Novel framework for modelling the cadmium balance and accumulation in farmland soil in Zhejiang Province, East China: Sensitivity analysis, parameter optimisation, and forecast for 2050

Tingting Fu, Ruiying Zhao, Bifeng Hu, Xiaolin Jia, Zhige Wang, Lianqing Zhou, Mingxiang Huang, Yan Li, Zhou Shi

Novel framework for modelling the cadmium balance and accumulation in farmland soil in Zhejiang Province, East China: sensitivity analysis, parameter optimisation, and forecast for 2050

Tingting Fu[1,2], Ruiying Zhao[1,2], Bifeng Hu[1,3,4,*], Xiaolin Jia[1,2], Zhige Wang[1,2], Lianqing Zhou[1,2], Mingxiang Huang[5], Yan Li[6], Zhou Shi[1,2]

[1] Key Laboratory of Environment Remediation and Ecological Health, Ministry of Education, College of Environmental and Resource Sciences, Zhejiang University, Hangzhou 310058, China

[2] Institute of Agricultural Remote Sensing and Information Technology Application, Zhejiang University, Hangzhou 310058, China

[3] Unité de Recherche en Science du Sol, INRAE, Orléans 45075, France

[4] Sciences de la Terre et de l'Univers, Orléans University, 45067 Orléans, France

[5] Information Center of Ministry of Ecology and Environment, Beijing 100035, China

[6] Institute of Land Science and Property, School of Public Affairs, Zhejiang University, Hangzhou 310058, China

Corresponding author: Dr. Bifeng Hu
Tel./Fax: +86 571 88283816
Email: bifeng.hu@inra.fr

Abstract

Modelling the mass balance and forecasting the temporal variations of cadmium (Cd) in farmland soil play a critical role in the development of mitigation strategies for Cd pollution. In this study, a novel framework integrating the mass balance model with model-independent parameter estimation, geostatistics, and bagging algorithms were integrated to simulate the long-term changes in the Cd content of farmland soil in Zhejiang Province, China. The predicted Cd content in farmland soil in 2013 was

compared to observed data (R value=0.568 and root-mean-square error=0.177 mg kg^{-1}), demonstrating the feasibility of our model. The prediction results for 2050 indicated that the average concentration of Cd in farmland soil from Zhejiang Province will increase to 0.30 mg kg^{-1} if the current trend continues, and that 37.4% of the farmland soil in the province will be classified as a "security utilisation region", indicating great risk of soil Cd contamination in these areas. Reducing industrial emissions and soil acidification to reduce the Cd pollution risk should receive great attention. This study provides a new perspective for forecasting the temporal trends of Cd accumulation in farmland soil and facilitates improved management and risk prevention of Cd pollution in agricultural soils and products.

Key words:

Cadmium; Mass balance model; Model-independent parameter estimation; Farmland soil; Parameter optimisation

1. Introduction

Heavy metal (HM) contamination of soil has been a broad environmental concern worldwide due to its various sources, high toxicity, and non-biodegradability (Jia et al., 2019; Xia et al., 2019; Hu et al., 2020a; Sciarrillo et al., 2020; Van Pelt et al., 2020). Cd is the most widespread HM pollutant in China, with 7.0% of soil samples polluted by Cd, according to the Chinese government (NSPCIR, 2014). Due to its high mobility, Cd can be easily absorbed by plants and then accumulate in the

human body via the food chain, inhalation of contaminated dust, dermal contact, and consumption of crops grown in polluted soil (DalCorso et al., 2008; Guo et al., 2011; Hu et al., 2017a; Ronzan et al., 2018). Cd a carcinogen that can damage the functioning of some human organs (Sawut et al., 2018; Shen et al., 2019; Xiang et al., 2019). Hence, modelling Cd accumulation and forecasting its temporal trends in agricultural soil are essential.

However, the accumulation of Cd in soil is a highly complicated and variable process. Numerous pathways, including atmospheric deposition, wastewater irrigation, rock weathering, agricultural inputs, and industrial waste, contribute to the accumulation of soil Cd (Shao et al, 2018). The mass balance model is a useful tool for modelling and forecasting the accumulation of HMs in agricultural soil (Keller et al., 2001; Six and Smolders, 2014), and some attempts have been made to simulate the long-term trends of HM accumulation in agricultural soil using mass balance models. Keller et al. (2001) used the mass balance model to estimate the accumulation of Cd and Zn in agricultural soils from Sundagau City, south of Basel, Switzerland. Chen et al. (2007) also analysed the long-term transfer of Cd in cropland soils from the Tongzhou and Daxing Districts in southern Beijing, China, using the mass balance model, and found that the strong variations in model parameters contributed to the uncertainty of the prediction result. Six and Smolders (2014) analysed the mass balance of soil Cd across European agricultural land using empirical models, and the uncertainty analysis of five regional scenarios suggested that leaching was the most uncertain flux.

These studies demonstrated the feasibility of using the mass balance model for modelling and forecasting the accumulation of HMs in agricultural soil. However, previous studies had several drawbacks: (1) existing studies limited by time, labour, and economic costs for soil sampling usually lacked independent validation; (2) most existing studies were impeded by high uncertainty related to the variability of model parameters, leading to biased prediction results; and (3) existing studies were impeded by extremely high values when exploring the spatial pattern of HMs, leading to the overestimation of the HM content in nearby regions (Keller et al., 2001; Chen et al., 2007; Six and Smolders, 2014). This study implemented several measures to resolve these issues. First, sensitivity analysis was conducted to identify the parameters with the largest impacts on model performance using the model-independent parameter estimation (PEST) approach (Doherty et al., 2004). The values of several of the most impactful parameters were then optimised through PEST. Second, we utilised the results of extensive soil surveys conducted in 2003 and 2013 to train and independently validate the model, respectively. Third, we utilised the bagging algorithm to conduct 50 simulations, and the results of the 20 simulations with the highest accuracy were selected and averaged to produce a final forecast map of the Cd content in farmland soil in 2050. This approach effectively eliminated the negative influence of local extreme values on the prediction results. Finally, spatial autocorrelation and correlation analyses were conducted to ensure that there was no systematic bias in our model.

The area studied here has undergone rapid industrialisation and urbanisation over

the past four decades, which has led to many environmental issues, particularly soil HM pollution. In this study, we developed a novel framework that combines the mass balance model, PEST method, geostatistical method, spatial autocorrelation analysis, grid computing, and bagging algorithms to model the mass balance of Cd in 2003 and 2013, and forecast the pollution trend of Cd in the soil of Zhejiang Province, southeast China, in 2050. The objectives of our research were to (1) apply the PEST method to identify the most sensitive parameters affecting the prediction results and optimise the values of these parameters to reduce the model uncertainty and bias; (2) model the mass balance of Cd, estimate the spatial pattern of soil Cd in the study region, and validate the model performance based on data for 2013; and (3) forecast the trend of Cd contamination in the study area during 2050. Our study will help policymakers and stakeholders to understand the temporal trends of Cd pollution and more efficiently control Cd pollution in the soils of the research area.

2. Material and methods

2.1 Study area

Zhejiang Province, located in southeast China, covers the area between the latitudes of 27°02′ and 31°11′N and longitudes of 118°01′ and 123°01′E, has a total land area of 104,141 km^2 and a population of 58.50 million people. The elevation decreases from the southwest to the northeast, and rivers cross the whole province. Zhejiang experiences a typical northern subtropical monsoon climate, with a mean annual temperature of 15 to 18 °C and mean annual precipitation of 980 to 2000 mm. Zhejiang Province is located in the southern Yangtze River Delta, and its economy

and industry are highly developed. Additionally, Zhejiang is an important grain production area, with farmland covering an area of 19,746 km^2. The main crops and agricultural products cultivated in Zhejiang are rice, cabbage, broccoli, sweet rice, and watermelon. However, due to the rapid development of local industries, environmental pollution in the study area has become increasingly severe, and soil pollution by HMs is a severe environmental problem.

