction results of heavy metal pollution risk based on random forest model, (a) Cd, (b) As, (c) Pb, (d) Cr, (e) Cu, (f) Ni, and (g) Zn.
The weights of different factors in the random forest model, (a) Cd, (b) As, (c) Pb, (d) Cr, (e) Cu, (f) Ni, and (g) Zn.

3.2. Contamination Evaluation of Heavy Metals

3.2.1. Single Pollution Index Method

The mean values of $P_i$ for each heavy metal, from highest to lowest, were as follows: Cd (2.53), Zn (0.49), As (0.47), Cr (0.36), Cu (0.26), Pb (0.21), and Ni (0.19). Cd was the highest pollutant among all the tested metals. Based on the classification criteria for different levels of pollution, the $P_i$ values indicated that the pollution level of Cd was in the moderately to heavily contaminated class. The pollution levels of the other heavy metals (Zn, As, Cr, Cu, Pb, and Ni) were at uncontaminated levels based on the mean values of $P_i$ (Figure 4). However, when considering the maximum value, $P_{max}$ of Cd, Pb, and Cr were at low to moderately contaminated levels, while Cd and Zn were at heavy to extremely contaminated levels. The other heavy metals (As, Cu, and Ni) showed an unpolluted tendency.

Overall, the study area was contaminated with Cd, and there might be some contamination with Zn, Pb, and Cr, depending on the maximum values of $P_{max}$. The other heavy metals were not likely to cause significant pollution in the area.

3.2.2. Nemero Index

The mean value of the Nemero index in the study area was 1.86, which indicated that the area was slightly polluted, but the precaution levels for the samples had varying levels of pollution, with some samples being only slightly polluted and others being seriously polluted. The results indicated that more than 56.5% of the samples were above the safety level (34 warning level samples, 14 mild pollution level samples, 7 moderate pollution level samples, and 23 severe pollution level samples).

The spatial interpolation and classification of the Nemero index values were carried out using inverse distance weighting (IDW) interpolation [29,36]. Figure 5 shows that the pollution levels in the study area ranged from the safety level to the seriously polluted level. The study identified regions with slightly polluted levels, moderately polluted levels, and seriously polluted levels, which covered areas of 34.3 km$^2$, 11.2 km$^2$, and 22.0 km$^2$, respectively. The polluted areas were mainly concentrated in the central-eastern region, which correlated well with relevant Cd concentrations, as shown in Figure 4.
level (34 warning level samples, 14 mild pollution level samples, 7 moderate pollution level samples, and 23 severe pollution level samples).

The spatial interpolation and classification of the Nemero index values were carried out using inverse distance weighting (IDW) interpolation [29,36]. Figure 5 shows that the pollution levels in the study area ranged from the safety level to the seriously polluted level. The study identified regions with slightly polluted levels, moderately polluted levels, and seriously polluted levels, which covered areas of 34.3 km², 11.2 km², and 22.0 km², respectively. The polluted areas were mainly concentrated in the central–eastern region, which correlated well with relevant Cd concentrations, as shown in Figure 4.

Figure 5. IDW interpolation of the Nemero index.

4. Discussion

4.1. Pollution Feature and Possible Sources

Based on the statistical results of the samples (Table 5), it could be concluded that the main pollutant of heavy metal pollution in cultivated soil in the study area was Cd, followed by Zn. The coefficient of variation (CV) of Cd and Zn exceeded 100%, which suggested that the variability of these two elements was high. This high variability indicated that external factors had a significant influence on the accumulation of these metals in the soil [46].

The RF model, which was built using six independent variables, including factory distance, slope, road distance, annual rainfall, average annual temperature, and elevation, provided accurate predictions of heavy metal concentrations in the soil for each grid with a size of 500 m in the study area. The results indicated that most heavy metals, especially Cd, Pb, and Zn, had similar spatial distribution (Figure 2), with high pollution risk in the middle–east of the study area and low risk in the west and north-east regions, which was consistent with the locations of the industrial enterprises. The spatial distribution of heavily polluted areas derived from RF was comparable to that mapped from the single pollution index method and Nemero index (Figures 2, 4, and 5).

