Article

Comparison of Models for Spatial Distribution and Prediction of Cadmium in Subtropical Forest Soils, Guangdong, China

Xiaogang Ding ¹, Zhengyong Zhao ^{2,3,*}, Zisheng Xing ⁴, Shengting Li ⁵, Xiaochuan Li ¹ and Yanmei Liu ⁶

¹ Guangdong Academy of Forestry, Guangzhou 510520, China; 27267152@sinogaf.cn (X.D.); lixiaochuan@sinogaf.cn (X.L.)

² Guangxi Key Laboratory of Forest Ecology and Conservation, College of Forestry, Guangxi University, Nanning 530004, China

³ Faculty of Forestry, University of British Columbia, Vancouver, BC V6T 1Z4, Canada

⁴ AAFC-Portage, BRDC, Brandon, MB R1N 3V6, Canada; zisheng.xing@agr.gc.ca

⁵ State Key Laboratory for Conservation and Utilization of Subtropical Agro-Bioresources, College of Life Sciences, South China Agricultural University, Guangzhou 510642, China; lishengting@stu.scau.edu.cn

⁶ School of Biological Engineering and Technology, Tianshui Normal University, Tianshui 741001, China; lym-78118@163.com

* Correspondence: z.zhao@unb.ca

Citation: Ding, X.; Zhao, Z.; Xing, Z.; Li, S.; Li, X.; Liu, Y. Comparison of Models for Spatial Distribution and Prediction of Cadmium in Subtropical Forest Soils, Guangdong, China. *Land* **2021**, *10*, 906. <https://doi.org/10.3390/land10090906>

Academic Editors: Assem Abu Hatab, Nasem Baderldin and Zhen Liu

Received: 22 June 2021

Accepted: 17 August 2021

Published: 27 August 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Cadmium (Cd) is a toxic metal and found in various soils, including forest soils. The great spatial heterogeneity in soil Cd makes it difficult to determine its distribution. Both traditional soil surveys and spatial modeling have been used to study the natural distribution of Cd. However, traditional methods are highly labor-intensive and expensive, while modeling is often encumbered by the need to select the proper predictors. In this study, based on intensive soil sampling (385 soil pits plus 64 verification soil pits) in subtropical forests in Yunfu, Guangdong, China, we examined the impacting factors and the possibility of combining existing soil information with digital elevation model (DEM)-derived variables to predict the Cd concentration at different soil depths along the landscape. A well-developed artificial neural network model (ANN), multi-variate analysis, and principal component analysis were used and compared using the same dataset. The results show that soil Cd concentration varied with soil depth and was affected by the top 0–20 cm soil properties, such as soil sand or clay content, and some DEM-related variables (e.g., slope and vertical slope position, varying with depth). The vertical variability in Cd content was found to be correlated with metal contents (e.g., Cu, Zn, Pb, Ni) and Cd contents in the layer immediately above. The selection of candidate predictors differed among different prediction models. The ANN models showed acceptable accuracy (around 30% of predictions have a relative error of less than 10%) and could be used to assess the large-scale Cd impact on environmental quality in the context of intensifying industrialization and climate change, particularly for ecosystem management in this region or other regions with similar conditions.

Keywords: artificial neural network (ANN); heavy metal; multi-variate analysis; partial correlation analysis; soil survey

1. Introduction

Toxic metal contamination in soil has become a global issue due to its environmental impacts and potential risks and hazards to human beings. Among toxic metals in soils, Cadmium (Cd) has been of great concern and is extensively studied around the world [1–3]. Cd affects human health by leaching into streams and groundwater but is also of concern in agricultural soils [4]. Cd has been listed as one of the most serious soil contaminants in China [5–7]. Cd may participate in many processes in soil-plant biota systems (e.g., triggering the synthesis of reactive oxygen species), and it can hinder the utilization, uptake, and transport of essential nutrients and water and modify the photosynthetic

machinery, thereby resulting in plant tissue death [8]. On the other hand, its build-up at high concentrations in soil may also pose risks and hazards to human and ecosystem health through the food chain, contaminated groundwater, and microorganism processes [9], thus affecting the density and diversity of meso- and macro-fauna [10]. For example, Cd may influence the health of women with low nutritional stores and affect fetal development [11], and it may potentially pose both non-carcinogenic and carcinogenic risks, such as kidney disease, skeletal damage, and cancers [7,12] in humans. Marked cadmium contamination was reported in areas where food had been grown in Japan in the 1950s and 1960s, with Cd concentrations ranging from 0.2 to 3 mg kg⁻¹ [13].

The average natural abundance of Cd in the Earth's crust is normally reported to range from 0.1 to 0.5 mg kg⁻¹ across the world [14–16], with some exceptions: some reports indicate a significant percentage increase in various soils, with up to 25 mg kg⁻¹ detected in soil derived from sedimentary rocks. These variations may be attributed to many factors, such as climatic, topographic, ecological, soil, and biological conditions. For example, a recent review study in China indicated that Cd ranged from 0.003 mg kg⁻¹ to 9.57 mg kg⁻¹ in Chinese soils and varied among different locations [7]. The spatial heterogeneity in soil Cd has been associated with varying land uses, from high concentrations in agricultural soil to low concentrations in undisturbed forest soil [3,17–19]. Cadmium in soils can be derived from the underlying bedrock, transported parent material (e.g., glacial till and alluvium), or aerial deposition and sewage sludge, manure, and phosphate fertilizer application [7]. Many factors, such as geological formation processes, climate conditions, soil formation factors, and interactions among soil-plant systems, can affect Cd distribution along the landscape or soil profile [4,13,20,21]. A study on the distribution of toxic metals down to the BC or C horizon of soils at 95 locations among five forest production regions of Switzerland suggested that most toxic metal contamination originated from the parent rock [22], while anthropogenic input was detected in most topsoils. Therefore, Cd accumulation in the components of terrestrial ecosystems is conditioned by the deposition intensity, soil characteristics, mineral composition, and vegetation type [23]. In general, the transfer of chemical elements/compounds within the soil-plant chain is a part of biochemical cycling, which determines the final mobility and availability of chemicals, particularly within topsoils. Cadmium may be transferred from the soil to plant parts through the biogeochemical cycle [24] or from plant residues to soils. Cd may vary with different soil pH because of its impact on the uptake of Cd by plants and changes in solubility and adsorption [17], and Cd can also vary by depth, as indicated by a previous study [18]. A study on some acidic forest soils in Hunan Province, China, concluded that the release and potential active speciation of Cd in the tested contaminated red soil (CRS) and yellow-red soil (CYRS) increased significantly with decreasing pH and increasing ion concentrations of simulated acid rain [17] within the soil profile.

As a rule, the spatial distribution of Cd is determined through soil surveys. However, this approach is more expensive and labor-intensive and should be limited to finding the ground truth only. However, the complexity of soil characteristics leads to great uncertainty and difficulties in the extrapolation of field survey results [25]. Therefore, seeking an alternative approach to determining Cd distribution has become an important topic. Among the alternatives to soil surveys, statistical analysis, modeling, or modeling + GIS techniques, and remote sensing technology have found applications. For example, Alam et al. [26] used statistical and GIS techniques to assess and predict concentrations of Cd in the soil of Lahore City, Pakistan. Asmaryan et al. [27] applied a remote sensing method to assess and map soil pollution with Cd. Frequently, multi-variate regression is the first choice for estimating Cd with the use of various easily obtained predictors [18], although this method may not perform well in many cases due to difficulties in selecting proper predictors and other complexities, such as the interactions among various impacting factors. As a statistical method, principal component analysis (PCA) has had some success in distinguishing the impacting factors of Cd, although this method still suffers from difficulties in the selection of the variables used for predictions [28,29]. Using anomaly analysis, correlation

analysis, and gray relational analysis to explore the relationship and the occurrence of Cd in the soil environment in the Tianshan Mountain regions of China, Zhang et al. [30] successfully illustrated the distribution patterns of Cd along land use gradients in this region with statistical approaches. Cao et al. [3] used regression kriging to map the spatial distribution of Cd in an area of China where Cd contamination was identified.

Among modeling approaches, ANN has been widely used to estimate some soil properties including Cd [31–33]. In a recent conference, Ghazi [31] presented an approach in which a neural network is used to estimate the concentrations of Cd with variables such as soil retention, release reactions of solutes, and the soil matrix. However, his method only demonstrated the Cd variability with soil depth and lacked the ability to link the results to landscape variability. Some studies, although few, have been conducted to address the most complicated but important task of estimating spatial and temporal concentrations of Cd. For example, an ANN-GA model (artificial neural network-genetic algorithm), which has been proven to have better performance than ANN alone, was developed by Nadri et al. to estimate Cd removal and determine heavy metal contamination [32]. However, due to the black-box property of this method, many networks developed from this algorithm lack ecological, biological, geological, and topological principles. Even though this method can generate formulae based on sampled information, the extrapolation of the final results may be limited to the sampled data range, with limited potential for application beyond the region for which the model is developed. Therefore, the development of methods to translate or extrapolate valuable existing soil data, such as the existing soil map and collected soil sampling data, is a valuable study area for gaining very useful knowledge required by researchers.