2.2 Sampling and Data processing

In this study, multi-source heterogeneous data were used to conduct the superposition analysis. Over the past few decades, the grid computing environment has been widely used as a powerful tool for data-intensive and computation-intensive modelling (Hu et al., 2004). Pouchard et al. (2003) reported the growing use of grid computing for obtaining, comparing, and analysing data and synthesising grid computing with spatial data processing methods. Although points may have low spatial density, grid computing can maximise data utilisation, particularly in large-scale studies, and it is also suitable for multi-source, heterogeneous data processing. Sampling points from two time periods can be overlaid on the grids. Although one grid covers more than one point, the value of the grid can be calculated using the average value of several points. On this basis, the forecast results can be improved and the model can be optimised through parameter modification, improving the credibility of the prediction results.

Figure 1.

In this study, topsoil (0–20 cm) samples from the farmland in Zhejiang Province were collected in 2003 and 2013, respectively. The total concentration of Cd was obtained by acid digestion with HF–HNO_3–$HClO_4$. To integrate the soil samples collected in 2003 and 2013, two points with separation distances of less than 300 m were considered to share the same soil pool. Based on this assumption, the study area was divided into grids, and the value of each grid was the average Cd concentration overlaid on the grid by the sampling points for each survey. After data matching, 413 grids were retained in the study area (Fig. 1).

2.3 Novel framework based on the mass balance model

2.3.1 Model description

The flux of Cd in the topsoil constitutes of several mass-balance procedures, including the motion of Cd between soil, water, and plants (Peng et al., 2016). In this study, we considered the main input and output flows of soil Cd, and the pathways with negligible contributions to Cd accumulation in the farmland soil were not considered. Figure 2 presents the main potential input, internal, and output flows of Cd in farmland soil. The main inputs of Cd in soil include wind transportation and deposition, rock weathering, water irrigation, fertiliser application, straw return, air deposition, and solid waste landfilling, while the main outputs include surface drainage, wind transportation, plant uptake, and water leaching. Additionally, solid–liquid conversion occurs in the root-zone soil pool.

Figure 2.

In this study, the mass balance model developed by Keller et al. (2001) was used to model and forecast the movement and accumulation of Cd in the farmland soil of the study area. This method can be described as an empirical stochastic balance model that estimates the HM content in soil at a regional scale considering the spatial variations of the various pathways through which HMs accumulate in soil. Thus, the change in the Cd content (g ha^{-1}) of topsoil (0–20 cm) during a certain period of time Δt (yr) can be obtained as follows (Keller et al., 2001):

$$\frac{\Delta Cd_{ij}}{\Delta t} = I_{ir} + I_{fer} + I_{air} - O_u - O_L + E_r \quad (1)$$

$$I_{ir} = \sum_1^N (N_i \times T_i \times \alpha) \quad (2)$$

$$I_{fer} = \sum_1^N (N_i T_i) \quad (3)$$

where I_{ir} is the total input flux of Cd through irrigation water, considering the effective agricultural water consumption; I_{fer} denotes the input metal flux through pesticide and fertiliser application, including manure, sludge, and lime; I_{air} is the input metal flux through atmospheric deposition; O_u is the metal flux removed from soil through the harvested crop; O_L represents the HM flux through the vertical movement of water in soil, defined as leaching; N_i is the quantity of fertiliser, irrigation, pesticides, and limestone, and is a fixed value according to local farming customs; T_i is the concentration of Cd in the fertiliser, irrigated water, pesticides, and limestone; and α is the agricultural water consumption. The input data and default values of the model parameters are listed in Table S1.

The major outputs of the model include the movement of HMs in the solution and solid phases, the output fluxes of HMs via leaching and plant uptake, and a forecast of the probable soil Cd concentration in the near future, which can be compared to the actual value of the predicted year. More details of the mass balance model were provided by Keller et al. (2001) and Peng et al. (2016).

2.3.2 Parameter sensitivity analysis and optimisation

As indicated in previous studies, uncertainty related to parameter variability introduces errors and biases to prediction results (Chen et al., 2007; Six and Smolders., 2014). To determine the high-impact parameters to be used in the model calibration process, sensitivity analysis was conducted to identify parameters that would critically impact the model outcomes. Such analysis measured the combined variations in the model estimates due to a partial variation in the studied parameter (Doherty, 2004). In this study, we employed the PEST method to identify and optimise parameters with the greatest impact on the model construction.

The mass balance parameters were calibrated following a grid-based optimisation method to describe the nonlinear procedures by computing the Jacobian matrix of the sensitivities of the model parameters as implemented in PEST (Doherty et al., 1994). The model communicated through input and output documents, reading parameter values from the input documents and writing the outputs to the output documents. The PEST method then read the output document, calculated the squared error between the observed and predicted values, and revised the parameter values to minimise the weighted sum of the squared deviations. A novel mass balance input

document was then written, the model was executed again, and the same procedure was applied numerous times. Detailed information regarding the PEST method is available in the studies by Doherty et al. (2004) and Rose et al. (2007) and is also presented in the Supplemental Materials.

2.3.3 Model simulation based on bagging algorithm

Over the last several decades, ensemble methods, such as bagging (Breiman, 1996), a bootstrap aggregation method, have been widely used to solve numerous issues related to real-world predictions (Peng et al., 2019; Zhou et al., 2019; Yan et al., 2020). The bagging method employs bootstrap to reproduce the training data set and produce multiple editions of classifier and optimum training data sets. The new classifiers are then combined through plural voting to establish an aggregated classifier for class prediction (Pham and Prakash., 2019). In this study, we employed the bagging algorithm to conduct model simulations and produce prediction results.

2.3.4 Cd content in 2050 based on the bagging algorithm

In this study, subsampling was repeatedly conducted from the training set, forming 50 base models, and the results of the base models that met the required precision were integrated to generate the final prediction results. This aided in smoothing the effects of outliers or extreme values.

Figure 3.

The process of forecasting the soil Cd in 2050 based on the bagging method is shown in Fig. 3 and described as follows: according to the bootstrap method, 70% of

the samples were utilised as a subtraining set through replacement, and the remaining 30% of the samples were used as a subtesting set. After repeating this 50 times, one-to-one subtraining and subtesting set data were obtained. The subtraining set was then input into the mass balance equation to obtain the prediction results. The spatial prediction results were then obtained using the inverse distance weighted (IDW) interpolation operation. This process was repeated 50 times, and the results of the 20 simulations with the highest accuracy were selected and averaged to produce the final forecast map of Cd content in the farmland soil of Zhejiang Province in 2050.

2.4 Geostatistical and spatial autocorrelation analysis

Geostatistics were first used in studies on the spatial variations of soil in the 1970s, and ordinary kriging is a representative geostatistical method (Webster and Oliver, 2011), which was developed for predicting the values of target variables at unvisited locations with data obtained from observation points (Li and Heap, 2011). The ordinary kriging method considers the spatial positions of the observation and estimated points, and considers the relative positions of the observation points, thus, its interpolation effect is better than that of the IDW method with a small sample size. The IDW method is readily employed and more suitable for datasets that deviate significantly from normal distribution (Hu et al., 2017b; Jia et al., 2020). Numerous studies have been published on the application of ordinary kriging and IDW in soil science, environmental science, social sciences, ecology, and other disciplines (Ismail et al., 2020; Lin et al., 2020; Singh et al., 2020; Jin et al., 2020).