Based on the relative importance of six independent factors in the RF model, the factory distance factor was identified as having the highest weight in predicting pollution risks for all seven heavy metals, with particular significance for Cd and Zn (Figure 3). The modeling process showed that precipitation and annual average temperature were important factors in predicting the pollution risks of Cr and Ni. On the other hand, the slope factor was found to be significant in predicting the pollution risks of Cd, Pb, and Cu.

Figure 4. The spatial distribution of 7 heavy metal concentrations, (a) Cd, (b) As, (c) Pb, (d) Cr, (e) Cu, (f) Ni, and (g) Zn.

Figure 5. IDW interpolation of the Nemero index.
4. Discussion

4.1. Pollution Feature and Possible Sources

Based on the statistical results of the samples (Table 5), it could be concluded that the main pollutant of heavy metal pollution in cultivated soil in the study area was Cd, followed by Zn. The coefficient of variation (CV) of Cd and Zn exceeded 100%, which suggested that the variability of these two elements was high. This high variability indicated that external factors had a significant influence on the accumulation of these metals in the soil [46].

The RF model, which was built using six independent variables, including factory distance, slope, road distance, annual rainfall, average annual temperature, and elevation, provided accurate predictions of heavy metal concentrations in the soil for each grid with a size of 500 m in the study area. The results indicated that most heavy metals, especially Cd, Pb, and Zn, had similar spatial distribution (Figure 2), with high pollution risk in the middle–east of the study area and low risk in the west and north–east regions, which was consistent with the locations of the industrial enterprises. The spatial distribution of heavily polluted areas derived from RF was comparable to that mapped from the single pollution index method and Nemero index (Figures 2, 4 and 5).

Based on the relative importance of six independent factors in the RF model, the factory distance factor was identified as having the highest weight in predicting pollution risks for all seven heavy metals, with particular significance for Cd and Zn (Figure 3). The modeling process showed that precipitation and annual average temperature were important factors in predicting the pollution risks of Cr and Ni. On the other hand, the slope factor was found to be significant in predicting the pollution risks of Cd, Pb, and Cu.

The results obtained in this study were consistent with those derived from source analysis of heavy metals using principal component analysis, cluster analysis, and positive matrix factor analysis methods (PMF) [49,50]. Through these methods, five potential sources, including soil parent material, coal burning, agricultural and industrial activity, electroplating, and transportation, were identified. Previous research had shown that the two major pollutants, Cd and Zn, mostly originated from agricultural and industrial activities and were distributed around industrial enterprises [51], which was in line with the finding that Cd and Zn pollution had the highest weight with the factory distance factor. Relevant studies had also indicated that there was a common source between Cr and Ni in soil, and these two heavy metals both were iron-philic elements in the epigenetic geochemical process. The content of these metals in soil was likely dominated by the soil-forming process [34,35], which was similar to the RF results indicating that the two metals resolved from the soil parent material correlated with precipitation and annual average temperature. This confirms that the combination of RF, single pollution index, and Nemero index could effectively determine the pollution degree of heavy metals, providing valuable information for the formulation of soil remediation measures.

4.2. Human Health Risks

The potential risks to children and adults posed by contaminated soil were assessed through ingestion, inhalation, and dermal contact pathways (Table 6). The HI for children was 1.21, which was nearly 10 times higher than that for adults (HI = 1.42 \times 10^{-1}). The sum of the HQ values for each pathway of the seven heavy metals showed that the risk of ingestion was the highest for both adults and children, followed by dermal contact and inhalation. However, the risk of inhalation was considered negligible due to its low HQ value compared to the other two pathways.
In addition to non-carcinogenic risks, the carcinogenic risk (CR) of exposure to heavy metals, including Cd, As, Cr, and Ni, was calculated (Table 6), and it was found that the potential CR risks for children were greater than those for adults. However, none of the two indices (CR and HI) were greater than $1.00 \times 10^{-5}$, which indicated that there were no severe potential carcinogenic risks in the area.