Although some approaches have been developed and tested in various countries, quantifying the Cd concentrations in soils across a landscape and soil profile remains a challenge for the following reasons: (1) the complexity of Cd distributions (i.e., affected by large-scale parent materials resulting from geological formation and weathering processes, contamination from local industrial operations and atmospheric pollution, land use changes, and climate change) for using a process-based model and (2) the lack of an approach to determine the spatial distribution of Cd using existing soil data to scale up due to either insufficient soil sampling data or improper spatial analysis methodology. In this paper, we present a study to (1) analyze the Cd distribution pattern based on an intensive soil survey and (2) explore and compare several potential approaches for estimating the vertical and spatial distributions of Cd concentrations with a dataset obtained in a subtropical forest ecosystem in South China. The data include information from a coarse soil map, DEM-derived variables, and several soil layer properties (e.g., soil organic contents, chemical concentrations, soil particle size). The compared methods include ANN, multi-variate analysis, principal component analysis, and Pearson correlation. The results can provide valuable insight and information for the assessment of Cd contamination potential at the landscape level in this region or other similar regions while being used as a measurement tool for assessing the large-scale soil environmental quality and forest management in this region.

2. Materials and Methods

2.1. Study Site

The study site in the Yunfu forest region (22°22′–23°19′ N, 111°03′–112°31′ W) is located in west Guangdong, China, with an area of 7785 km², and it is characterized by typical mountainous (69%) and hilly (31%) topographic conditions (Figure 1). The combination of high air temperature (mean annual 22.4 °C) and abundant precipitation (1900 mm) makes the conditions in this region highly suitable for plant growth (Guangdong Soil Census Office, Guangzhou, China, 1993) and supports typical tropical monsoon forests, such as natural secondary evergreen broad-leaved forest, coniferous forest, and mixed forest, with major species including *Cunninghamia lanceolata*, *Acacia* spp., *Eucalyptu* spp., and *Pinus massoniana* (Guangdong Soil Census Office, Guangzhou, China, 1993).

Figure 1. Research site, DEM and soil sampling location.

Udults (a suborder of Ultisols, covering 86% of the region) are the major soil type, with some coverage of Entisols and Inceptisols (Guangdong Soil Census Office, Guangzhou, China, 1993). These soils are the result of many years of interactions among the hot and humid climate, geological evolution, and forests. Udults are the dominant soils beyond the humid tropics and subtropics of southern China [34–36]. Strong desilication and enrichment in iron and aluminum, an argillic or kandic horizon (B horizon), and distinct vertical patterns of soil texture are other typical features of the soils [37].

2.2. Soil Sampling, Processing, and Application

Soil samples for Cd concentration analysis were collected in Yunfu from 449 forest plots established by Guangdong Academy of Forestry Sciences through a project assessing forest soil nutrients, which started in 2015 [38]. In total, 385 plots within two of five sub-areas in Yunfu were used to develop the models, while 64 additional verification points were sampled for model verification (Figure 1). To improve the representativeness of each soil sampling pit, sampling stratification was implemented on a coarse-resolution soil texture map to allocate each plot geospatially. Georeferenced information for each sample point was extracted from the Atlas of Guangdong Soils (1:2,800,000, Guangdong Soil Survey Team, Guangzhou, China, 1993) using remote sensing and GIS tools. In addition, the soil properties from the coarse soil map database were spatially extrapolated to generate information on soil texture, nutrient contents, and soil organic matter content for each soil sampling site, as shown in Supplementary Table S1.

At each sampling location, a soil pit (about 50 × 50 × 100 cm) was dug to a 1 m depth for soil with a depth greater than 1 m or a maximum depth to the parent material horizon for plots with shallow topsoil <1 m. The vertical profile of each pit was divided

into five layers from the soil surface down to 1 m depth. The layers were D1: 0–20 cm; D2: 20–40 cm; D3: 40–60 cm; D4: 60–80 cm; and D5: 80–100 cm. A soil sample with a weight of 1 kg was collected from each layer following the standard procedure (ISSS 1929) for the determination of soil chemical and physical properties and texture.

Each soil sample was air-dried and ground, passed through a sieve of 2 mm, and stored in glass bottles for further analysis. After further processing the soil samples from each layer, a composite soil sample per soil pit was generated by subsampling from each layer sample while retaining independent layer samples as well. The concentrations of Cd and other heavy metals in soil samples were determined based on the methods in GB/T 17140-1997 and measured with an Atomic Absorption Spectrometer (AAS, Wincom, Shanghai, China). The soil pH, bulk density, particle size, and texture were determined following standards described in Determination of Soil Physical Parameters (Soil Physics Research and Education Group, Nankai Soil Research Institute, Chinese Academy).

2.3. DEM-Derived Topo-Hydrologic Variables

The DEM-derived variables listed in Table 1 were obtained through the following steps:

Table 1. DEM-derived topo-hydrologic variables *.

Variable	Description	Range/Number of Classes	Reference
Topographic position index (TPI)	The relative topographic position	6 classes	[39]
Slope	Slope gradient (degrees)	0.0–56.0	
Aspect	Slope aspect (0–360°) reclassified into 9 general directions flat, north, northeast, east, southeast, south, southwest, west, north west.	9 classes	(ESRI 1999–2013)
Potential solar radiation (PSR)	The PSR is a calculated total of potential solar radiation (kWh) over a year by taking more than three atmospheric factors affecting heterogeneity into consideration such as topographic changes, position changes and cloudiness	484–1888	[40]
Soil terrain factor (STF)	A modified version of hydrological similarity index	3.7–34.1	[41,42]
Sediment delivery ratio (SDR)	The ratio of the sediment transported to the outlet out of total erosion in a watershed	0–100	[43]
Vertical slope position (VSP)	Elevation difference between the upland and nearest water surfaces (m)	0.0–268.1	[44]
Flow direction (FD)	The steepest descent direction of each pixel, which is one of the keys to obtaining surface hydrological features	8 classes	[45,46]
Flow length (Length)	The length of the maximum ground distance along the FD projected to the nearest water channel		[47]

Note: * the table is adapted from Zhao et al. [48] with some modifications.

- Extracting DEM data that were derived from stereo images of Cartosat-1 (IRS-P5) with 12.5 m-resolution [48];
- Subsampling the obtained DEM data and recasting to 10 m resolution DEM;
- Deriving nine topo-hydrologic variables as defined in Table 1. Some detailed information has been reported in previous studies [48]. The spatial analyst extension tools and forest hydrology tools of ArcGIS were used; and
- Linking DEM dataset with soil sampling information.

2.4. Statistical Methods and Models

2.4.1. Principal Component Analysis, Multi-Variate Regression Analysis, and Correlation Analysis

To examine the relationship between Cd concentrations and DEM-derived variables and other soil properties collected in this study, principal component analysis (PCA) was

conducted using data obtained from each soil depth, including the mixed layer (0–100 cm). However, using the same values of DEM-derived variables for each layer limited the application of the method because the results from each layer tended to reveal no difference. Therefore, we only report the results from the mixed soil layer (0–100 cm). The analysis served as a reference for the selection of candidate variables in ANN model development.

Multi-variate regression and correlation analyses were also conducted to analyze the Cd concentration variations with DEM-derived variables and other soil properties, such as the contents of other heavy metals, pH, and sand and clay contents in the soil. The results were used to examine the contribution of each variable to inform the decision about whether to exclude or include a predictor. Backward stepwise regression and correlation analysis were conducted with Statistix 10 software (Analytical Software, Tallahassee, FL, USA).

2.4.2. Artificial Neural Network Model

Development Principle of Artificial Neural Network Model

Similar to many studies on ANN model development, the current study used a back-propagation ANN modeling approach [48,49] by taking advantage of its nonlinear mapping capabilities, which are suitable for limited discontinuous points between the input and output variables without a hypothesis [50]. Each developed model was tested, and better results were obtained with a model with three layers, including an input layer, a hidden layer, and an output layer. The input layer consists of independent variables (i.e., one or more of the DEM-derived variables in Table 1, depending on the best model performance). The output layer is the dependent variable (i.e., Cd concentration) in this study. The input and output layers are connected through a hidden layer. Input variables in the input layer are connected to all nodes of the next layer. The input weight matrix is composed of all links between the input and hidden layers, and the output weight matrix is composed of all links between the hidden and output layers. The weight (w) matrix controls the propagation value (x) and the output value (o) from each node and is modified based on the obtained value from the preceding layer according to Equation (1).

$$o_i = f(T + \sum w_i \cdot x_i) \quad (1)$$

where T is a specific threshold (bias) value for each node and f is a nonlinear sigmoid function, which increases monotonically.