The global Moran's I index was issued by Moran (1948), and detects the spatial

autocorrelation of a variable. The global Moran's I index is obtained as follows:

$$I = \frac{\sum_{i=1}^{N}\sum_{j=1}^{n} W_{ij}(X_i-\bar{X})(X_j-\bar{X})}{\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2 \sum_{i=1}^{N}\sum_{j=1}^{n} W_{ij}} \quad (4)$$

where $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, n represents the sample number, i and j denote the samples, and W_{ij} is the spatial weight matrix based on distance. Moran (1948) and Cliff and Ord (1981) provided a detailed description of the global Moran's method, which is presented in the Supplemental Material.

2.5 Data analysis and utilisation of the model

In this study, the summary statistics were processed using R Studio software, while the global Moran's I index was computed in GeoDa 1.12.1.59 (Anselin et al., 2006). The map developed through ordinary kriging and IDW was produced in ArcGIS 10.2 (ESRI Inc., USA), and the parameter sensitivity analysis and optimisation were conducted in PEST (Doherty, 2004). The model simulation based on the bagging algorithm was conducted using the bagging function of the ipred package in R Studio (Peters et al., 2002). The main steps of this study are presented in Fig. 4.

Figure 4.

3. Results and discussion

3.1 Summary statistics

The mean contents of Cd in the farmland soil of the research region were 0.19 and 0.22 mg kg^{-1} in 2003 and 2013, respectively (Table 1). The average soil Cd

content in 2013 was higher than that in 2003, indicating an increasing trend of Cd accumulation in the surveyed region between 2003 and 2013. Zhejiang Province is one of the most developed provinces in China, with numerous industrial enterprises that have rapidly expanded between 2003 and 2013 under rapid economic development. Large amounts of solid waste and sewage generated by industrial activities are discharged into the natural environment. Additionally, some pollutants accumulate in agricultural soil through atmospheric deposition and wastewater irrigation. These activities may contribute to increases in the soil Cd in Zhejiang Province.

In this study, the coefficient of variation (CV; %) indicates the extent of the variation in the Cd content. The high CV values in both 2003 (70.37%) and 2013 (76.50%) represent strong variations in the Cd content of the study area, further demonstrating that anthropogenic activities contribute to the accumulation of Cd in the soil (Hu et al., 2017c; Hu et al., 2018). The high values of kurtosis and skewness indicate the existence of anomalously high Cd values, particularly in 2003 (Table 1).

Table 1.

Some of the abnormally high Cd concentrations observed in grids during 2003 and 2013 can be explained by the changes in the surrounding enterprises. Regional soil pollution is typically due to the joint action of two processes: (1) the continuous background process, which is mainly affected by differences in the parent material and diffusion source; and (2) the quasi-point process, which is mainly attributable to

point-source pollution caused by anthropogenic activities. Variations due to the continuous background and random processes at a small scale are referred to as the background and point source effects (i.e., artificial random effects such as changes in polluting enterprises), respectively. Both of these effects can generate outliers and affect the prediction accuracy.

3.2 Sensitivity analysis and model parameter optimisation

In this study, the input data of the mass balance model included the initial soil Cd concentration measured in 2003, the soil Cd concentration measured in 2013, soil properties, sorption isotherm features, plant growth and uptake, and the amount of Cd added to soil from various sources. The Cd content in 2003 was used to train the model, and the predicted value of Cd in 2013 was used to validate the model by comparison with the Cd content measured in 2013.

In this study, we set the range of the parameter values (10^{-10}–10^{10}), and the default values of all parameters in PEST were used to define the searching space of the parameters. The most sensitive parameters affecting the prediction results were also identified by PEST. In this study, the most sensitive parameters were E_r, d, and R_p (Fig. S1), representing plant uptake, soil depth, and water leaching, respectively. The soil depth was assumed to be 20 cm, and the values of E_r and R_p were optimised and used in the mass balance model.

3.3 Model performance

As the original soil Cd concentration data exhibited a skewed distribution, the original data were pre-processed through logarithmic conversion to satisfy the normal

316 distribution before constructing the mass balance model. Several indices, including
317 the Pearson correlation coefficient (R), root-mean-squared-error (RMSE), mean
318 absolute error (MAE), and mean error (ME), were calculated to assess the model's
319 performance for estimating the Cd content at the different sampling sites in 2013. As
320 shown in Fig. 5, the R and RMSE values of our framework were 0.568 and 0.177 mg
321 kg^{-1}, respectively. Considering the high variation and complex sources of Cd in soil,
322 these results confirm the feasibility of our framework for predicting the soil Cd
323 content in farmland at a regional scale. Nevertheless, additional efforts should be
324 devoted to increasing the model performance in a further study.
325
326 **Figure 5.**
327
328 **3.4 Residual error analysis**
329 **3.4.1 Relationships between residual errors and input variable values**
330 As reported by Smith et al. (2015), when residual errors exhibit an obvious
331 structural relationship with the input data, a systematic bias exists and the model
332 neglects some important factors. In contrast, when the relationships are random, there
333 is no systematic bias and the residuals of the model represent random errors.
334 Therefore, in most studies, scatter diagrams of residuals are plotted to assess
335 heteroscedasticity (Smith et al., 2015). Autocorrelation was assessed using the test
336 statistic and the visual assessment of correlograms (Ljung and Box, 1978).
337 In this study, the bulk density (BK) and volume water content (V) of every

sample were measured. Therefore, we analysed the relationships between the residual errors of the prediction results and input values of BK and V to detect whether there was systematic bias in the prediction results. As shown in Table S2, the input values of V and BK were not significantly correlated with the residual errors. The relationships between the input values and residual errors of the target grids are also shown in Fig. 6. These results confirm that the structure of the mass balance model is satisfactory, and the model could approximately predict the future pollution scenario.

Figure 6.

3.4.2 Spatial pattern and autocorrelation test of residual errors

The residual errors of the prediction results in 2013 were the differences obtained by deducting the measured value from the predicted value in 2013 (Fig. S2). The global Moran's I value of the residual error of the predicted Cd content in 2013 was 0.00092 (Fig. S3), which indicates an almost random spatial distribution of the residual error of the prediction result. Additionally, as shown in Fig. S2, there was no clear spatial trend of the Cd content prediction residuals in 2013, indicating that there was no systematic bias in the predicted Cd content in 2013 as the residuals were randomly distributed across the research region.

The map of the residual errors of the prediction results for 2013 (Fig. 7) was produced by ordinary kriging. As shown in Fig. 7, high residual errors mainly occurred in the central and eastern coastal regions, with highly developed economic activities. In these areas, the higher strength of human activities leads to various

sources of Cd contamination and strong variations in soil Cd, increasing the difficulty in modelling the mass balance of soil Cd. Moreover, the soil Cd values for 2003 and 2013 that we used were the total Cd content, and the influences on the effectiveness of soil and plant absorption were not as direct as the effective state of Cd in soil based on the pH, organic matter, cation exchange capacity, and other soil properties (Hu et al., 2020b); thus, different forms of Cd may require a transition. This topic requires further in-depth studies.

Figure 7.

3.5 Forecast Cd pollution status in 2050 based on the bagging algorithm

On May 28, 2016, the Chinese government issued the National Soil Pollution Control Plan (http://www.gov.cn/zhengce/content/2016-05/31/content_5078377.htm), which strives to ensure that more than 95% of areas classified as security utilisation and priority protection zones should be made safe for use through landuse transformation by 2030. As we verified the feasibility of our developed model for estimating the temporal changes in the Cd content of farmland soil, the model was used to forecast the Cd content in sample grids of the study area in 2050. The summary statistics of the Cd content and corresponding degree of pollution in the farmland soil of Zhejiang Province in 2050 are shown in Table 2. In 2050, under the current trend, the mean Cd content in farmland soil will increase to 0.30 mg kg^{-1}.

Table 2.

Regarding the pollution classification, referring to the national soil environmental quality standard for farmland (GB 15618-2018), when the soil HM concentration is below the risk screening values, the risk for soil HM pollution is considered to be low, and this case is referred to as the priority protection class (Table S3). When the soil HM concentration exceeds the risk intervention value, the risk of soil contamination in agricultural land is considered to be high, and this is referred to as the strictly controlled class (Table S4). The neutral case is the security utilisation class.