The results suggest that while there are potential non-carcinogenic risks for children due to exposure to heavy metals, there are no severe potential health risks posed to adults. It is important to continue monitoring and regulating exposure to heavy metals to ensure that these risks remain low and do not pose a significant threat to human health.

### 4.3. Potential Ecological Risks

According to the results, Cd was found to be the most dangerous heavy metal in the study area, with the highest mean risk value (75.8) and maximum risk value (1090) among all the heavy metals (Cd > As > Pb > Cu > Ni > Cr > Zn). The IDW interpolation and classification criteria were used to map the potential ecological risk index values, and the spatial distribution of the ecological risk index showed that the area with low risk was the largest, followed by the area with moderate risk (Figure 6). The areas with considerable and very high-risk areas were smaller in size but still significant.

![Figure 6. Map of IDW interpolated potential RI values for the study area.](image)

It appears that the study area has an uneven distribution of RI, with the lowest, mean, and highest values being 16.9, 85.0, and $1.11 \times 10^3$, respectively. Cd is identified as the most significant contributor to the RI, accounting for 89.2%. The spatial distribution of RI is similar to the distribution of Cd, Pb, and Zn concentrations. The study area is divided into the following four categories based on RI values: low risk (RI < 150); moderate risk...
(150 ≤ RI < 300); considerable risk (300 ≤ RI < 600); and very high risk (RI > 600). The area and ratio of each category were 93.3 km$^2$ and 43.4%, 78.2 km$^2$ and 36.3%, 22.4 km$^2$ and 10.4%, and 21.3 km$^2$ and 9.9%.

The results concluded that heavy metals posed noticeable potential ecological risks in the study area, and Cd was the most polluted heavy metal and, therefore, should be evaluated specifically for the relevant risks.

5. Conclusions

This study measured the levels of seven heavy metals (Cd, As, Pb, Cr, Cu, Ni, and Zn) in soils in the Ganhe Industrial Park in the Upper Yellow River and found that Cd and Zn were the main pollutants in the soils. A total of 65 samples (47.1%) for Cd, 13 samples (9.40%) for Zn, 2 samples (1.40%) for Cr, and 1 sample (0.70%) for Pb were proved to exceed the soil environmental quality standard values. Cd was more concentrated than the other elements, and the polluted areas were mainly concentrated in the central–eastern region, which was spatially correlated with the factories, such as zinc smelting plants and chemical plants. The pollution index and Nemero index results showed that more than 56.5% of the samples were beyond the safety level, indicating that the soil was slightly polluted.

The study also found that there were serious potential non-carcinogenic risks for children (HI = 1.21) but no severe potential health risks posed to adults (HI = 0.14). Similarly, the potential carcinogenic risk (CR) of heavy metals for children (CR = 2.27 × 10$^{-5}$) was greater than those for adults (CR = 1.20 × 10$^{-5}$). The individual index values of potential ecological risk assessment of heavy metals indicated that Cd was the main contributor to ecological risk as it recorded the highest E$^i_r$ values.

Overall, this study highlights the potential ecological risks posed by heavy metals in the study area and the need for specific evaluation of Cd for relevant risks. The results provide theoretical support for pollution control and environmental management in the study area and the Upper Yellow River in China.

Author Contributions: F.Y.: conceptualization, writing, project administration, review, writing—original draft; W.M.: methodology, review; L.L.: project administration, review, editing; K.F.: writing and editing, software; C.Y.: review and supervision. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (42071258).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References
2. Imin, B.; Abliz, A.; Shi, Q.; Liu, S.; Hao, L. Quantitatively Assessing the Risks and Possible Sources of Toxic Metals in Soil from an Arid, Coal-Dependent Industrial Region in NW China. J. Geochem. Explor. 2020, 212, 106505. [CrossRef]


33. Fan, Y.; Fang, C. A Comprehensive Insight into Water Pollution and Driving Forces in Western China—Case Study of Qinghai. J. Clean. Prod. 2020, 274, 123950. [CrossRef]


Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.