During the back-propagation training process, the ANN model structure was adjusted by changing the weight and bias values along a negative gradient descent to minimize the mean squared error (MSE) between the network outputs (predicted values) and the targeted values (measured values) [51] using the Levenberg-Marquardt algorithm.

In addition, the collected dataset of 385 soil pits was divided into training and testing sub-datasets. The training data (10% of total data) were used to calculate the gradient, update the network weights, estimate the biases, and calculate MSE. The testing dataset was used to prevent the training process from “over-fitting”. Once the training MSE decreased and the testing MSE increased, the training of the ANN model was stopped under the assumption that the best-fitting model coefficients had been obtained.

Selection of Model Inputs

In the selection of input variables (as predictors) for the ANN models, we assume that the Cd concentration can be related to on-site conditions (DEM-related variables) [35]. However, at different soil depths, the leading factors affecting the Cd amount may be different when using 1 or more of the 9 variables in Table 1. Due to a lack of knowledge about the role of each input variable in determining Cd distribution, we adopted a comprehensive approach to select the best combination of variables for a given number of input variables. For example, for a model constructed with only one input variable, we performed 9 runs to select the best variable. If 2 input variables were used, then there were C_9^2 combinations or 36 runs to accommodate all combinations of 2 variables out of 9. After 36 runs of all combinations of all pairs of variables, the best 2-variable model from these

runs was determined based on the lowest MSE, highest r^2 , and lowest relative error. For each soil depth, there was a maximum of 9 best models, each with variables from 1 to 9, as shown in Table 2. The best model for each soil depth is highlighted in Table 2.

Table 2. Ten-fold ANN model comparison.

Depth	Parameter Number	RMSE	r^2	ROA10% (Ratio)	ROA20% (Ratio)	Variables Included in Modelling
0–20 cm	1	0.379	0.655	0.166	0.288	Slope
	2	0.41	0.769	0.213	0.327	Slope, Direction
	7	0.623	0.882	0.332	0.488	TPI, Slope, Aspect, VSP, Length, FD, PSR
20–40 cm	1	391.543	0.424	0.106	0.247	SDR
	2	313.05	0.519	0.145	0.281	SDR, Direction
	7	108.356	0.869	0.319	0.475	TPI, Slope, STF, VSP, Length, FD, PSR
40–60 cm	1	325.248	0.474	0.119	0.26	SDR
	2	235.787	0.662	0.143	0.27	Slope, Direction
	7	173.355	0.777	0.275	0.423	Slope, Aspect, STF, VSP, Length, FD, PSR
60–80 cm	1	249.519	0.419	0.131	0.241	Slope
	2	158.023	0.691	0.131	0.276	Slope, Aspect
	7	114.598	0.796	0.212	0.365	TPI, Slope, Aspect, STF, VSP, Length, PSR
80–100 cm	1	249.519	0.419	0.131	0.241	Slope
	2	196.055	0.623	0.158	0.276	Slope, VSP
	7	114.598	0.796	0.212	0.365	TPI, Slope, Aspect, STF, VSP, Length, PSR
0–100 cm	1	799.402	0.497	0.13	0.236	SDR
	2	663.318	0.613	0.14	0.278	SDR, Length
	7	360.45	0.831	0.332	0.488	TPI, Slope, Aspect, SDR, VSP, Length, PSR

Note: The last column are the final inclusion of variables for the best model through ANN modeling. The denotation of the variables can be found in Table 2 and Table S1. RMSE = Root mean squared error, r^2 = Coefficient of determination, ROA10%(ratio) = Relative overall accuracy, a prediction ratio out of predictions with relative overall accuracy threshold 10%, ROA20%(ratio) = a prediction ratio out of predictions with relative overall accuracy threshold 20%.

In addition, soil type information extracted from the CSM was included in all constructed models, regardless of the number of variables used to fit ANN models.

Model Fitting and Validation

Our dataset was composed of a total of 385 depth-specific soil Cd concentration measurements from two of the five sub-areas in Yunfu, the study area. To obtain the best model for each soil sampling depth, a k-fold cross-validation or rotation estimation approach (i.e., $k = 10$ in this study) was used with the following steps: (a) split the entire dataset into 10 approximately equal subsets; (b) take one subset as a hold-out or test dataset; (c) use the remaining 9 datasets as the training dataset; (d) fit an ANN model with the first training subset and evaluate the model with the testing dataset; (e) retain the evaluation score and discard model; and (f) rotate to the second subset of the 10 sub-datasets and use it as the training subset, repeating steps c–e. By following this procedure, the performance of about 10 models can be assessed, and the average of each assessment criterion can be regarded as the reported model performance.

A small dataset consisting of 64 soil pits was also used to further validate the models. The small dataset was independent of the model development dataset and was used to calibrate the models.

Accuracy Assessment of ANN Model

The screening of ANN models was based on three indicators of model accuracy commonly used in model development for k-fold cross-validation.

- Root-mean-square error (RMSE) or root-mean-square deviation (RMSD), as described by Hyndman and Koehler [52], as follows:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \quad (2)$$

where $\hat{y}_i$ is the predicted value, y_i is the measurement, and n is the total number of data points.

- Coefficient of determination (r^2): The proportion of variation explained by each ANN model:

$$r^2 = \frac{\sum_{i=1}^n (y_i - \hat{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (3)$$

- Relative overall accuracy (ROA): A relative accuracy indicator of model predictions. Model predictions were considered to be relatively accurate if they were within a certain variation range of measurements [33]. In other words, if we use a threshold of relative error as a criterion, out of all prediction points, the total number of model predictions with relative error within $\pm\%$ will be a relatively good indicator of model prediction accuracy. For example, ROA $\pm 5\%$ was calculated by counting all predictions within a $\pm 5\%$ range of the measurement over the total measurement points as follows:

$$\text{ROA}\% = \frac{\sum_{i=1}^n \begin{cases} 1 & \text{if } \text{abs}\left(\frac{\hat{y}_i - y_i}{y_i}\right) \times 100 < T \\ 0 & \text{else} \end{cases}}{n} \quad (4)$$

where T is the accuracy threshold (e.g., 5 for 5%), determined based on a target to fit the model. In this study, we set T to both 10% and 20% to calculate the model accuracy parameters, as shown in Table 2. This study revealed that ROA% at two thresholds (i.e., 10 and 20%) had similar patterns.

Therefore, we can conclude that successful models have high values of ROA and r^2 and low values of RMSE.

Best ANN Model Determination

To assist in the objective selection of the generated ANN models, a determinant index of model selection was adopted. The index is defined as follows:

$$Si = \frac{\text{ROA}10\%}{\text{RMSE}} r^2 \quad (5)$$

Si was calculated for each model (containing a certain number of predictable variables) based on average ROA 10%, average RMSE, and r^2 for each soil layer.

The model with the highest Si was the selected model for each layer. The selected model was used to estimate Cd concentrations for the layer in the study area and produce a spatial map of Cd concentrations. In addition, the mean estimated Cd concentration from all 10-fold runs was used to generate a distribution map of Cd to demonstrate the spatial pattern of the soil Cd for the study area.

2.5. Other Statistical Analyses

Analysis of variance (ANOVA) was used to examine the significance of Cd concentration variations along the soil profile. Pearson correlation analysis was also used to examine the Cd concentration with other impacting factors. These analyses were performed with Statistix 10 software (Analytical Software, Tallahassee, FL, USA).

3. Results

3.1. Cd Variation in Soil Samples

As shown in Table 3, the average Cd concentration across different soil depths was 0.018 mg kg^{-1} . The differences among layers were statistically significant ($p < 0.05$), and the mean Cd concentration decreased with increasing soil depth. The highest mean concentration of 0.021 mg kg^{-1} was measured in the 0–20 cm layer, and the lowest of $0.0144 \text{ mg kg}^{-1}$ was in the deep layer (80–100 cm). However, the maximum concentrations in our soil survey varied from a minimum of 0.152 mg kg^{-1} in the 60–80 cm layer to a maximum of 0.338 mg kg^{-1} in the 0–20 cm soil layer without a vertical trend. The minimum mean Cd concentration did not show any variation pattern with soil depth. Overall, the Cd concentration showed high and similar variation, regardless of the layer, as indicated by its CV%, which ranged from 113% to 125% with an average of 119%.

Table 3. Measured Cd concentration and its variation with depth.