Our results indicated that 346 grids were forecast as the priority protection class, accounting for 83.78% of all of the sampled grids. Furthermore, 67 grids were forecast as the security utilisation class, accounting for 16.22% of all grids. Spatial interpolation was subsequently employed to obtain the spatial pattern of the forecast Cd content in the farmland soil of Zhejiang Province in 2050. As the original content and content after the logarithmic conversion of soil Cd in the study area deviated from a normal distribution, the IDW was utilised to conduct spatial interpolation (Jie et al., 2013). A map of the forecast Cd content of farmland soil in Zhejiang Province during 2050 is presented in Fig. 8.

Figure 8.

As shown in Fig. 8 (a), the concentration of soil Cd in the farmland of some areas in eastern and northern Zhejiang Province will exceed 0.45 mg kg^{-1}, or even 0.55 mg kg^{-1}, in 2050, which exceeds the screening value of Cd in agricultural land. The highest soil Cd concentrations were mainly observed in the central and eastern coastal areas of Zhejiang Province. These areas, characterised by high-intensity industrial, commercial, and transportation activities, are the most developed regions of the province. Cd originating from traffic emissions, wastewater irrigation, industrial waste, and atmospheric deposition contributes to the relatively high Cd concentrations of these regions.

Considering the map of the Cd concentration in 2050 and latest Chinese national standard (GB 15618-2018), in combination with the soil pH and land-use types, the degree of Cd pollution in Zhejiang Province during 2050 was mapped, and the result is shown in Fig. 8 (b). As shown in Fig. 8 (b), if the current trend continues, 37.4% of the farmland in Zhejiang Province will be classified as a security utilisation region in 2050, and the other 62.3% of farmland will be classified as a priority region. Therefore, there is the potential risk of Cd pollution in the edible agricultural products planted in 37.4% of the farmland in Zhejiang Province, and the results suggest that measures, such as agronomic management and alternative planting strategies, should be adopted to reduce the risk (Hu et al., 2017b; Hu et al., 2019).

Moreover, some extremely high values were smoothed in the simulations based on bagging algorithms. This explains why the maximum Cd content in 2050 is 0.65, which is lower than the maximum Cd contents in 2003 and 2013. Therefore, some

local sites that were severely polluted by Cd are not presented in Fig. 8. Additionally, the negative effects of soil acidification on Cd pollution control should receive special attention. As significant soil acidification has been detected in most farmland soil of China (Guo et al., 2020), and acidic soil has a lower intervention value for Cd, the pollution risk of Cd in soil could increase. Numerous studies have revealed that Cd is more mobile in acidic soil; therefore, crops can absorb it more easily and it can accumulate in the human body via food chains (McBride, 2002; Li et al., 2005; Houben et al., 2013), posing a greater threat to human health. Therefore, substantial effort must be devoted to reduce Cd accumulation in soil and to prevent soil from becoming significantly acidified.

3.6 Limitations and perspectives

Although we adopted several measures to reduce model uncertainty and bias, and improve the model performance, our study has several limitations. First, although we included as many main sources and outputs of Cd in agricultural soil as possible, not all sources and outputs of soil Cd could be included in our mass balance model. Therefore, some uncertainty remains in our prediction results. Second, limited by time, labour, economic costs, and equipment, the values of some parameters were obtained from similar previous studies. Third, although we collected over 20,000 soil samples from the study area, after data pre-processing, only 413 grids contained soil samples in both 2003 and 2013. Most of these soil grids were distributed in the central, northern, and eastern coastal regions of Zhejiang Province, which also negatively affected the results. The inconsistencies in the strategies of the two surveys conducted

in 2003 and 2013 contributed to this limitation. Therefore, to improve the model performance and confirm our results, more soil samples should be collected referencing the sampling strategies of the surveys conducted in 2003 and 2013 in future studies to fully utilise the information of soil samples that were already collected. Additionally, field experiments must also be conducted to refine the parameter values. Finally, information related to additional input and output fluxes, such as crop straw, rock weathering, household solid waste, electronic waste, and surface runoff, should also be collected, which will facilitate more accurate prediction results.

Our results also have implications for the government and policymakers. As indicated by our results, Cd accumulation is highly related to the socio-economic development level, as confirmed by the overlap of areas with high Cd content and high economic development. This also demonstrates that anthropogenic activities, particularly industrial activities, contribute to Cd pollution the most. Numerous related studies have also reached this conclusion (Gan et al., 2019; Deng et al., 2020; Hu et al., 2020c). Additionally, soil acidification is a threat that could offset efforts to control Cd pollution, as acidic soil has a lower intervention threshold for Cd pollution. Therefore, we should focus on accurately appointing the sources of soil Cd and implementing specific measures to reduce the accumulation of Cd in farmland soil, and remain alert to the issue of soil acidification.

4. Conclusion

In this research, we constructed a novel framework combining the mass balance

model, PEST method, geostatistical method, spatial autocorrelation analysis, and bagging algorithm to model the mass balance of Cd in the farmland soil of Zhejiang Province, eastern China, during 2003 and 2013, and forecast the Cd concentrations in 2050. The sensitivity analysis found that E_r, d, and R_p were the main factors affecting the prediction results. Therefore, the two parameters (excluding soil depth) were optimised to ensure more robust and reliable performance. For the predicted Cd content in the farmland soil of the survey region in 2013, the R and RMSE values of our model were 0.568 and 0.177 mg kg^{-1}, respectively.

The results showed that the average Cd content of farmland soil in Zhejiang Province will increase to 0.30 mg kg^{-1} in 2050 under the current trend, and 37.4% of the farmland soil in Zhejiang Province will be classified as a security utilisation region. Measures, including agronomic regulations and alternative planting, should be adopted in such areas. However, additional surveys and experiments are required to confirm our results and improve the model performance. Furthermore, attention should be paid to issues such as local extremely polluted sites, soil acidification, and waste produced from industrial activities. Our study provides new insight for modelling HMs, particularly the Cd balance, and forecasting the temporal trends of HM pollution at the regional scale. The results of this study also provide critical information for the government and policymakers, and contribute to the development of more efficient and reasonable strategies and policies for preventing and controlling HM contamination in the survey region. For example, different policies could be implemented in priority protection, security utilisation, and strictly controlled areas.

In agricultural and rural regions, more effort should be devoted to reducing the use of fertilisers and pesticides. In industrial regions, traditional industries (energy and pollution-intensive enterprises) could be upgraded into environmentally friendly enterprises. Finally, the results also provide inspiration for conducting similar analyses in other similar regions.

Acknowledgements

This work was supported by the National Key Research and Development Program of China (2018YFC1800105), the National Natural Science Foundation of China (41771244), and the Consulting Research Project of the Chinese Academy of Engineering (2019-XZ-24). We acknowledge the support received by Bifeng Hu from the China Scholarship Council (201706320317) for three years of Ph.D. study in the French National Institute for Agriculture, Food, and Environment and Orléans University in France.

References

Anselin, L., Syabri, I., Kho, Y., 2006. GeoDa: an introduction to spatial data analysis. Geographical analysis. 38(1), 5-22.

Breiman, L., 1996. Bagging predictors. Machine Learning. 24(2), 123-140.

Chen, W., Chang, A.C., Wu, L., 2007. Assessing long-term environmental risks of trace elements in phosphate fertilizers. Ecotox. Environ. Safe. 67(1), 48-58.

DalCorso, G., Farinati, S., Maistri, S., Furini, A., 2008. How Plants Cope with Cadmium: Staking All on Metabolism and Gene Expression. J. Integr. Plant. Biol. 50(10), 2368-2380.