Depth (cm)	N	Cd (mg kg^{-1} , Mean $\pm$ SE)	CV (%)	Max (mg kg^{-1})	Min (mg kg^{-1})
0–20	385	0.022 ± 0.001 ^A	122	0.338	0.001
20–40	385	0.018 ± 0.001 ^B	117	0.185	0.002
40–60	385	0.017 ± 0.001 ^{BC}	119	0.163	0.001
60–80	373	0.015 ± 0.001 ^{BC}	113	0.152	0.001
80–100	352	0.014 ± 0.001 ^C	125	0.175	0.001
Mean		0.018			
0–100	352	0.029 ± 0.002	117	0.272	0.001
p-value		0			
CV%		122.89	119		

3.2. Major Variables Influencing the Cd Concentration

Further analysis clearly showed that the Cd concentration in each layer was significantly correlated with adjacent layers (all $p < 0.05$), although the correlations between layers 1 and 5 and between layers 2 and 4 were relatively weak. This further verifies the decreasing trend over depth in Table 3.

By considering all measured factors in the soil samples and analyzing the results with Pearson correlation, we found that the Cd concentration in each layer was more or less correlated with some impacting factors of each category, including forest condition, surface soil properties (soil texture, pH, bulk density, stone percentage, etc.), and some layer properties, as shown in Table 4. (1) Most soil properties on the coarse soil map were significantly correlated with Cd concentration, except for TP20 and AK20 in the 0–60 cm layer and the sand percentage in the 80–100 cm layer. (2) Layer-specific Cd concentrations were more or less correlated with some soil properties in the particular layer, such as (i) other metal contents in the same layer, except for Ni, which was only correlated within the 60–100 layer; (ii) TK and TP contents in most soil layers, except for TP in the 0–20 and 80–100 cm layers and TK in the 80–100 cm layer; and (iii) particle size of soil only for the 0–20 cm layer. The soil organic contents and pH had very limited correlations with Cd concentration. (3) The DEM-derived variables and forest conditions were less correlated with Cd concentration, especially slope, PSR, and VSP in some layers. It should be noted that, when different numbers of variables from different impacting factor categories were included in the correlation analysis, the results varied. For example, DEM-derived variables contributed less to the Cd concentration when compared with the soil properties in the specific layer. However, when directly correlating only DEM-derived variables with Cd concentration, some DEM-derived variables contributed more than other variables.

Table 4. Pearson correlation between Cd concentration and surface soil properties from the coarse soil map and soil layer properties measured in soil samples at different depths.

Factor Categories	Variable	Cd					
		0–20 cm	20–40 cm	40–60 cm	60–80 cm	80–100 cm	0–100 cm
Forest condition	forest1	0.0388	−0.0006	0.0802	0.0892	0.0656	−0.0565
	forest2	0.0921	0.0418	0.1068 *	0.1208 *	0.0523	0.0181
	forest3	0.0885	0.0034	0.0302	0.0912	0.0358	0.0308
	forest4	0.0438	0.0597	0.0145	−0.0021	0.0494	0.017
Properties defined by coarse soil map	TK20	0.1819 *	0.1374 *	0.1766 *	0.115 *	0.1251 *	0.1485 *
	TP20	−0.0477	−0.0298	−0.072	−0.0528	−0.0542	−0.0555
	TN20	0.2026 *	0.1782 *	0.1388 *	0.1741 *	0.2422 *	0.2936 *
	AP20	0.0792	0.1304 *	0.0286	0.061	−0.1753 *	0.1593 *
	AN20	0.2399 *	0.2562 *	0.1455 *	0.2076 *	0.0195	0.2964 *
	AK20	−0.0937	−0.0718	−0.0943	−0.1337 *	0.1723 *	−0.0589
	Clay20	0.3114 *	0.2607 *	0.1946 *	0.2107 *	0.2023 *	0.2643 *
	Sand20	−0.1155 *	−0.1342 *	−0.121 *	−0.1356 *	−0.0683	−0.0432
	SOM20	0.1631 *	0.1202 *	0.1282 *	0.1617 *	0.2532 *	0.2108 *
DEM derived landscape feature	Aspect	0.0608	0.046	0.0363	0.0134	0.04	0.0183
	Slope	−0.0932	−0.0359	−0.0449	−0.0918	−0.0849	−0.1624 *
	SDR	0.0191	0.016	0.0595	−0.0245	−0.0471	−0.004
	VSP	0.0461	0.112 *	0.093	0.023	0.0318	−0.0284
	STF	0.0323	0.0308	0.0612	0.0345	0.0493	0.0121
	PSR	0.102 *	0.0906	0.0784	0.0929	0.1242 *	0.1231 *
	Length	−0.0364	−0.0412	−0.0353	0.0186	0.0336	−0.006
	TPI	0.0189	−0.0224	−0.0586	−0.0697	−0.056	−0.0584
	FD	−0.0782	−0.0897	−0.0607	−0.0192	−0.0517	−0.0356
Soil layer properties	bulk	−0.1092 *	0.0245	0.0045	0.0287	−0.041	n/a
	pH	0.0044	0.0769	0.0552	0.0239	0.1237 *	0.1026 *
	som1	0.0913	0.0066	0.142 *	0.0487	0.0037	0.1218 *
	st051	−0.0774	−0.0622	−0.0516	0.0181	0.0161	0.0141
	st011	−0.1297 *	−0.0158	−0.0724	−0.0466	−0.0271	−0.0822
	st0051	−0.1492 *	−0.0424	−0.0598	−0.0575	−0.0245	−0.0784
	st0011	−0.1435 *	−0.0455	−0.0562	−0.0526	−0.0189	−0.0928
	TN	0.0904	−0.0221	−0.0006	0.0588	0.0231	0.0819
	TP	0.0697	0.1666 *	0.2246 *	0.0531	0.1536 *	0.1897 *
	TK	0.2511 *	0.1873 *	0.1925 *	0.2022 *	0.0859	0.2097 *
	AN	0.0108	−0.1247 *	−0.0558	−0.0593	0.0539	0.0153
	AP	0.053	0.0504	0.0711	0.0238	0.0287	0.002
	AK	0.0924	0.0469	0.0381	0.085	0.0666	0.0863
	Cu	0.1403 *	0.3705 *	0.1939 *	0.2204 *	0.2065 *	0.1505 *
	Zn	0.2404 *	0.2168 *	0.2423 *	0.2998 *	0.2301 *	0.1863 *

Table 4. Cont.

Factor Categories	Variable	Cd					
		0–20 cm	20–40 cm	40–60 cm	60–80 cm	80–100 cm	0–100 cm
	Pb	0.2484 *	0.4872 *	0.344 *	0.4605 *	0.4652 *	0.3534 *
	Ni	0.0688	0.0749	0.0395	0.1393 *	0.1753 *	0.1226 *

Note: * with shaded cells indicates significant correlation at the 0.05 level. The first column indicates the variable group analyzed in the Pearson correlation analysis. More detailed information about 368 the variable names can be found in Supplementary Table S1.

3.3. ANN Model Performance for Predicting Soil Cd Concentrations in Different Soil Layers in the Study Area

Based on the analysis, soil texture information included in the coarse soil map and DEM-derived variables were used to develop ANN models for different soil layers.

As shown in Table 2, to achieve a reasonably good estimate (low RMSE, $r^2 > 0.6$, and highest ROA 10%) of Cd concentration in each layer, about seven DEM-derived variables were needed, although there was a slight change in the excluded variables in models for different depths. (1) For the 0–20 cm layer, SDR and STF were excluded, and the rest of the variables remained. However, it should be noted that the most important variable was slope, which, when used as the sole variable, could achieve an r^2 of 0.655 and the lowest RMSE, although ROA 10% and ROA 20% were lower than those obtained when more variables were included. Then, with more variables included, the model was continuously improved until seven variables were used. When SDR and STF were introduced, model performance was significantly degraded. (2) For the 20–40 and 40–60 cm layers, SDR was the best explanatory variable compared with other variables. However, aspect and SDR were excluded in the 20–40 cm and 40–60 cm layers when seven variables were used to predict Cd concentration. This is interesting and may mean that the SDR's influence may be replaced by a combination of other variables. According to ROA and r^2 , the model for the 20–40 cm layer was more effective than that for the 40–60 cm depth. (3) For the 60–80 and 80–100 cm layers, slope regained importance in predicting Cd concentration. In both layers, SDR and Flow direction were excluded from the best model and showed almost identical predictive capabilities. (4) For the 0–100 cm layer, the flow direction and STF were excluded from the selected model.

In summary, either slope or SDR appeared to be the most important factor in predicting soil Cd concentration when a single variable was used, depending on the soil layer, and they achieved an r^2 value of 0.40–0.60. The best model was obtained when seven candidate variables were included, achieving an r^2 value of 0.77–0.88, which varied among different soil layers. The commonly included variables were slope, VSP, Flow length, and PSR for all soil layers. According to the 10-fold ANN model results, with seven DEM-derived variables included, these models can produce good predictions. Around 21–33% of predictions had an error less than ROA 10%, while 36–49% had an error less than ROA 20%. Given the complexity of the Cd distribution within the forest in the study area, these predictions are very good.