Deng, M.H., Zhu, Y.W., Shao, K., Zhang, Q., Ye, G.H., Shen, J., 2020. Metals source apportionment in farmland soil and the prediction of metal transfer in the soil-rice-human chain. J. Environ. Manage. 260, 110092.

Doherty, J., Brebber, L., Whyte, P., 1994. PEST: Model-independent parameter estimation. Watermark Computing, Corinda, Australia. 122, 336.

Doherty, J., 2004. PEST model-independent parameter estimation user manual. Watermark Numerical Computing, Brisbane, Australia. 3338-3349.

Gan, Y.D., Huang, X.M., Li, S.S., Liu, N., Li, Y.C.C., Freidenreich, A., Wang, W.X., Wang, R.Q., Dai, J.L., 2019. Source quantification and potential risk of mercury, cadmium, arsenic, lead, and chromium in farmland soils of Yellow River Delta. J. Clean. Prod. 221, 98-107.

Guo, H. Y., Zhu, J. G., Zhou, H., Sun, Y. Y., Yin, Y., Pei, D. P., Ji, R., Wu, J.C., Wang, X.R., 2011. Elevated CO_2 Levels Affects the Concentrations of Copper and Cadmium in Crops Grown in Soil Contaminated with Heavy Metals under Fully Open-Air Field Conditions. Environ. Sci. Technol. 45(16), 6997-7003. doi:10.1021/es2001584.

Guo, J.H., Liu, X.J., Zhang, Y., Shen, J.L., Han, W.X., Zhang, W.F., Christie, P., Goulding, K.W.T., Vitousek, P.M., Zhang, F.S., 2010. Significant acidification in major Chinese croplands. Science. 327(5968), 1008-1010.

Houben, D., Evrard, L., Sonnet, P., 2013. Mobility, bioavailability and pH-dependent leaching of cadmium, zinc and lead in a contaminated soil amended with biochar. Chemosphere. 92(11), 1450-1457.

Hu, B.F., Chen, S.C., Hu, J., Xia, F., Xu, J.F., Li, Y., Shi, Z., 2017a. Application of portable XRF and VNIR sensors for rapid assessment of soil heavy metal pollution. PLoS ONE. 12, 1–13.

Hu, B.F., Wang, J.Y., Jin, B., Li, Y., Shi, Z., 2017b. Assessment of the potential health risks of heavy metals in soils in a coastal industrial region of the Yangtze River Delta. Environ. Sci. Pollut. R. 24(24), 19816-19826.

Hu, B.F., Jia, X.L., Hu, J., Xu, D.Y., Xia, F., Li, Y., 2017c. Assessment of heavy metal pollution and health risks in the soil-plant-human system in the Yangtze river delta, China. Int. J. Environ. Res. Pub. He. 14(9), 1042.

Hu, B.F., Zhao, R.Y., Chen, S.C., Zhou, Y., Jin, B., Li, Y., Shi, Z., 2018. Heavy Metal Pollution Delineation Based on Uncertainty in a Coastal Industrial City in the Yangtze River Delta, China. Int. J. Environ. Res. Pub. He. 15(4), 710.

Hu, B.F., Shao, S., Fu, Z.Y., Li, Y., Ni, H., Chen, S.C., Zhou, Y., Jin, B., Shi, Z., 2019. Identifying heavy metal pollution hot spots in soil-rice systems: A case study in South of Yangtze River Delta, China. Sci. Total. Environ. 658, 614-625.

Hu, B.F., Shao, S., Fu, T.T., Fu, Z.Y., Zhou, Y., Li, Y., Qi, L., Chen, S.C., Shi, Z., 2020a. Composite assessment of human health risk from potentially toxic elements through multiple exposure routes: A case study in farmland in an important industrial city in East China. J. Geochem. Explor. 210, 106443

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

Hu, B.F., Shao, S., Ni, H., Fu, Z.Y., Hu, L.S., Zhou, Y., Min, X.X., She, S.F., Chen, S.C., Huang, M.X., Zhou, L.Q., Li, Y., Shi, Z., 2020c. Current status, spatial features, health risks, and potential driving factors of soil heavy metal pollution in China at province level. Environ. Pollut. 262, 114308.

Hu, Y., Yong, X., Wang, J., Sun, X., Zhang, A., 2004. Feasibility Study of Geo-spatial Analysis Using Grid Computing. Paper presented at the Computational Science -

ICCS 2004, 4th International Conference, Kraków, Poland, June 6-9, 2004, Proceedings, Part IV.

Ismail, H.Y., Shirazian, S., Singh, M., Whitaker, D., Albadarin, A., Walker, G.M., 2020. Compartmental approach for modelling twin-screw granulation using population balances. Int. J. Pharmaceut. 576, 118737.

Jia, X.L., Hu, B.F., Marchant, B. P., Zhou, L.Q., Shi, Z., Zhu, Y.W., 2019. A methodological framework for identifying potential sources of soil heavy metal pollution based on machine learning: A case study in the Yangtze Delta, China. Environ. Pollut. 250, 601-609.

Jia, X.L., Fu, T.T., Hu, B.F., Shi, Z., Zhou, L.Q., Zhu, Y.W., 2020. Identification of the potential risk areas for soil heavy metal pollution based on the source-sink theory. J. Hazard. Mater. 393, 122424.

Jin, X., Wang, G., Tang, P., Hu, C., Liu, Y., Zhang, S., 2020. 3D geological modelling and uncertainty analysis for 3D targeting in Shanggong gold deposit (China). J. Geochem. Explor. 210, 106442.

Keller, A., Von, S.B., Se, V.D.Z., Schulin, R., 2001. A stochastic empirical model for regional heavy-metal balances in agroecosystems. J. Environ. Qual. 30(6), 1976.

Kikuchi, T., Okazaki, M., Toyota, K., Motobayashi, T., Kato, M., 2007. The input–output balance of cadmium in a paddy field of Tokyo. Chemosphere. 67(5), 920-927.

Li, J., Heap, A.D. 2011. A review of comparative studies of spatial interpolation methods in environmental sciences: Performance and impact factors. Ecol. Inform. 6(3-4), 228-241. doi:10.1016/j.ecoinf.2010.12.003.

Li, Z., Li, L., Chen, G.PJ., 2005. Bioavailability of Cd in a soil–rice system in China: soil type versus genotype effects. Plant. Soil. 271(1-2), 165-173.

590 Lin, X., Hu, Y., Meng, G., Zhang, M., 2020. Geochemical patterns of Cu, Au, Pb and
591 Zn in stream sediments from Tongling of East China: Compositional and
592 geostatistical insights. J. Geochem. Explor. 2020, 210: 106457.
593 Ljung, G.M., Box, G.E., 1978. On a measure of lack of fit in time series models.
594 Biometrika. 65(2), 297-303.
595 McBride, M.B., 2002. Cadmium uptake by crops estimated from soil total Cd and pH.
596 Soil. Sci. 167(1), 62-67.
597 Ministry of Ecology and Environment of the People's Republic of China, and State
598 Administration for Market Regulation, 2018. Soil Environmental Quality-risk
599 Control Standard for Soil Contamination of Agricultural Land (GB 15618-2018).
600 China Environment Press (in Chinese).
601 Moran, P. A., 1948. The Interpretation of Statistical Maps. Journal of the royal
602 statistical society series b-methodological, 10(2), 243-251.
603 NSPCIR, Ministry of Environmental Protection, Ministry of Land and Resources.,
604 2014. The National Soil Pollution Condition Investigation Report [EB/OL].
605 http://www.gov.cn/foot/site1/20140417/782bcb88840814ba158d01.pdf.Availanle
606 at 15th March 2020.
607 Peng, C., Wang, M., Chen, W., 2016. Modelling cadmium contamination in paddy soils
608 under long-term remediation measures: Model development and stochastic
609 simulations. Environ. Pollut. 216, 146-155.
610 Peng, J., Biswas, A., Jiang, Q.S., Zhao, R.Y., Hu, J., Hu, B.F., Shi, Z., 2019.
611 Estimating soil salinity from remote sensing and terrain data in southern Xinjiang
612 Province, China. Geoderma. 337, 1309-1319.
613 Peters, A., Hothorn, T., Lausen, B., 2002. ipred: Improved predictors. R news. 2(2),
614 33-36.