To further examine the ANN model performance, we mapped Cd concentrations using the best model for each layer and the mean of 10-fold runs, as shown in Figures 2 and 3. Overall, the single model prediction showed a higher Cd concentration for each layer in comparison with the mean of 10-fold runs. However, these maps generally displayed good agreement with the patterns in soil survey data. Overall, the ANN models developed for each layer achieved at least 70% accuracy compared with verification measurements. While 10-fold model runs were comparable, the single run seemed to more reasonably estimate soil Cd concentration, regardless of soil depth.

To further explore the effect of the above factors on the Cd concentration in each layer, a multi-variate backward stepwise regression analysis was conducted to obtain further information. As shown in Table 5, with multi-variate models, only 26–36% of Cd variance could be explained by the variation in variables from each impacting factor

category, depending on the soil layer. For the current prediction accuracy, more variables were generally involved in the upper layers than the deeper layers, but more variables were used to predict the Cd concentration in the 40–60 cm layer. For the mixed layer, 13 variables with significant contribution to the model were used, which was the same as that for the 40–60 cm layer.

Figure 2. Prediction of soil Cd concentrations (μg mg⁻¹) for 5 soil depths with selected ANN model.

The partial correlation analysis of the soil layer Cd concentration with DEM-derived variables indicated that some DEM-derived variables were helpful in predicting the soil layer Cd concentration, while others were less useful. Variable PSR, STF, and slope each explained a certain degree of the variability in Cd concentration, regardless of the soil layer. SDR and VSP played relatively important roles in the determination of Cd concentration in the 0–60 cm soil layer. Some variables (i.e., SDR, PSR, Aspect, STF) were positively related to the soil concentration in each layer, and some were negatively related (i.e., slope, FD), while the rest were either positively or negatively (i.e., VSP, Length, TPI) related to the Cd concentration, depending on the soil depth. For the composite samples, slope and PSR were the dominant predictors of the Cd concentration, followed by VSP and FD, Aspect and STF, and SDR and Length.

The backward stepwise regression results in Table 5 show that the Cd distribution was significantly related to at least two predictors. However, the relatively highly related predictors determined through regression for each layer were not entirely consistent with the dominant predictors determined through Pearson correlation (Table 4). The regression analysis suggested that the most frequently used predictor was VSP, except for the 80–100 cm layer. SDR was the secondary determinant of Cd concentration, as it was a significant predictor for the 20–40, 40–60, and 80–100 layers. Slope was determined to be a significant predictor for the 0–20, 60–80, and 0–100 layers. In contrast, Pearson correlation analysis suggested that more predictors determined spatial and vertical Cd concentration, including Pb, Zn, Cu, TK, and TP contents and some DEM-derived variables in a specific layer. Interestingly, the Cd concentrations at different depths were well correlated with the corresponding soil layer texture (e.g., clay, sand content); soil TN, TP, and TK content; and organic matter content in the 0–20 cm soil layer recorded in the coarse soil map. In comparison, soil texture played a more important role than soil nutrients and DEM-derived variables in determining soil Cd concentration.

Figure 3. Mean prediction of soil Cd concentrations ($\mu\text{g mg}^{-1}$) at different depths with the selected best model from 10 runs.

Table 5. Coefficients for predicting soil Cd concentration at different depths with stepwise regression models *.

	Cd1 (0–20 cm)		Cd2 (20–40 cm)		Cd3 (40–60 cm)		Cd4 (60–80 cm)		Cd5 (80–100 cm)		Cd0 (0–100 cm)	
	Var. **	Coe.	Var.	Coe.	Var.	Coe.	Var.	Coe.	Var.	Coe.	Var.	Coe.
DEM derived variable	Con.	−52.72	Con.	−20.07	Con.	−35.65	Con.	−5.84	Con.	−41.245	Con.	−74.12
	VSP	0.23	VSP	0.18	VSP	0.27			PSR	0.02	STF	−1.58
	Slope	−0.33			Slope	−0.27					Slope	−0.45
Coarse soil map variable	Sand20	0.36	Sand20	−0.33	TK20	5.92	Sand20	−0.21			TK20	1.21
			TN20	12.06	TN20	177.04	TN20	13.40	TN20	17.51	TN20	89.24
					SOM20	−7.05					SOM20	−2.77
					AN20	−0.39					AK20	−0.22
					AK20	−0.57						
	Clay20	1.90			Clay20	−2.30						
Soil properties at the depth	SOM1	0.22	bulk2	12.68	SOM3	0.38	tk4	0.30			pH0	10.04
	tk1	0.67	tp2	11.23	tp3	21.67					SOM0	0.28
	Zn1	0.33	tk2	0.51	tk3	0.32					tk0	0.58
	Pb1	0.11	Zn2	0.25	Zn3	0.23	Cu4	0.40	Cu5	0.48	Cu0	0.35
	st0051	−0.21	Pb2	0.26	Pb3	0.40	Pb4	0.63	Pb5	0.75	Pb0	0.44
											st050	0.78
											st010	−0.71
Other factors	forest1	5.75			forest1	5.44						
Adjusted r^2		26.6		35.9		28.4		27.9		28.9		28.0

Notes: * The variable coding can be found in Supplementary Table S1. ** Var. is variable; Coe. is Coefficient; Con. Constant.

The principal component analysis (Supplementary Table S2) shows that, overall, 38–39 components were needed in order to explain the total variance in Cd concentration for different layers. However, the first three components explained about 29% of the total variance in the Cd concentration in the 0–100 cm soil layer, which is about the amount of variance in the Cd concentration explained in the multi-variate model for this layer. The first 10 components could explain more than 62% of the total variance in Cd concentration. Using the data of the 80–100 layer as an example (Table 5) and an absolute value of 0.3 as a loading threshold in PCA, the first component was dominated by the soil layer particle size and the second component was negatively related to TN in the coarse soil map. The third component consisted of clay content, TN, and organic content in the top 20 cm layer, which was represented by the information in the coarse soil map. The results of these analyses do not agree with multi-variate analyses and Pearson analyses, although they are similar analysis approaches. This highlights the determinant complexity of Cd distribution. However, this complexity also suggests that the spatial distribution of Cd may be explained by DEM-derived variables and could be more or less related to the coarse soil map information and soil properties within layers. On the other hand, the vertical variation was more or less related to the local variations in soil properties (e.g., the contents of soil heavy metal elements).

4. Discussion

4.1. Implication of Cd Measurement in the Forest Ecosystem in Yunfu Area

The soil survey showed that the average Cd concentration in this region was at the lower end of the global Cd concentration range, although the maximum concentration was as high as 0.4 mg kg^{-1} . Recently, Wuana and Okieimen [14] published a very intensive review indicating that the total Cd concentration in the soil varied from 0.1 to 345 mg kg^{-1} . Our measurements are generally lower than their reported range, which is explained by the measurement approach and less disturbed soil condition. In our study, the concentrations were only obtained from the solid portion. However, our data are consistent with most available Chinese data. For example, based on soil samples from 30 provinces of China, Wang et al. [7] indicated that the Chinese soil Cd concentrations ranged from 0.003 mg kg^{-1} to 9.57 mg kg^{-1} , and the soil Cd concentrations could be ranked from high to low as northwest > southwest > south central > east > northeast > north. Given the fact that Cd ore is mainly distributed in central, southwestern, and eastern China, our site is located in the low-Cd area, and our measurements are within the general range. In another study carried out in the typical industrial area of Shenzhen of Guangdong province, the reported mean soil Cd concentrations were 0.0169 to $0.4400 \text{ mg kg}^{-1}$, and there were generally no pollution concerns in comparison with the secondary pollution threshold of 0.3 mg kg^{-1} [53]. In contrast, the forest soils in this study showed a few spikes exceeding the secondary pollution threshold. Overall, the risk of Cd pollution in this region can be considered negligible.

Cd concentrations showed a vertical decreasing trend. The topsoil had the highest mean Cd concentration, which was 150% higher than the concentration in the 80–100 cm layer, and the maximum Cd concentration was measured in the top layer. This may suggest that Cd was either sourced from outside or from secondary sources resulting from other factors (e.g., land use). A vertical decreasing trend of Cd concentration in soil has been demonstrated in temperate forests in Japan [54], in which Cd concentrations decreased from 0.021 in the A horizon to 0.003 mg kg^{-1} in the C horizon. That study also concluded that the Cd content in soils can be changed by air and water pollution or heavy-traffic roads. Therefore, it is possible that the vertical decreasing trend of Cd concentration in soil profiles may also be partially attributed to factors other than the parent material [55]. While air and pollution could be responsible for the higher Cd level in the topsoil layer, heavy-travel roads have been considered an additional cause of deposition. Globally, approximately 85–90% of total airborne Cd emissions arise from anthropogenic sources, mainly from

the smelting and refining of nonferrous metals, fossil fuel combustion, and municipal waste incineration [56].

4.2. Leading Factors Impacting the Concentration and Distribution of Cd in Forest Soils

This study indicated that DEM-derived variables, soil layer properties, forest type, and topsoil properties defined by the coarse soil map were all more or less responsible for the soil Cd concentration to a certain extent. However, in each category of determining factors, some factors played a more important role than others. It should be pointed out that some unconsidered variables (human pollution, land use history of sampling site, etc.) may also play some role in the distribution of Cd in the study area, which we could not determine in the current study.

(1) The vertical distribution of Cd is more affected by the Cd contents from topsoil layers, as Cd showed a decreasing trend from the topsoil to the deeper soil, which can be confirmed by the high correlation between the Cd concentrations in different layers. This may suggest that Cd was initially or geologically sourced from horizontal transportation. Then, over time, plants or chemical reactions within the soil layers might have created vertical gradients of Cd concentrations along the soil profile. The correlation between the layer Cd concentration and other heavy metal contents could more or less explain some of the vertical variability in the Cd distribution, as reported in a previous study [57].

(2) The soil texture (e.g., clay %), soil pH, metal contents, and soil organic matter contents could be related to the layer Cd concentration. These findings are in good agreement with current documentation. According to Canadian soil quality guidelines [20], a variety of factors influence the mobility of Cd in soils, with pH and soil type (including particle size, content of metal oxides, hydroxides, oxyhydroxides, and organic matter content) probably being the most important. The same document indicates that a number of processes also have the potential to affect the fate of Cd in soils, including aeolian transport (wind erosion), fluvial transport, leaching, and uptake by terrestrial organisms. Another study also identified soil pH as an important factor influencing the mobility of Cd in soil [58].

(a) In our study area, the soil pH varied spatially from 2.91 to 6.07 with an average of 4.15, indicating that the soil is typically acidic. It might be possible that the vertical transport of Cd is partially controlled by the soil pH because the mean layer pH value showed an increasing trend: from 4.10 in the 0–20 cm soil layer to 4.3 in the 80–100 soil layer. However, neither Pearson correlation nor multi-variate modeling analysis supported a significant role for pH in determining the layer Cd concentration. This result re-affirms a previous study [4]. A specific study on the pH impact on soil Cd content may be beneficial.

(b) Our study further showed that the soil organic matter contents within the top 20 cm layer were correlated with the Cd concentration of each layer but less correlated with the organic content measured through soil sampling from a given layer. However, a linear regression analysis between soil layer organic matter content and the layer Cd concentration suggested a strong linear relation ($r^2 = 0.95$; Table 6). This seems to contradict the total correlation analysis with all factors included. This could possibly be attributed to the fact that the soil organic content in the top 20 cm from the coarse soil map was averaged and aggregated from a large area at a large scale, reflecting a spatial trend. The high correlation between the soil layer organic content and corresponding layer Cd concentration may have been overshadowed by other variables or interactions among variables. In a study in South Korea, simple and multiple linear regression modeling for forest soils suggested that 42% of the variability in Cd concentration could be accounted for by pH and organic matter content [19].

(c) There may be an interaction among soil organic matter contents, heavy matter contents, and pH, as suggested by a previous study [59].

Our study suggested that the Cd concentration in each layer was highly correlated with the layer Cu, Pb, Ni, and Zn contents, while they were less correlated with local layer organic matter contents. Our correlation and PCA analyses suggested that both a certain soil particle size in a given soil layer and the clay and soil organic contents in the top 20 cm

layer had a strong impact on the Cd concentration. The deeper soil contained less organic matter, which minimized its correlation with the Cd concentration. On the other hand, soil organic matter has a binding effect on heavy metals, and it is expected that stronger binding activities occur in the upper soil layers than deeper soil layers.

Table 6. Soil organic matter content and Cd concentration.

Layer (CM)	OM (1/000)			Cd		
	Mean	SE	C.V.	Mean	SE	C.V.
0–20	24.17	0.35	56.25	21.60	1.34	122.14
20–40	16.79	0.43	50.82	17.56	1.05	117.35
40–60	13.71	0.38	54.25	17.17	1.04	119.30
60–80	12.06	0.37	59.65	15.38	0.90	113.16
80–100	10.40	0.36	64.72	15.71	1.05	125.05

4.3. The Challenge for Predicting Soil Cd Concentration with Multiple-Variable Regression and ANN Model

Both multiple linear regression models (MLRMs) and ANN models showed limited alignment in terms of the variables used and predictive capability. For the 0–20 cm layer, MLRM provided the highest prediction, explaining 33% of the total Cd variation with variable slope, FD, and VSP. In comparison with MLRMs, ANN models achieved a better prediction with the addition of four more variables (i.e., TPI, aspect, length, and PSR), as illustrated by the statistical parameters ($r^2 = 0.882$, ROA 20% (ratio) = 48%, RMSE = 0.623). As shown in Table 7, further verification of 64 data points indicated that ANN models reasonably predicted Cd concentration. It is reasonable that slope, VSP, and FD are related to the landscape transport of solid material in soils. One interesting finding is that VSP was indicated as the most important factor in determining the vertical distribution of Cd for the 20–40, 40–60, and 60–80 cm layers in MLRM, and with a different number of variables incorporated, the models explained 21–24% of the variability in the Cd concentration. In the 80–100 cm layer, it was found that the slope and STF could explain 33% of the variability in Cd concentration using MLRM. On the other hand, all of the ANN models used seven variables and achieved an ROA 20% (ratio) of 36–48%. This comparison may suggest that the performance of the ANN model is better than that of MLRM. In a recent study, Nadari et al. [32] used stepwise multiple linear regression (MSLR) and a neural network-genetic algorithm model (ANN-GA) based on satellite imagery to identify the source of heavy metals, including Cd, with 300 soil samples. The authors found that the ANN-GA model demonstrated higher accuracy than multiple linear regression. They also found higher concentrations of Cd in soils under industrial lands and irrigation farming in comparison with barren and dryland farming owing to the accumulation of industrial wastes in roads and streams. In our case, the advanced industry with rice cultivation may have created a heterogenic distribution of the Cd concentration due to wind erosion or atmospheric deposition into the forest area [60]. A field monitoring study in the forest soils of the Bydgoszcz and Torun provinces, Poland, also showed that Cd accumulated mainly in the litter horizons and was correlated with soil acidity, organic matter and clay contents, and cation exchange capacity [61]. Given the low predictability of our models in this study, we can assume that some excluded variables were needed in order to improve the prediction of Cd concentrations in the forest soils in this region.

Our study positively affirmed the findings of many previous studies [4,5,7], which indicated some correlation between Cd concentration and soil pH in the studied forest soil. However, our ANN model and MLRM were designed to ignore those detailed factors, as shown in Tables 4 and 6, for the purpose of using spatial variables linked to soil Cd concentration, which may have constrained the improvement of the prediction accuracy of the models. Obviously, including factors such as distances to industrial centers, heavy-traffic roads, and irrigation farming areas, soil pH, and organic matter content may substantially improve the model performance. However, we intended to develop simplified tools for the

assessment of spatial variability in Cd distribution and contamination. Furthermore, the Cd distribution may have been potentially governed by many factors, such as large-scale landscape variables (DEM-related variables, soil-forming factors, and historical topographic changes). The difficulties in obtaining soil-forming data and topographic changes make it impossible to collect all possible predictors.

Table 7. Comparison of model predictions (single run of the selected model and the mean from 10 modeled maps) against field measurements of 64 verification points.

Layer (cm)	Measurements (mg kg ^{−1})	Single Model		Mean of 10 Fold Runs	
		Estimation (mg kg ^{−1})	Relative Error (%)	Estimation (mg kg ^{−1})	Relative Error (%)
0–20	0.022	0.022	0	0.016	−27
20–40	0.018	0.023	28	0.022	22
40–60	0.017	0.019	12	0.018	5.8
60–80	0.015	0.012	−20	0.008	−47
80–100	0.014	0.010	−29	0.012	−15
0–100	0.029	0.024	−17	0.021	−29

5. Conclusions

Toxic Cd contamination in soil has important implications for human and animal health. As the major ecosystem in Yunfu, Guangdong, China, the forest acts as either a source or a sink of Cd, playing a role in the life and history in the region.