615 Pham, B.T., Prakash, I., 2019. A novel hybrid model of Bagging-based Naive Bayes
616 Trees for landslide susceptibility assessment. B. Eng. Geol. Environ. 78(3),
617 1911-1925. doi:10.1007/s10064-017-1202-5
618 Pouchard, L., Cinquini, L., Drach, B., Middleton, D., Bernholdt, D., Chanchio, K.,
619 Foster, I., Nefedova, V., Brown, D., Fox, P., Garcia, J., Strand, G., Williams, D.,
620 Chervenak, A., Kesselman, C., Shoshani, A., Sim, A., 2003. An ontology for
621 scientific information in a grid environment: the Earth System Grid[C]//CCGrid.
622 3rd IEEE/ACM International Symposium on Cluster Computing and the Grid,
623 2003. Proceedings. IEEE. 626-632.
624 Ronzan, M., Piacentini, D., Fattorini, L., Della Rovere, F., Eiche, E., Riemann, M.,
625 Falasca, G., 2018. Cadmium and arsenic affect root development in Oryza sativa L.
626 negatively interacting with auxin. Environ. Expe. Bot. 151, 64-75.
627 Rose, K.A., Megrey, B.A., Werner, F.E., Ware, D.M., 2007. Calibration of the
628 NEMURO nutrient–phytoplankton–zooplankton food web model to a coastal
629 ecosystem: Evaluation of an automated calibration approach. Ecol. Model.
630 202(1-2), 38-51.
631 Sawut, R., Kasim, N., Maihemuti, B., Hu, L., Abliz, A., Abdujappar, A., Kurban, M.,
632 2018. Pollution characteristics and health risk assessment of heavy metals in the
633 vegetable bases of northwest China. Sci. Total. Environ. 642, 864-878.
634 Sciarrillo, R., Zuzolo, D., Cicchella, D., Iannone, F., Cammino, G., Gaurino, G., 2020.
635 Contamination and ecological risk assessment of the seaport of Naples (Italy):
636 Insights from marine sediments. J. Geochem. Explor. 210, 106449.
637 Shao S, Hu B F, Fu Z Y,Wang J Y, Lou G, Zhou Y, Jin B, Li Y, Shi Z. 2018.
638 Source identification and aportionment of trace elements in soils in the Yangtze

River Delta, China. International Journal of Environmental Research and Public Health, 15,1240.

Shen, Z.T., Fan, X.L., Hou, D.y., Jin, F., O'Connor, D., Tsang, D., Ok, Y., Alessi, D., 2019. Risk evaluation of biochars produced from Cd-contaminated rice straw and optimization of its production for Cd removal, Chemosphere. 233,149-156.

Singh, G., Rishi, M.S., Herojeet, R., Kaur, L., Priyanka, Sharma, K., 2020. Multivariate analysis and geochemical signatures of groundwater in the agricultural dominated taluks of Jalandhar district, Punjab, India. J. Geochem. Explor. 208, 106395.

Six, L., Smolders, E., 2014. Future trends in soil cadmium concentration under current cadmium fluxes to European agricultural soils. Sci. Total. Environ. 485-486(3), 319-328.

Smith, T., Marshall, L., Sharma, A., 2015. Modeling residual hydrologic errors with Bayesian inference. J. Hydrol. 528, 29-37.

Van Pelt, R.S., Shekhter, E.G., Barnes, M.A.W., Duke, S.E., Gill, T.E., Pannell, K.H., 2020. Spatial and temporal patterns of heavy metal deposition resulting from a smelter in El Paso, Texas. J. Geochem. Explor. 210, 106414.

Webster, R., Oliver, M.A., 2011. Geostatistics for Environmental Scientists, Second Edition. John Wiley & Sons.

Xiang, L., Liu, P., Jiang, X., Chen, P., 2019. Health risk assessment and spatial distribution characteristics of heavy metal pollution in rice samples from a surrounding hydrometallurgy plant area in No. 721 uranium mining, East China. J. Geochem. Explor. 207, 106360.

Yan, F.P., Shangguan, W., Zhang, J., Hu, B.F., 2020. Depth-to-bedrock map of China at a spatial resolution of 100 meters. Sci. Data. 7(1), 1-13.

664 Zhou, Y., Hartemink, A., Shi, Z., Liang, Z.Z., Lu, Y.L., 2019. Land use and climate
665 change effects on soil organic carbon in North and Northeast China. Sci. Total.
666 Environ. 647,1230-1238.

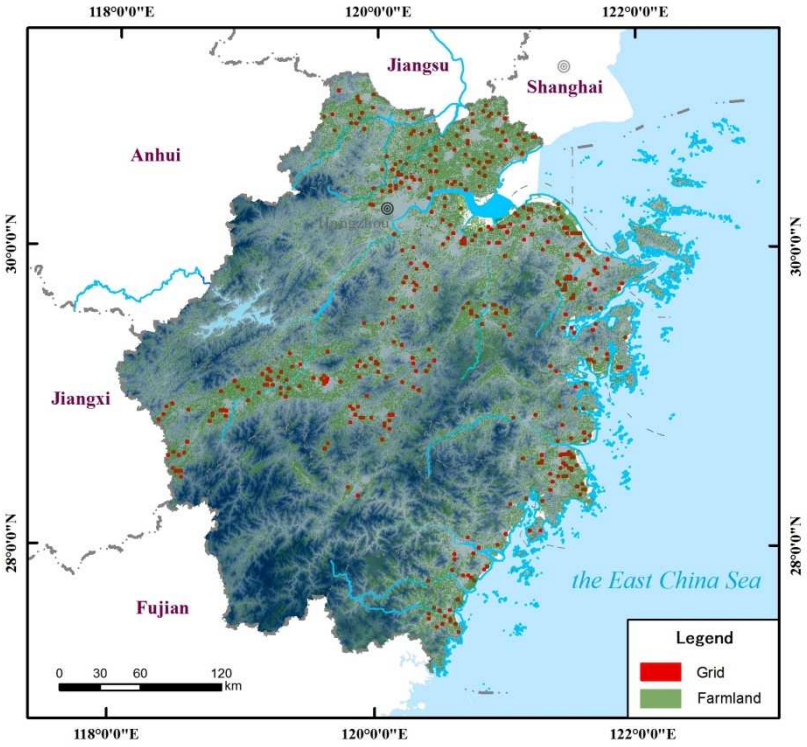


Figure 1. Map of the locations of the sampling grids

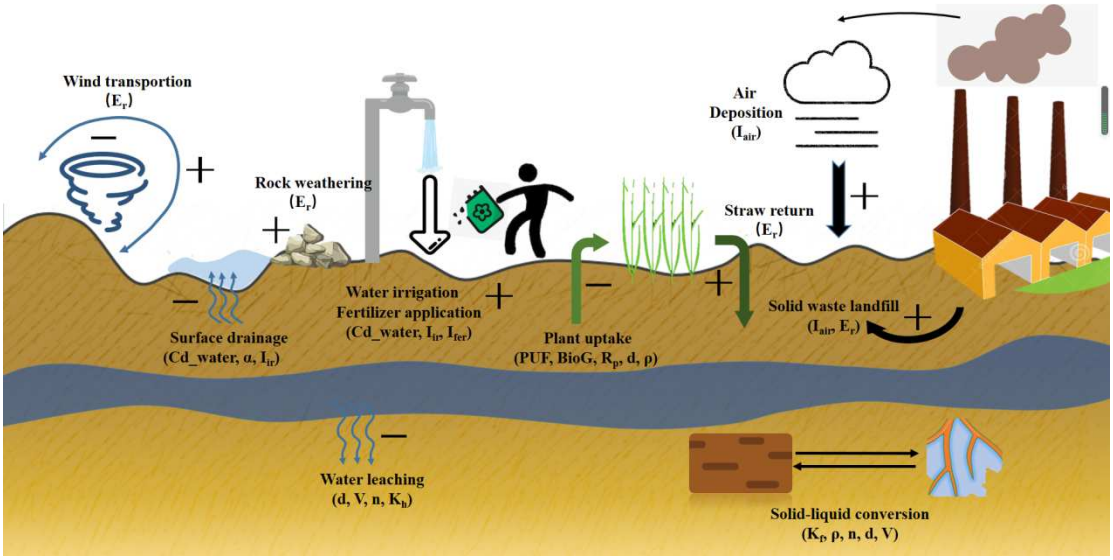


Figure 2. The main outputs and inputs of the source and sink equilibrium system of soil Cd

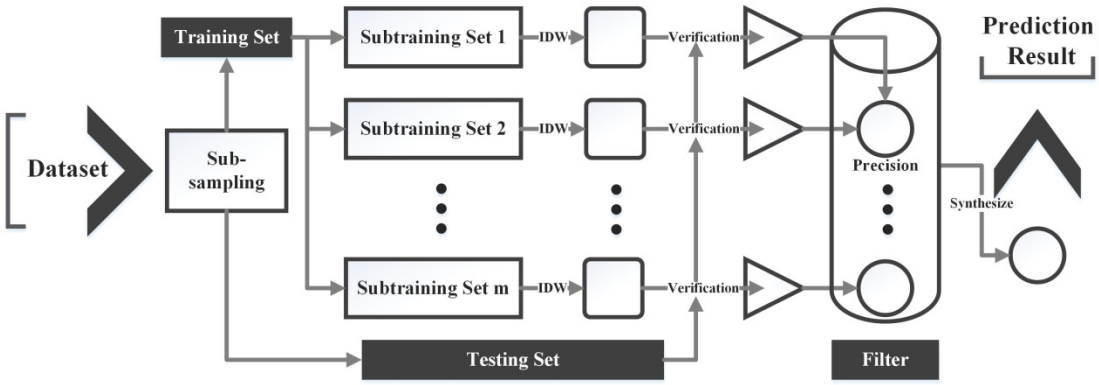


Figure 3. The forecast process of soil Cd in 2050 based on the bagging method

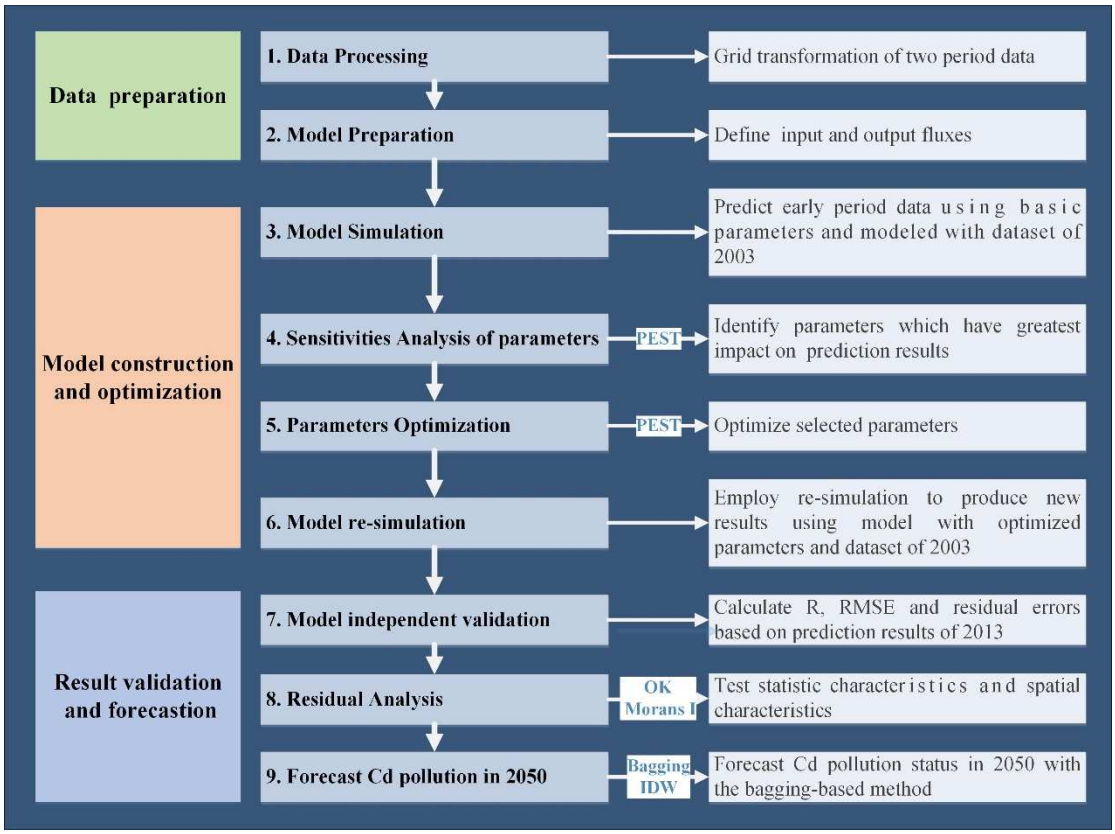


Figure 4. Flow chart of the framework for modeling soil cadmium balance

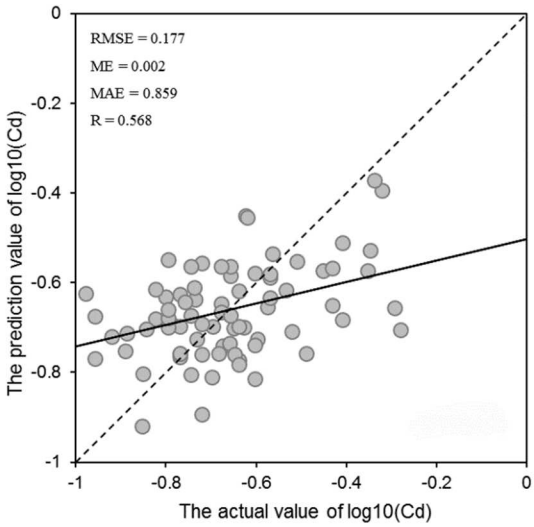


Figure 5. Relationships between measured and predicted values of Cd in 2013

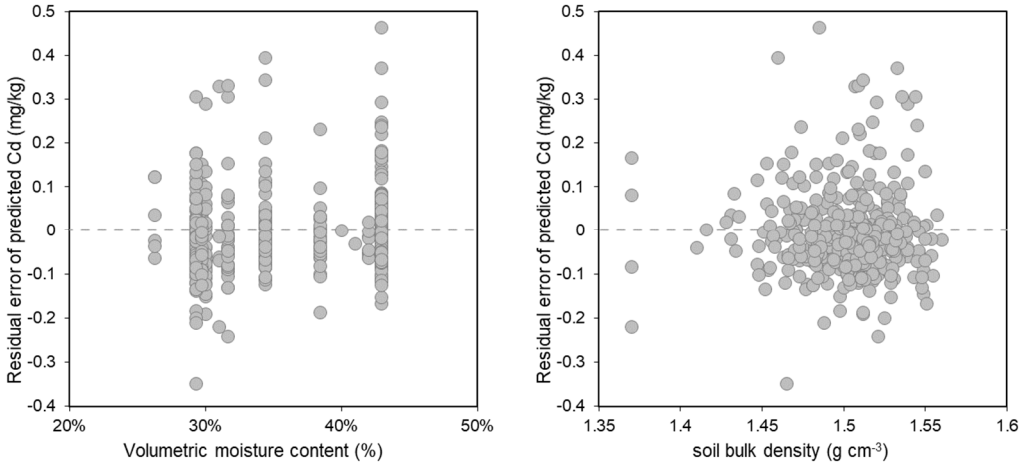


Figure 6. Relationships of the input values and residual errors

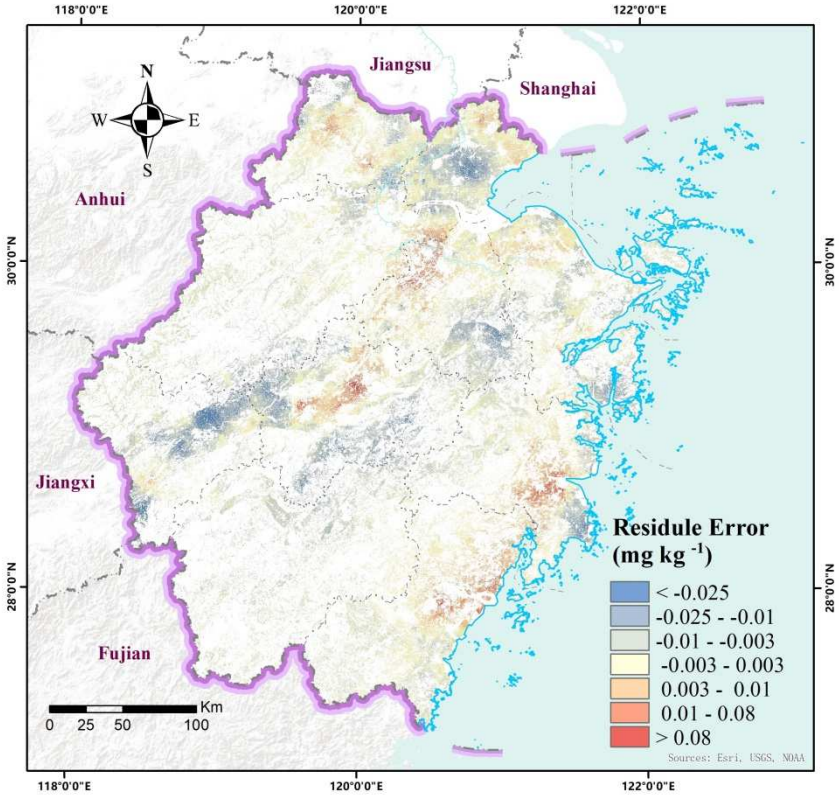


Figure 7. Map of the residual error of model simulation results

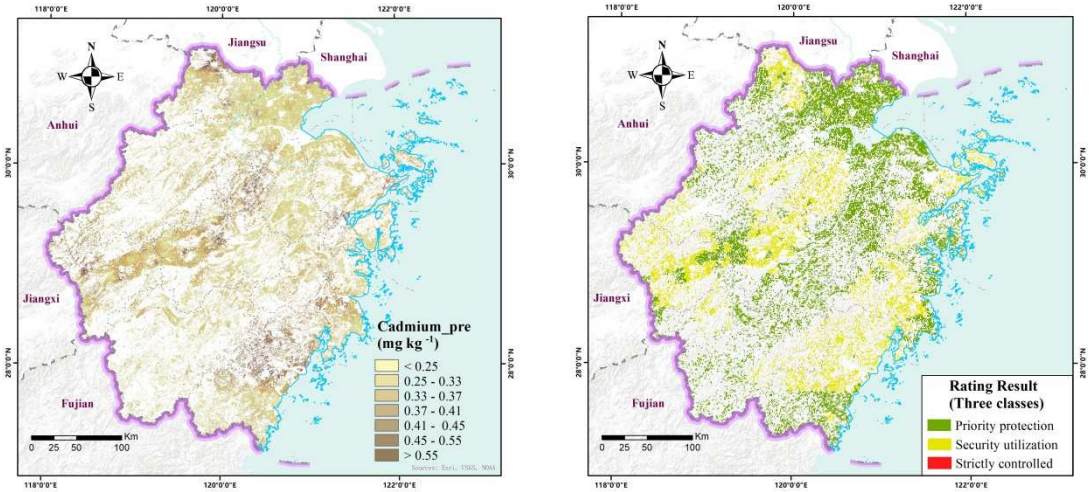


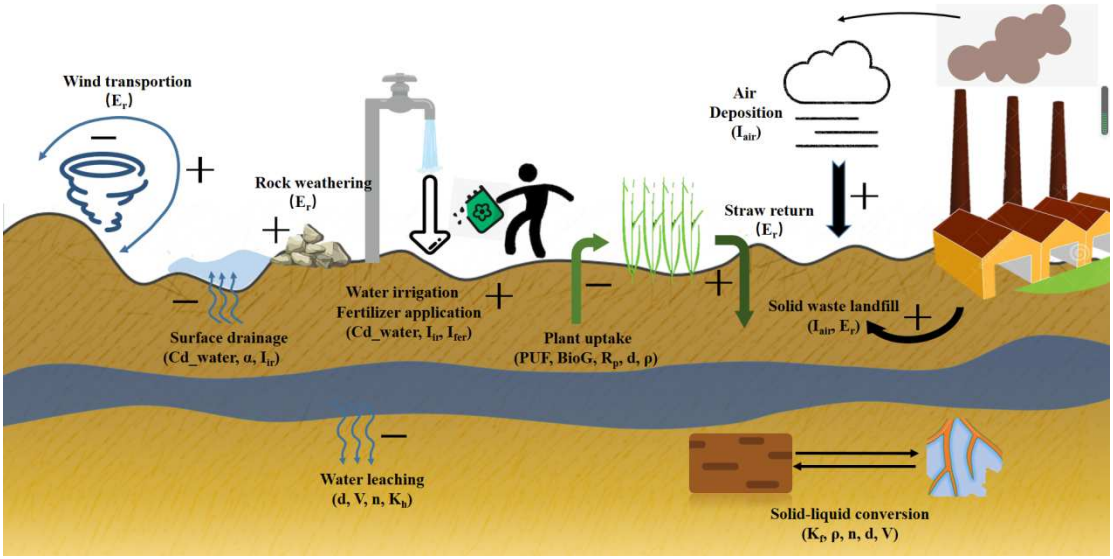
Figure 8. Map of (a) forecasted Cd content in farmland soil in the study area in 2050 and (b) classification of the environmental quality assessment of Cd

Table 1. Descriptive statistics of Cd concentrations in the research area in the years 2003 and 2013 (mg kg^{-1})

Year	Number of grids	Mean	Std	Min	Max	CV (%)	Skewness	Kurtosis
2003	413	0.19	0.14	0.07	2.21	70.37	9.90	130.10
2013	413	0.22	0.17	0.01	1.83	76.50	5.02	37.29

Table 2. Summary statistics of Cd content and degree of pollution in farmland soil in Zhejiang Province in 2050

Element	Number of grids	Content (mg kg^{-1})			Priority protection class		Security utilization class	
		Min	Mean	Max	Number	Proportion (%)	Number	Proportion (%)
Cd	413	0.24	0.30	0.65	346	83.78	67	16.22


Wind transportion
(E_r)
Rock weathering
(E_r)
Air
Deposition
(I_{air})
Straw return
(E_r)
Water irrigation
Fertilizer application
(Cd_water, I_{ir}, I_{fer})
Plant uptake
(PUF, BioG, R_p, d, ρ)
Solid waste landfill
(I_{air}, E_r)
Surface drainage
(Cd_water, α, I_{ir})
Water leaching
(d, V, n, K_h)
Solid-liquid conversion
(K_f, ρ, n, d, V)