According to our comprehensive survey of 385 soil pits, Cd in the studied forest soils showed a spatial pattern associated with the topographic condition along the landscape. The overall average concentrations of Cd were generally at the medium and low ends of the global distribution range with a very marginal pollution level, although some spike values may be beyond the current threshold of contamination level. The vertical decreasing trend of Cd concentration and its high dependence on the levels immediately above may also be associated with the increase in pH from the topsoil layers to the deeper soil layers, and it suggests that Cd may also be attributable to air deposition or geological transportation along landscapes. With the rapid urban sprawl and intensive farming, the Cd concentration in affected soils is expected to increase over time. Furthermore, the Cd concentration in a given layer was correlated with some soil properties (soil texture, particle size, etc.) in the top 20 cm layer and the content of other heavy metals in the analyzed layer.

Multiple regression and ANN models showed that some DEM-derived variables were correlated with the Cd distribution. However, multiple regression explained a maximum of 33% of the variation in Cd concentration with fewer than five variables, while ANN models provided better predictions of soil Cd with seven variables, regardless of soil depth. It should be noted that the performance of ANN models may be improved by introducing soil pH, organic content in the topsoil, and distances to industrial areas and irrigation farming, which may be a promising research focus for developing more accurate models in the future.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/land10090906/s1>, Table S1: Variable name and definition in the study, Table S2: Principal component analysis for 0–100 cm soil layer.

Author Contributions: Conceptualization, X.D.; methodology, Z.Z.; validation, X.L., formal analysis, Z.X., S.L. and X.L.; investigation, S.L. and X.L.; writing—original draft preparation, Z.X.; writing—review and editing, Z.X., Z.Z. and Y.L.; visualization, Z.Z. and Z.X.; supervision: X.D.; project administration, X.D.; funding acquisition, X.D. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by funding from Guangdong Forestry Science and Technology Plan of China (Grant No. 2019-07) and the Natural Science Foundation of Guangxi, China (Grant Nos. 2016GXNSFCA380029 and 2018GXNSFBA138035).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article or Supplementary Materials.

Acknowledgments: The authors are grateful for the support from the soil survey team of Forest Research Academy of Guangdong for their field efforts in sampling and lab processing of samples.

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

References

1. Mitrikeski, J.; Mitkova, T.; Stikova, E. Contents of soluble forms of heavy metals (cu, cd, pb, zn and co) in cinnamonic forest soils in kumanovo and prilep region. *Zb. Zemjod. Fak. Univ. St. Kiril Metod. Skopje* **1997**, *42*, 11–15. (In Macedonian)
2. Vuksic, N.; Speranda, M. Distribution of heavy metals (cd, pb, hg, as) and essential elements (fe, se) in forest soil and plant communities of the state open hunting area “krndija ii” xiv/23. [croatian]. *Šumarski List* **2016**, *140*, 147–153.
3. Cao, S.; Anxiang, L.; Jihua, W.; Lili, H. Modeling and mapping of cadmium in soils based on qualitative and quantitative auxiliary variables in a cadmium contaminated area. *Sci. Total Environ.* **2016**, *580*, 430–439. [[CrossRef](#)] [[PubMed](#)]
4. Andersen, M.; Refsgaard, A.; Raulund-Rasmussen, K.; Strobel, B.; Hansen, H. Content, distribution, and solubility of cadmium in arable and forest soils. *Soil Sci. Soc. Am. J. SSSAJ* **2002**, *66*, 1829–1835. [[CrossRef](#)]
5. Zhang, X.; Chen, D.; Zhong, T.; Zhang, X.; Cheng, M.; Li, X. Assessment of cadmium (cd) concentration in arable soil in china. *Environ. Sci. Pollut. Res.* **2015**, *22*, 4932–4941. [[CrossRef](#)] [[PubMed](#)]
6. He, X.; Yan, M.; Mo, R.; Lyu, J. Assessment of potential ecological hazards of heavy metals in the soils in the areas of edible forest products in xinyu city. *Acta Agric. Univ. Jiangxiensis* **2018**, *40*, 423–428. (In Chinese)
7. Wang, L.; Cui, X.; Cheng, H.; Chen, F.; Wang, J.; Zhao, X.; Lin, C.; Pu, X. A review of soil cadmium contamination in china including a health risk assessment. *Environ. Sci. Pollut. Res.* **2015**, *22*, 16441–16452. [[CrossRef](#)] [[PubMed](#)]
8. Qadir, S.; Jamshieed, S.; Rassol, S.; Ashraf, M.; Akram, N.A.; Ahmad, P. Modulation of plant growth and metabolism in cadmium-enriched environments. *Rev. Environ. Contam. Toxicol.* **2014**, *229*, 51–88. [[CrossRef](#)]
9. Yadav, S.K. Heavy metals toxicity in plants: An overview on the role of glutathione and phytochelatins in heavy metal stress tolerance of plants. *S. Afr. J. Bot.* **2010**, *76*, 167–179. [[CrossRef](#)]
10. Chrzan, A.A. The impact of heavy metals on the soil fauna of selected habitats in niepolomice forest. *Pol. J. Soil Sci.* **2017**, *50*, 291–300. [[CrossRef](#)]
11. Järup, L.; Berglund, M.; Elinder, C.G.; Nordberg, G.; Vanter, M. Health effects of cadmium exposure—A review of the literature and a risk estimate. *Scand. J. Work Environ. Health* **1998**, *24*, 1–51.
12. Johri, N.; Jacquillet, G.; Unwin, R. Heavy metal poisoning: The effects of cadmium on the kidney. *BioMetals* **2010**, *23*, 783–792. [[CrossRef](#)]
13. Elinder, C.-G. Uses, Occurrence, and Intake. In *Cadmium and Health: A Toxicological and Epidemiological Appraisal*; CRC Press Inc.: Boca Raton, FL, USA, 1985.
14. Wuana, R.A.; Okieimen, F.E. Heavy metals in contaminated soils: A review of sources, chemistry, risks and best available strategies for remediation. *ISRN Ecol.* **2011**, *2011*, 402647. [[CrossRef](#)]
15. Szymanska-Pulikowska, A.; Kucharzewski, A.; Nowak, L. Pollution of forest soils in lower silesia with heavy metals and sulfur. [polish]. *Acta Sci. Pol. Form. Circumiectionis* **2004**, *3*, 77–87.
16. Vytopylova, M.; Tejnecky, V.; Boruvka, L.; Drabek, O. Sorption of heavy metals in organic horizons of acid forest soils at low added concentrations. *Soil Water Res.* **2015**, *10*, 1–9. [[CrossRef](#)]
17. Guo, Z.-H.; Liao, B.-H.; Huang, C.-Y. Mobility and speciation of cd, cu, and zn in two acidic soils affected by simulated acid rain. *J. Environ. Sci.* **2005**, *17*, 332–334.
18. Atteia, O.; Thélin, P.; Pfeifer, H.R.; Dubois, J.P.; Hunziker, J.C. A search for the origin of cadmium in the soil of the swiss jura. *Geoderma* **1995**, *68*, 149–172. [[CrossRef](#)]
19. Byun, J.; Yoo, J.; Kim, C.; Jeong, J.; Lee, B. Estimation of heavy metal concentrations by soil property of forest soils in seoul. *FRI J. For. Sci.* **1999**, *61*, 97–101.
20. CCME. Canadian soil quality guidelines for the protection of environmental and human health: Cadminum. Winnipeg. In *Canadian Environmental Quality Guidelines*; Canadian Council of Ministers of the Environment: Winnipeg, MB, Canada, 1999.
21. Fisak, J.; Skrivan, P.; Tesar, M.; Fottova, D.; Dobesova, I.; Navratil, T. Forest vegetation affecting the deposition of atmospheric elements to soils. *Biologia* **2006**, *61*, S255–S260. [[CrossRef](#)]
22. Luster, J.; Zimmermann, S.; Zwicky, C.N.; Plienemann, P.; Blaser, P. Heavy metals in swiss forest soils: Modification of lithogenic and anthropogenic contents by pedogenetic processes, and implications for ecological risk assessment. *Geol. Soc. Lond. Spec. Publ.* **2006**, *266*, 63–78. [[CrossRef](#)]
23. Belanovic, S.; Kosanin, O.; Danilovic, M.; Kadovic, R. Accumulation of heavy metals as related to cation exchange in some forest and pasture soils of stara planina (serbia). *Istanb. Univ. Orman Fak. Derg. Ser. A* **2008**, *58*, 1–11.
24. Mohammed, F.A.E.; Bchitou, R.; Boulmane, M.; Bouhaouss, A.; Guillaume, D. Modeling of the distribution of heavy metals and trace elements in argan forest soil and parts of argan tree. *Nat. Prod. Commun.* **2013**, *8*, 21–23. [[CrossRef](#)]

25. Liu, X.; Wu, J.; Xu, J. Characterizing the risk assessment of heavy metals and sampling uncertainty analysis in paddy field by geostatistics and gis. *Environ. Pollut.* **2006**, *141*, 257–264. [[CrossRef](#)] [[PubMed](#)]
26. Alam, N.; Ahmad, S.R.; Qadir, A.; Ashraf, M.I.; Lakhan, C.; Lakhan, V.C. Use of statistical and gis techniques to assess and predict concentrations of heavy metals in soils of lahore city, pakistan. *Environ. Monit. Assess.* **2015**, *187*, 1–10. [[CrossRef](#)]
27. Asmaryan, S.G.; Muradyan, V.; Sahakyan, L.; Saghatelyan, A.; Warner, T. (Eds.) Development of remote sensing methods for assessing and mapping soil pollution with heavy metals. In *Global Soil Map: Basis of the Global Spatial Soil Information System*; CRC Press/Taylor and Francis Group: Abingdon, UK, 2014; pp. 429–432. [[CrossRef](#)]
28. Boruvka, L.; Vacek, O.; Jehlicka, J. Principal component analysis as a tool to indicate the origin of potentially toxic elements in soils. *Geoderma* **2005**, *128*, 289–300. [[CrossRef](#)]
29. Hong, Y.; Shen, R.; Cheng, H.; Chen, S.; Chen, Y.; Guo, L.; He, J.; Liu, Y.; Yu, L.; Liu, Y. Cadmium concentration estimation in peri-urban agricultural soils: Using reflectance spectroscopy, soil auxiliary information, or a combination of both? *Geoderma* **2019**, *354*, 354. [[CrossRef](#)]
30. Zhang, Z.; Abuduwaili, J.; Jiang, F. Relationship of heavy metals and soil N, P, K and total salts in tianshan mountains, central asia. *Asian J. Chem.* **2013**, *25*, 8971–8975. [[CrossRef](#)]
31. Ghazi, F.F. Estimation of heavy metals contamination in the osil of zaafaraniya city using the neural network. *J. Phys. Conf. Ser.* **2018**, *1003*, 012058. [[CrossRef](#)]
32. Nadari, A.; Mohammad Amir, D.; Babak, K.; Mohammad Sagdegh, A. Assessment of spatial distribution of soil heavy metals using ann-ga, mslr and satellite imagery. *Environ. Monit. Assess.* **2017**, *189*, 214. [[CrossRef](#)]
33. Zhao, Z.; Chow, T.L.; Rees, H.W.; Yang, Q.; Xing, Z.; Meng, F.-R. Predict soil texture distributions using an artificial neural network model. *Comput. Electron. Agric.* **2009**, *65*, 36–48. [[CrossRef](#)]
34. Birkeland, P. *Soils and Geomorphology*, 3rd ed.; Oxford University Press: New York, NY, USA, 1999.
35. Brady, N.C.; Weil, R.R. *The Nature and Properties of Soil*, 14th ed.; Pearson Education Inc.: Upper Saddle River, NJ, USA, 2008.
36. China Soil Survey Office. *Chinese Soil*; Chinese Agriculture Press: Beijing, China, 1988.
37. Guangdong Soil Census Office. *Guangdong Soil*; Science Press: Beijing, China, 1993.
38. Li, X.; Ding, X.; Ceng, S.; Zhang, C.; Yang, H. *Forest Soil Survey of Yunfu, Guangdong Province*; China Forestry Publishing House: Beijing, China, 2018.
39. De Reu, J.; Bourgeois, J.; Bats, M.; Zwervaegher, A.; Gelorini, V.; De Smedt, P.; Chu, W.; Antrop, M.; De Maeyer, P.; Finke, P.; et al. Application of the topographic position index to heterogeneous landscapes. *Geomorphology* **2013**, *186*, 39–49. [[CrossRef](#)]
40. Meng, F.-R.; Castonguay, M.; Ogilvie, J.; Murphy, P.N.C.; Arp, P.A. Developing a gis-based flow-channel and wet areas mapping framework for precision forestry planning. In Proceedings of the Iufro Precision Forestry Symposium 2006, Stellenbosch, South Africa, 5–10 March 2006; pp. 43–55.
41. Ambroise, B.; Beven, K.; Freer, J. Toward a generalization of the topmodel concepts: Topographic indices of hydrological similarity. *Water Resour. Manag.* **1996**, *32*, 2135–2146. [[CrossRef](#)]
42. Zhao, Z.; Benoy, G.; Chow, T.L.; Rees, H.W.; Daigle, J.-L.; Meng, F.-R. Impacts of accuracy and resolution of conventional and lidar based dems on parameters used in hydrologic modeling. *Water Resour. Manag.* **2010**, *24*, 1363–1380. [[CrossRef](#)]
43. Ferro, V.; Minacapilli, M. Sediment delivery processes at basin scale. *Hydrol. Sci. J.* **1995**, *40*, 703–717. [[CrossRef](#)]
44. Meng, F.-R.; Arp, P.A.; Zelazny, V.F.; Colpitts, M.C.; Schivatcheva, T.; Fahmy, S.H. Spatial and temporal variation of soil moisture. *Prog. Rep. Fundy Model For.* **1997**, *4*, 517.
45. Jenson, S.K.; Domingue, J.O. Extracting topographic structure from digital elevation data for geographic information system analysis. *Photogramm. Eng. Remote Sens.* **1988**, *54*, 1593–1600.
46. Greenlee, D.D. Raster and vector processing for scanned linework. *Photogramm. Eng. Remote Sens.* **1987**, *53*, 1383–1387.
47. Akhtar, M.K.; Corzo, G.A.; van Andel, S.J.; Jonoski, A. River flow forecasting with artificial neural networks using satellite observed precipitation pre-processed with flow length and travel time information: Case study of the ganges river basin. *Hydrol. Earth Syst. Sci.* **2009**, *13*, 1607–1618. [[CrossRef](#)]
48. Zhao, Z.; Yang, Q.; Sun, D.; Ding, X.; Meng, F.-R. Extended model prediction of high-resolution soil organic matter over a large area using limited number of field samples. *Comput. Electron. Agric.* **2020**, *169*, 105172. [[CrossRef](#)]
49. Werbos, P.J. *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences*; Harvard University: Cambridge, MA, USA, 1975.
50. Li, Z.Y. Supervised classification of multispectral remote sensing image using b-p neural network. *J. Infrared Millim. Waves* **1998**, *17*, 153–156.
51. Sigillito, V.G.; Hutton, L.V. *Case Study II: Radar Signal Processing Neural Network PC Tools*; Academic Press: Cambridge, MA, USA, 1990; Chapter 11.
52. Hyndman, R.J.; Koehler, A.B. Another look at measures of forecast accuracy. *Int. J. Forecast.* **2006**, *22*, 679–688. [[CrossRef](#)]
53. Feng, Y.; Liu, L.; Xiao, H.; Wu, Q. Pollution characteristics and health risk assessment of heavy metals in soil of typical industrial district of shenzhen. *Ecol. Environ. Sci.* **2017**, *26*, 1051–1058.
54. Memon, A.R.; Itō, S.; Yatazawa, M. Distribution of zinc and cadmium in the temperate forest taxa of central Japan. *Soil Sci. Plant Nutr.* **1980**, *26*, 281–290. [[CrossRef](#)]
55. Ding, Y.; Yu, X.; Zhao, G.; Zhang, Y.; Sun, N.; Du, Y. Analysis of heavy metals in soil of forests of mt. Lushan at different altitudes. *Environ. Sci. Technol.* **2013**, *36*, 191–194. (In Chinese)

-
56. WHO Regional Office for Europe. Cadmium. In *Air Quality Guidelines*, 2nd ed.; European Series, No. 91; WHO Regional Publications: Copenhagen, Denmark, 2000.
 57. Ruan, X.; Zhang, G.; Ni, L.; He, Y. Distribution and migration of heavy metals in undisturbed forest soils: A high resolution sampling method. *Pedosphere* **2008**, *18*, 386–393. [[CrossRef](#)]
 58. Chanmugathas, P.; Bollag, J.M. Microbial mobilization of cadmium in soil under aerobic and anaerobic conditions. *J. Environ. Qual.* **1987**, *16*, 161–167. [[CrossRef](#)]
 59. Quenea, K.; Lamy, I.; Winterton, P.; Bermond, A.; Dumat, C. Interactions between metals and soil organic matter in various particle size fractions of soil contaminated with waste water. *Geoderma* **2009**, *149*, 217–223. [[CrossRef](#)]
 60. Wang, Y.-y.; Liu, D.-m.; Sun, T.; Yu, Z.-w.; Xu, H.-P.; Tang, B. Characteristic analysis and evaluation of heavy metal content in soils around a smelter. *Guangzhou Chem. Ind.* **2019**, *47*, 128–132.
 61. Malczyk, P. Heavy metals content in soils of selected forest ecosystems. *Zesz. Probl. Postępow. Nauk. Rol.* **1996**, *434*, 599–603. (In Polish)